Author’s response to reviews

Title: The effects of aerobic, resistance, and meditative movement exercise on sleep in individuals with depression: protocol for a systematic review and network meta-analysis

Authors:

Gavin Brupbacher (gavin.brupbacher@unibas.ch)
Heike Gerger (heike.gerger@unibas.ch)
Monika Wechsler (monika.wechsler@unibas.ch)
Thea Zander-Schellenberg (thea.zander@unibas.ch)
Doris Straus (doris.straus@oberwaid.ch)
Hildburg Porschke (hildburg.porschke@oberwaid.ch)
Markus Gerber (markus.gerber@unibas.ch)
Roland von Känel (roland.vonkaenel@usz.ch)
Arno Schmidt-Trucksäss (arno.schmidt-trucksaess@unibas.ch)

Version: 1 Date: 11 Jan 2019

Author’s response to reviews:

Please see the uploaded point-to-point response for tracked changes in colour.

Reviewer #1

Overall, clearly written, sound plan of analysis. This is an important undertaking. The section that seems to require the most attention is the background. I have provided several suggestions to help clarify the background. The remainder of the paper is quite clear.

Point 1: Line 47-lifetime prevalence where? worldwide? Using many terms in the first paragraph and throughout background, i.e. unipolar depression, depression, major depression. I would suggest explaining how these terms differ and how you are distinguishing their use.
Reply 1: We have specified the prevalence (indeed, this source refers to worldwide prevalence) rate. We have used the term “unipolar depression” more consistently.

Worldwide lifetime prevalence of unipolar depression is estimated to be 10% [1]. Unipolar depression is the leading cause of burden of disease…

Point 2: Line 50: change to $250 billion

Reply 2: We are not sure whether you refer to the missing “$” sign or whether you recommend to change the number (from 210 to 250). We have included the dollar sign in the manuscript. As to the number: we cite Greenberg et al. 2010 (http://dx.doi.org/10.4088/JCP.14m09298).

We have made the following change:

was estimated to be $210 billion in the USA in 2010

Point 3: Line 51 change increase to increases.

Reply 3: We have made this change in the manuscript.

depression also substantially increases the risk of all-cause mortality

Point 4: Line 55: This is a huge range of insomnia prevalence, how will you address this in your review?

Reply 4: We address this issue in our review by reporting sleep quality at baseline in the summary table of included studies.

This is indeed a large range. As we have mentioned this is likely due to heterogeneous operationalization of insomnia and depression. We have omitted one source and changed the statement (see below). We have chosen this approach since this is a methodological discussion which we do not want to enter at this stage of the manuscript.

This is reflected by the high prevalence (up to 90%) of co-occurring insomnia in individuals with depression [8].
Point 5: Line 60-61: one of the most... is insomnia OR Some of the most prevalent... ARE insomnia symptoms

Reply 5: We have made this change in the manuscript.

Some of the most prevalent residual symptoms after treatment response or remission in this clinical population are insomnia symptoms.

Point 6: Line 64: Whose treatment guidelines? Are they different depending on geographic location or are there universal guidelines?

Reply 6: We have clarified this point in the manuscript.

According to the British national (NICE) treatment guidelines.

Point 7: Lines 78-79: Include the citations of the meta-analyses here

Reply 7: We have made this change in the manuscript.

The effects of exercise on depressive symptoms have been summarized in multiple meta-analyses [27–30].

Point 8: Lines 99-100: consider changing to "chronic exercise improves thermoregulation"

Reply 8: We have made this change in the manuscript.

Chronic exercise improves thermoregulation

Point 9: Lines 107-109: I gather this information is regarding circadian rhythms, but the explanation is very unclear as a person unfamiliar with sleep physiology. Define phase shifts as an attribute of sleep, consider changing 'in heterogeneous population' - do you mean that across any population this occurs or within a highly diverse population? How has this been studied?
Independent of light—this needs clarification as well. Do you mean that exercise can induce a phase shift regardless of the presence of sunlight? Briefly define entrainment. This all may be well understood among sleep researchers, but less so for readers whose focus is on exercise or depression.

