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)	Does Improving Sleep Lead to Better Mental Health? A Protocol for a Meta-Analytic Review of Randomised Controlled Trials
AUTHORS	Scott, Alex; Webb, Thomas; Rowse, Georgina

VERSION 1 - REVIEW

REVIEWER	John Cape University College London, UK
REVIEW RETURNED	28-Mar-2017

GENERAL COMMENTS	The paper is a proposal for a meta-analytic review of RCTs of interventions for sleep disorders which included estimates of mental health outcomes. The authors make a case that a more comprehensive review of studies than has been undertaken to date will improve our understanding of the circumstances where improving sleep leads to better mental health outcomes and also argue that this could provide evidence of a causal relationship between sleep problems and poor mental health. The authors' argument that intervention studies can demonstrate a causal relationship between sleep and mental health outcomes is incorrect. In exactly the same way that they point out that longitudinal studies of the relationship between poor sleep at time 1 and mental health problems at time 2 do not prove causality, a third unmeasured variable in intervention studies (e.g. therapeutic relationship in psychological interventions for insomnia) rather than improvement in sleep could be responsible for improved mental health outcomes. Mediational analyses can help to elucidate causality, but the authors do not indicate they are planning to identify and extract these from studies. My major concern with the review relates to the very comprehensive nature of its scope that the authors see as a strength. All sleep disorders are included in the scope (this is clear from the search terms, but is not explicit in the inclusion criteria) and these are a heterogenous range of disorders each with potentially very different relationships with mental health (e.g. nightmares and post traumatic stress disorder) and each with often disorder specific different interventions. The interventions included in the scope are also heterogenous – different medications, different psychological interventions and different physical interventions (e.g. CPAP for sleep apnoea). The authors indicate that they will deal with this in moderator analyses and give a list in Table 2 of 26 variables to be extracted for moderator analyses. But some of these variables are themselves heterogenous ("theoretical basis" of intervention appears to cover the full range of different medications, psychological and physical interventions) and how they plan to proceed with these

	analyses is unclear. Is the plan to explore the effect of each moderator across all studies (e.g. duration of intervention across all studies), which to this reviewer makes little sense clinically, given the different disorders and interventions? Or is the plan to explore moderators within subsets of studies (e.g. duration of CBT in insomnia disorder) which makes sense clinically, but with the number of disorder X intervention cells would require a very large number of separate moderator analyses? Either way, we need clarity as to all the pre-planned analyses for this to be an informative protocol. Otherwise, with the very large number of potential analyses that could be undertaken, the protocol will not serve its function of enabling other researchers to check whether the analyses eventually reported on in the study findings are those that were pre-planned rather than opportunistic and how many analyses were undertaken with potential for chance findings from multiple analyses. Limiting the scope just to interventions targeting insomnia / insomnia disorder and excluding interventions targeted at specific sleep disorders, would simplify these issues considerably, without in the opinion of this reviewer any significant loss, and this is one option the authors might wish to consider. Missing from the methodology is how sleep outcomes will be analysed. The authors indicate that only studies which show significant effects on sleep outcomes will be included in the meta-analyses. Some specific sleep disorders have rather specific outcomes (e.g. nightmares, apnoea attacks, phase shift, sleep attacks and excessive daytime sleepiness). Will these outcomes be used as well as measures of sleep efficiency (and its components – SOL and WASO), total sleep time and composite questionnaire measures (e.g. ISI, SCI, PSQI)? And given common reporting of multiple sleep outcomes in intervention studies, how will this be dealt with? And is it statistical significance or clinical significance of the impact of the intervention on sleep outcomes that will be the inclusion criterion. A couple of other aspects that need to be covered in the methods section are: 1. Whether for each study the post-treatment means and SDs used to calculate Hedges g for each sleep, mental health and wellbeing variable will be adjusted for baseline differences between groups if reported and whether, where both are reported or can be calculated from study data, just the means and SDs from participants who completed the study will be used or the means and SDs for all participants from an intention-to-treat analysis (in which case is a particular type of ITT analysis favoured) or some alternative? 2. Whether dichotomous outcomes where reported in studies will be used and, if so, whether just when these are the only outcome measures reported in a study or also when both dichotomous as well as continuous measures are reported? And, if so, how these will be analysed. A couple of minor points: 1. The in other respects good introduction to the paper could be read as suggesting that the traditional view that mental health problems cause sleep disturbance has been replaced by a view that the causal relationship is entirely in the other direction, rather than by a view that it is a bidirectional relationship 2. The reference list needs checking – there are duplicates and missing information.

	Finally, the paper ends rather abruptly without any discussion that might even if briefly reflect on the proposal and its strengths and weaknesses, which readers might find helpful.

REVIEWER	Harriet Hiscock Murdoch Children's Research Institute, Australia
REVIEW RETURNED	10-Apr-2017

GENERAL COMMENTS	This is a worthy topic of a meta-analysis. However, I have a number of concerns which need to be addressed: 1. What age group are you including? It looks like adults only but there is a growing literature on RCTs of sleep interventions in children and adolescents with mental health outcomes as well. These studies should be included. 2. Objectives: what is your rationale for including sleep quality only as an outcome? Why not sleep duration or sleep scheduling - both of which have been shown to affect mental health and behaviour in children and youth. 3. Inclusion criteria: - why have you not included ADHD? It is commonly associated with sleep problems and RCTs of sleep interventions in children with ADHD have shown benefits to their mental health too. 4. Will you include non-English articles? What years will your search cover? 5. I would suggest also searching large international trial registries (not just UK National Research Register) eg ISRCTN Registry. 6. Data extraction - will the second code of the subset of articles (pg 10) be a random subset? 7. Proposed analysis - you now include "wellbeing" in your inclusion criteria whereas earlier in the manuscript it was just mental health. What do you mean by "wellbeing"? How will you define this? For example, will you include QoL? 8. Please state your principal summary measure clearly (as per PRISMA guidelines). Is it means and SD or effect sizes? 9. Table 1 - consider including other common mental health disorders such as ADHD which is also common in adults. Will you include developmental disorders such as autism spectrum disorder?

VERSION 1 – AUTHOR RESPONSE

Reviewer 2.

Response to Reviewers

Editorial Requests

1. The Editor requested that we provide more detail in the abstract, particularly in the methods and analysis section.

In response, we took your advice and looked at the abstracts of other protocols for meta-analyses that have recently been published in BMJ Open(1 2). We then revised our abstract to include more information about the proposed review. For example, we now state that we will search multiple bibliographic databases in addition to a search of the grey literature. Furthermore, we are explicit that we will compute Hedges effect size g and associated confidence intervals based on means and standard deviations alongside a moderation and mediation analyses. Finally, we have outlined our approach to assessing study quality and the risk of publication bias and included the Prospero review registration number.

2. The Editor (and Reviewer 2) requested that we provide the dates of coverage for each database searched.

We apologise for omitting this