Reply 9: We have added a brief description of phase shift and have clarified what we mean when referring to a heterogeneous population (i.e., in terms of age and fitness) as well as an effect independent of light. We have also added a description of entrainment.

Exercise has been shown to consistently produce phase shifts (i.e. changes in circadian rhythm within the 24h cycle) in individuals of different ages and fitness levels. This effect has been found in individuals irrespective of age and cardiorespiratory fitness as well as independent from the effect of light. [58]. Therefore, exercise may act as a so-called 'zeitgeber' positively affecting entrainment (i.e. the synchronization of the endogenous and exogenous rhythms).

Point 10: Line 115: the first point should become the last point—building up to the gap in the literature from the importance instead of the other way around.

Reply 10: We have made this change in the manuscript.

The rationale for this review can be summarized in four points. 1) Sleep disturbances are of high prognostic relevance for remission in depression [11]. 2) Current therapies have a dissatisfactory benefit-to-risk-ratio. 3) Exercise has been shown to have positive effects on depression [28, 36, 83] as well as sleep [39, 41, 49, 51, 53]. 4) To the best of our knowledge, no systematic review has been performed to ascertain the effects of aerobic, resistance, and meditative-movement exercise on sleep in people with depression.

Point 11: Line 118. H with serious adverse effects? What is H?

Reply 11: This sentence should not have been included in the manuscript. Please accept our apologies. The change can be seen in the response to the previous point.

Point 12: Lines 120-122: Hypnotics come out of nowhere, but its presence in your abstract signifies its importance. This should be defined in the background prior to this point.
Reply 12: We have now addressed the adverse effects of hypnotics earlier in the text (last paragraph of the section “Description of the condition”):

Hypnotics are frequently prescribed for insomnia. However, they have a poor benefit-to-risk ratio with serious adverse effects including cognitive impairment, injury from falls and automobile accidents (including in younger individuals), cancer, suicide, and hypnotic withdrawal insomnia [25, 26].

Point 13: Line 155: Explain rationale for including acute and chronic and how these might differ or how you will account for their difference in your analysis.

Reply 13: We have addressed this point as follows:

We have not placed restrictions on the duration of the intervention period in order to include the maximum number of trials in this review. Potential statistical heterogeneity or inconsistency due to this factor will be explored (see below).

We had already addressed the analytical issue in the section “Data synthesis”: “If the number of studies allows it, theoretically driven subgroup analyses will be done according to population (e.g. severity of depression), duration of intervention…”

Point 14: Lines 166-167: Since this is your primary outcome, I think it would be useful to include the scales you expect (the way you did for the depression outcomes above). What scales are commonly used?

Reply 14: We had already addressed this point and further elaborated it: The primary outcome will be standardized mean differences (SMD) of sleep quality at post-exercise-intervention and at the last available follow-up assessment, measured by self-reports (e.g. PSQI [80], ISI [81]) or clinician ratings (sleep related HAM-D items [82]).

Point 15: Lines 223-224: needs punctuation to clarify: "Disagreement will be resolved by consensus. If no consensus can be reached, disagreement will be resolved by adjudication of a designated third reviewer.

Reply 15: We have made this change in the manuscript.
Disagreement will be resolved by consensus. If no consensus can be reached, disagreement will be resolved by adjudication of a designated third reviewer (AST).

Point 16: Line 248: remove first sentence
Reply 16: We have made this change in the manuscript.

Amongst others, the following information will be extracted from each study.

Point 17: Lines 281-284: This information is important, consider including in your background
Reply 17: We have included an additional sentence in the background: Insomnia has a negative impact on health-related quality of life and daytime functioning [9, 10].

Point 18: Line 323: comparator treatments- do you mean control group treatments?
Reply 18: We used the word “comparator” since we consider the meaning of the term “control group” to be too narrow in the context of network meta-analyses. Some of the studies which will be included in the network meta-analysis are multi-arm trials. Especially in these cases we deem the term “comparator” to be more appropriate.

Data Synthesis: Clear, detailed. Well-written.
Discussion: Suitable.

Reviewer #2

GENERAL COMMENTS Thank you for the opportunity to review this protocol to conduct a systematic review with network meta-analysis on aerobic exercise and sleep quality in patients with unipolar depression. This is an important topic and the authors, overall, have done a good
job in utilizing available and up-to-date resources to conduct their proposed network meta-analysis.

However, a major concern that I have is that I believe that it will be highly unlikely that they will retrieve a sufficient number of studies to conduct and type of meta-analysis, network or otherwise. This is not their fault, but rather, the hypothesized lack of available randomized trials that will meet their inclusion criteria. I elaborate on this in my specific comments that follow.

Finally, much of the paper reads more like an outline than a manuscript that flows well. Lines 57 through 60 is one of many examples of this.

My specific comments appear below with reference to the line numbers that I believe the reviewer versus journal inserted.

SPECIFIC COMMENTS *

Point 1: Line 14 (Abstract) - Since MEDLINE is nested in PubMed, why not just search PubMed since it will provide you with access to additional biomedical content? *

Reply 1: This is a good point. We have made this change in the manuscript.

following online databases: PubMed,

Point 2: Lines 19 and 20 - Suggest that you delete the two commas in this sentence. Also, since you plan on conducting a network meta-analysis, don't you mean "randomized trials" versus "randomized controlled trials" here? I'm asking because one of the very reasons for conducting a network meta-analysis is for the purpose of including studies that compare the effects of two or more interventions head-to-head and which do not include a comparative control group. *

Reply 2: We concur and have indeed otherwise used the term “comparator” throughout the manuscript. We have made this change in the manuscript.

Randomized trials using aerobic, resistance or meditative movement exercise interventions which target sleep as a primary or secondary outcome will be included.

Point 3: Lines 30-31 - Since there are multiple PRISMA checklists, suggest that you be specific in saying that you will follow the PRISMA Guidelines for network meta-analysis. *
Reply 3: We have made this change in the manuscript.

The systematic review and network meta-analysis will be presented according to the PRISMA for Network Meta-Analyses (PRISMA-NMA) guideline.

Point 4: Lines 39 - While I traditionally see a PROSPERO registration protocol number upon submission of a protocol for publication, your approach is perfectly logical in my opinion. *

Point 4: We have submitted the PROSPERO registration, but at the time of submission there was a backlog of 20 days. We will include the registration number as soon as we receive it.

Point 5: Line 40 - Assuming my earlier comment is correct, replace "randomized controlled trial" with "randomized trial". *

Reply 5: Yes, your comment was correct. We have made this change in the manuscript.

Keywords: exercise, depression, sleep, randomized trial, systematic

Point 6: Line 50 - Suggest you insert "dollars" after the word "billion". *

Reply 6: Reviewer 1 also made this comment. We have inserted a “$” before the number. Estimated to be $210 billion in the USA in 2010

Point 7: Line 51 - Replace "increase" with "increases" *

Reply 7: We are not sure if you recommend to replace “increase” with “increases”. If this is the case, we have changed the manuscript accordingly.

In addition to the direct and often debilitating symptoms, unipolar depression also substantially increases the risk of all-cause mortality
Point 8: Line 75 - Did you mean to say "decades" or "decade" here? *

Reply 8: Thank you for your correction. We meant the last 10 years and have changed the manuscript accordingly.

In the light of this evidence, interest in adjuvant and alternative therapies, especially exercise, has increased in the last decade.

Point 9: Lines 78 through 84 - If you're summarizing the general literature on the effects of exercise on depressive symptoms, there are multiple other systematic reviews with meta-analyses that should probably be cited and mentioned. *

Reply 9: We agree and have added three references (doi: 10.1016/j.jad.2016.03.063, 10.1016/j.neubiorev.2015.11.012 and 10.1002/da.22842) and consider this to be a sufficient summary of the literature.

Point 10: Lines 83 - Delete the comma after "relevant". *

Reply 10: We have made this change in the manuscript.

This is relevant because depression is known to increase the risk of cardiovascular mortality and morbidity.


In addition to the above, did you search for systematic reviews of previous systematic reviews with meta-analysis on exercise and sleep? Here’s one: 1. Kelley GA, Kelley KS. Exercise and sleep: A systematic review of previous meta-analyses. J Evid Based Med. 2017;10:11. *

Reply 11: Thank you for the provided citations. We are confident that this only strengthens our argument for the necessity of our review. We have included all citations mentioned in point 11.

These findings have been replicated in populations with sleep complaints [39, 41–43] and chronic disorders [44–49], and confirmed by a meta-analysis of previous meta-analyses [50].

Point 12: Line 98 - You probably need to fix the following in accordance with journal guidelines "(cf. [34, 35] for review)" *

Reply 12: We have changed it to the following:

Multiple mechanisms of action, including ones which involve hyperarousal, have been proposed to explain the effect exercise has on sleep (confer the reviews of Buman and King (2010) [65] and Uchida et al. 2012 [66] for aerobic exercise).

Point 13: Line 107 - Depending on what you mean here, either replace "population" with "populations" or insert "a" before the word heterogeneous. Also, please provide a reference for this statement. *

Reply 13: Reviewer 1 also commented on this sentence. We have changed the sentence to the following:

Aerobic exercise has also been shown to consistently produce phase shifts (i.e. changes in circadian rhythm within the 24h cycle) in individuals of different ages and fitness levels. This effect has been found in individuals irrespective of age and cardiorespiratory fitness as well as independent from the effect of light. [76].
To the best of our knowledge, no systematic review has been performed to ascertain the effects of aerobic, resistance, and meditative movement exercise on sleep in people with depression.

A secondary goal is to ascertain the effects of exercise on sleep duration, sleepiness, daytime functioning, use of hypnotics, and adverse events (injuries, cardiovascular incidences, etc.).

The combination of aerobic and resistance training is possible. Please note the following sentences (lines 199-202 in the new submission). “Exercise can be part of a multicomponent intervention. Multicomponent interventions in which exercise was not a dominant part (i.e., exercise was one of four or more intervention modules) will be excluded.”
We have now included the definition of aerobic, resistance and meditative movement exercise, as follows:

Included trials must allocate subjects to at least one of the following: aerobic, resistance or meditative-movement exercise intervention. Aerobic exercise is defined as "any exercise that primarily uses the aerobic energy-producing systems, can improve the capacity and efficiency of these systems, and is effective for improving cardiorespiratory endurance" [89]. Resistance exercise is defined as “is exercise that causes muscles to work or hold against an applied force or weight” [90]. Meditative-movement exercise is defined as a combination of some form of movement or body positioning, breathing, and relaxation [91].

Point 18: Line 159 - What was your rationale for four or more intervention modules? Why not 3 or 5 or 10, etc.? *

Reply 18: We discussed this issue at length before the initial submission. As we could not find data which would have informed this decision, the choice was based on expert opinion.

Point 19: Line 160 - Along the same as above, why greater than 2 hours per day? Why not 1 or 3 or 4, etc.? Personally, I don't see a need for this here. *

Point 20: We included this aspect on theoretical grounds. We agree that it is not a central element of the screening criteria. Upon further thought we came to the conclusion that such intervention durations are also unlikely in patients with depression. Therefore, we have deleted this criterion.


Reply 21: Thank you for this interesting input. As you say, there needs to be a balance between these issues. We have included two data bases for grey literature (proquest.com and opengrey.eu). In our scoping search we found approximately 250 hits. Hence, the additional screening effort is minimal. Considering the limited number of trials in this topic, we do not want to a priori exclude additional data from PhD or MSc theses which we might find in these data bases or when asking authors of included studies (see last sentence of the section “information sources”). Furthermore, methodological quality is considered when deciding whether the network meta-analysis is valid. Lastly, if the number of studies allows it, theoretically driven subgroup analyses will include methodological quality.
Point 22: Line 177 - Replace "afore mentioned" with "aforementioned" *

Reply 22: We have made this change in all the cases of the manuscript (lines 177 and 217).

Point 23: Line 181 - As previously mentioned, since MEDLINE is nested in PubMed, why not just search PubMed since it will provide you with access to additional biomedical content? *

Reply 23: We agree and have changed the manuscript accordingly.

Point 24: Line 193 - Please tell the reader who will be responsible for conducting the searches. *

Reply 24: We have included this information in the manuscript at a later stage in order to reduce redundancy:

Both stages will be performed independently by two reviewers (GB and TZS) who will not be blinded to any information (author, journal, institutions, etc.).

Point 25: Lines 211 through 214 - Regardless of the method used to remove duplicates electronically, I would strongly suggest that you also look for duplicates manually. Why? Because electronic approaches almost always miss one or more duplicate citations. *

Reply 25: The additional work of manual deduplication at the beginning does not seem to justify the benefit of removing the few duplicates which might remain. According to our experience, manual deduplication is, at the very latest, concluded during the full text screening.

Point 26: Line 216 - Replace "are" with "will be". On the same line, insert "the" before the word "title". *

Reply 26: We have made these changes in the manuscript.

Any article which might be relevant, but could not be included due to the aforementioned language constraints will be listed in an appendix.

Point 28: Line 217 - Replace "are" with "will be". On the same line, and as previously mentioned replace "afore mentioned" with "aforementioned" *

Reply 28: We have made these changes in the manuscript.
Upon deduplication, records will be screened in two stages. Firstly, the title and the abstract of all records will be screened against the aforementioned inclusion and exclusion criteria.

Point 29: Line 222 - Suggest that you include the following reference to support not being blinded: Berlin JA. Does blinding of readers affect the results of meta-analyses? Lancet. 1997;350:185-186. On the same line, please provide the names of the two people who will screen and select the studies. *

Reply 29: Thank you for this interesting reference. We have included it in the manuscript.

Both stages will be performed independently by two reviewers (GB and TZS) who will not be blinded to any information (e.g., author, journal, institutions). We do not blind the reviewers, since there is empirical evidence that blinding has little to no effect in meta-analyses [92].

Point 30: Line 224 - Please provide the name of the third reviewer. *

Reply 30: We have made this change in the manuscript.

Disagreement will be resolved by consensus. If no consensus can be reached, disagreement will be resolved by adjudication of a designated third reviewer (AST).

Point 31: Line 232 - Insert "the" before "number" *

Reply 31: We have made this change in the manuscript.

A flow diagram according to the PRISMA guidelines [79] will illustrate the number and the reasons for excluded and included citations.

Point 32: Line 238 - Please provide the names of the two reviewers who will extract data from the studies. *

Reply 32: We have changed the manuscript accordingly.

Both reviewers (GB and TZS) will extract data independently.
Point 33: Lines 239 through 242 - Please rewrite the sentence that begins "In case of..." as it is not written correctly. *

Reply 33: We have changed the manuscript accordingly.

Authors will be contacted should data be missing. (The corresponding author will initially be contacted via e-mail with one additional reminder e-mail, should there be no response within two weeks. Subsequently the other authors will be contacted).

Point 34: Line 243 - Please provide the name of the third reviewer. *

Reply 34: We have changed the manuscript accordingly.

Disagreement will be resolved by consensus upon consulting the original paper or if no consensus can be reached, disagreement will be resolved by adjudication of a designated third reviewer (AST).

Point 35: Lines 243 and 245 - What will be your decision rule about which one to include? The most recent one, the one with the most information, or something else? *

Reply 35: We will use the publication which has the most information pertinent to the meta-analysis.

We will include the publication which has the most information pertinent for the meta-analysis.

Point 36: Line 248 - Suggest that you replace "points" with "information". *

Reply 36: We have changed the manuscript accordingly.

Amongst others, the following information will be extracted from each study:

Point 37: Line 289 - Please provide the names of the two reviewers who will be responsible for assessing the risk of bias. Also, please replace "on" with "at". *

Reply 37: The risk of bias will be evaluated independently by two reviewers (GB and TZS) at the study level.
Point 38: Line 291 - Please provide the name of this third reviewer. *

Reply 38: We have changed the manuscript accordingly.

Disagreement will be resolved by consensus or if no consensus can be reached, disagreement will be resolved by adjudication of a designated third reviewer (AST).

Point 39: Lines 291 through 294 - Will you use this version or the more recent and revised one that was completed in October of 2018? *

Reply 39: Thank you for this comment. We will use the most recent version and have changed the citation accordingly.

Point 40: Lines 307 and 308 - If a network meta-analysis is not possible, what about a traditional pairwise meta-analysis as a possible next step followed by a narrative review if a pairwise meta-analysis is not possible. Given your topic, I don't believe that you will find enough studies to conduct any type of meta-analysis. One way to address this is to conduct a preliminary search. While my suggestion does not necessarily fit the Cochrane ideal, I deal with reality versus ideality. *

Reply 40: Thank you for this important comment. We have discussed this issue and have decided to broaden the inclusion criteria. We will now include trials which used aerobic, resistance or mind-body exercises (qi gong, tai chi, yoga) as interventions. Our initial scoping search has identified eight, possibly nine trials which we can include in our meta-analysis. Also, we have already drawn the network and are confident that a network meta-analysis is possible. The major advantage of a network meta-analysis is that studies using different comparators can be included in one meta-analytic model. This is in contrast to the standard pairwise meta-analysis, where individual meta-analyses have to be conducted per comparator. For instance, if treatment A was compared with Comparator B and Comparator C in two trials each, the network meta-analysis will include 4 trials while each pairwise meta-analysis would include only 2 trials. Therefore, the network meta-analysis is the more parsimonious approach, and especially in this case, where we expect to find multiple comparators, with a small number of studies per comparator, the approach of network meta-analysis constitutes the most appropriate synthesis approach. Furthermore, the initial scoping search was performed a year ago. It is therefore possible that even more trials on this topic are now available. Nevertheless, we concur with your stepwise approach and have adapted the manuscript as follows:

If network meta-analysis results must be deemed methodologically inaccurate, a pairwise meta-analysis will be considered. Should a pairwise meta-analysis also not be possible, studies will be summarized narratively.
Point 41: Lines 314 through 317 - A need exists for more specific information here. For example, would you examine for transitivity of potential effect modifiers across the different pairwise comparisons for each outcome using chi-squared tests for categorical variables and one-way ANOVA tests for continuous variables, with appropriate follow-up tests when necessary? *

Reply 41: To the best of our knowledge, there is no published statistical test for the transitivity assumption available at the moment. Thus, evaluating the transitivity assumption requires epidemiological judgements (Linde et al J Clin Epidemiology 2016, 71, 86-96; Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. Res Synth Methods 2012;3:80e97.)

We have added a sentence concerning effect modifiers:

A priori we have defined depression severity, comorbidities, age, and gender as potential effect modifiers and will evaluate the comparability of the respective characteristics across comparisons qualitatively.

Point 42: Lines 318 and 319 - Please provide a reference for how you will calculate the standardized mean difference. For example, will you use Cohen's effect size, Hedge's effect size, or something else? Also, will you use any type of small-sample adjustment such as Hedge's approach? *

Reply 42: We will use Hedge’s g with adjustment for small sample size and have amended the manuscript accordingly:

We expect considerable diversity of outcome measures and will, therefore, calculate standardized mean differences (SMD) using Hedge’s g with 95% confidence intervals [99].

Point 43: Lines 321 and 322 - It's important to note that Cohen's cutpoints for standardized mean difference effect sizes are somewhat arbitrary. Also, alternative cutpoints have been proposed (see for example: Sawilowsky SS. New effect size rules of thumb. J Mod Appl Stat Methods. 2009;8(2):597-599). Also, please delete "random-effects before the word "pairwise ..." Why? Because random-effects has to do with pooling, not the calculation of each individual effect size.

*  

Reply 43: Thank you for this interesting reference. We have adjusted the protocol as follows:
The conventional and somewhat arbitrary classification of SMD proposed by Cohen (1988) [97] has been expanded to include very small (.01), small (0.2), medium (0.5), large (0.8), very large (1.2), and huge (2.0) effect sizes [98].

We meant the pooled effect size between treatments and comparators across studies here. We have specified the sentence as follows:

Random-effects pairwise SMDs across studies will be calculated based on the available comparisons between treatment and comparator treatments [99].

Point 44: Line 325 - I'm assuming here that you mean random-effects models will be used to pool findings. Assuming the former, please provide the reference for the random-effects model you will use given that several exist, all of which calculate between-study variances differently.

Reply 44: We have added two citations:


Random-effects pairwise SMDs across studies will be calculated based on the available comparisons between treatment and comparator treatments [99]. Inverse variance weighting is used for pooling. In addition, indirect evidence will be estimated using the entire network of evidence. Random-effects netmeta accounts for dependencies between comparisons in case of multi-arm trials [100].

Point 45: Line 330 and 331 - What will be your decision rule for deciding on which scale is the “most common” one? In other words, what is you a priori list? Another way to approach this is to pool multiple measures of the same outcome from the same study into one overall effect size. After doing that you could also treat them separately in order to examine any potential association(s) between the instruments used and changes in sleep quality. The same for your other outcomes.

Reply 45: We define the “most common scale” as being the one which is used the highest number of times across the included studies. The goal is to reduce between-study variability. We therefore do not have an a priori list.
Analyzing the association between instruments and changes in sleep quality is an interesting idea. We will include this analysis if the number of studies allows it.

We have adapted the manuscript as follows:

If more than one primary outcome is reported, the most frequently used scale will be included in the analysis to reduce between-study heterogeneity. If possible, we will assess the association between instruments and changes in sleep quality.

Point 46: Lines 335 and 336 - What about also including 95% prediction intervals to see what effects might be expected in a new trial? *

Reply 46: This is a good point and we will include the 95% prediction interval. We have changed the manuscript accordingly:

Whenever possible, measures of uncertainty will be reported in the form of the 95% confidence interval and 95% prediction interval.

Point 47: Lines 339 through 343 - It's important to understand that the Q statistic measures statistical heterogeneity while I-squared, an extension of Q, assesses inconsistency (see: Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. Br Med J. 2003;327(7414):557-560). *

Reply 47: We have changed the manuscript accordingly:

To calculate statistical heterogeneity between studies on the pairwise level the Q statistic will be used [73]. … I2 and the corresponding confidence interval can be interpreted as the percentage of overall heterogeneity that is due to variation of the true effects. An I2-value of 0% to 40% might not be important, 30% to 60% may represent moderate heterogeneity, 50% to 90% may represent substantial heterogeneity, and 75% to 100% considerable heterogeneity [73].

Point 48: Lines 346 and 347 - You will split the direct and indirect evidence and then compare them for any statistically significant differences, correct? If so, then please state this. *

Reply 48: We have added information in the manuscript:

The presence of inconsistency will be evaluated using the following approaches: (1) locally using the netsplit command (i.e., testing the difference between estimates derived from direct evidence and estimates derived from indirect estimates for statistical significance) and (2) globally using the decomp.design command (i.e., using the design-by-treatment interaction
model). For this purpose, the total Q statistic (i.e., the measure of total heterogeneity / inconsistency in the network) will be decomposed to an inconsistency factor (between designs) and a heterogeneity factor (within designs).

Point 49: Lines 348 and 350 - So, how will you do this? A global test of inconsistency using a chi-squared test and an alpha value <0.05, or something else? *

Reply 49: We will compare the magnitude of heterogeneity between consistency and inconsistency models to determine how much heterogeneity will be explained by inconsistency. We will do this by testing the residual inconsistency, which remains under the assumption of a full design by treatment interaction model for statistical significance.

Point 50: Line 353 - What do you mean by "study precision"? *

Reply 50: By study precision we mean the 95% confidence interval of the effect estimate which is directly linked to the sample size (see e.g. DOI: 10.1093/ije/dyt074). We have changed the manuscript as follows:

In case of statistical heterogeneity or inconsistency between results from individual studies, we will investigate the potential impact of the following trial-level effect modifiers: (1) year of publication, (2) study precision (i.e., sample size), …

Point 51: Line 354 - This is the first time that I've seen "imputed standard deviations". This should probably be explained back where you talk about calculating your standardized mean difference effect size. I'm assuming here that you're talking about imputing change outcome standard deviations according to an approach such as Follman et al. (see: Follmann D, Elliot P, Suh I, Cutler J. Variance imputation for overviews of clinical trials with continuous response. J Clin Epidemiol. 1992;45:769-773). *

Reply 51: We have included a paragraph on this topic in the section “Data Items”:

For the calculation of relative treatment effects group means, corresponding standard deviations and group sizes will be extracted primarily. In case one of these values was missing, other statistical data that can be converted into means and standard deviations will be extracted. Conversions will be calculated according to formulas provided e.g. [82, 83]. If standard deviations cannot be calculated from the available study information, we will impute them using the standard deviations reported in the other included studies [84]. We will conduct sensitivity analyses excluding studies in which standard deviations had to be imputed. If the N was missing in the table of analysis, we will use the N of the descriptive statistics. If studies report medians
and interquartile ranges, a normal distribution will be assumed, if not indicated otherwise, to convert these values to means and standard deviations [82]. If studies only report adjusted outcome values, data will be extracted, but sensitivity analyses will be calculated without these studies to check for possible bias. We plan to extract the effect size provided by the study authors only if no other information was available for effect size calculation. If it is not possible to impute appropriate measures for the calculation of effect sizes, and no effect sizes are reported we will contact the authors.

Point 52: Lines 351 through 358 - Regardless of whether statistical heterogeneity or inconsistency between results from individual studies is found, an examination of potential effect-modifiers using something like meta-regression will still be important. In addition to the factors listed, another covariate that is important to examine is the association between changes in sleep and changes in depressive symptoms as a result of the intervention. Why? Because to the best of my knowledge, we don't know for sure if an intervention improvement in something like sleep quality is the direct result of aerobic exercise or aerobic exercise reducing depressive symptoms, thereby resulting in improved sleep quality. *

Reply 52: This is a very interesting point. Unfortunately, we consider this analysis not to be feasible. We anticipate an insufficient amount of studies to run a meta-regression (confer also https://www.ncbi.nlm.nih.gov/pubmed/15160401). Furthermore, we consider this question to be out of the scope of our review.

Point 53: Line 363 - You mention publication bias but there is no description back in your methods about how you will assess publication bias, more appropriately termed, small-study effects. *

Reply 53: We use the term “publication bias” in accordance with the CINeMA framework (https://doi.org/10.1371/journal.pone.0099682) to refer to non-publication and have added the following sentence:

Publication bias will be assessed according to the GRADE guideline [102] and by comparing eligible trials identified in registries (e.g. clinicaltrials.gov) with published data.

Point 54: Lines 375 - Suggest that you insert "best of our" before the word "knowledge". *

Reply 54: We have included this phrase and deleted the phrase “of the authors”:

To the best of our knowledge, this will be the first systematic review concerning this topic.
Point 55: Line 384 - I don't see the avoidance of Google Scholar for grey literature as a limitation at all. Given the large number of false-positives currently generated when searching Google Scholar, most librarians would agree that searching Google Scholar is a waste of time. *

Reply 55: We agree with this statement and have deleted this sentence.

Point 56: Lines 390 - Since you're limiting your study to aerobic exercise, please insert the word "aerobic" before the word "exercise".

Reply 56: We have specified the types of exercises accordingly.

Considering the importance of sleep disturbances in depression, we hope that this systematic review can accelerate the consolidation of evidence, such that decision makers (patients, health-care professionals, and policy makers) are provided with high quality evidence to facilitate decisions on whether and how to implement aerobic, resistance or mind-body exercises as a treatment module for patients with depression.

END OF REVIEW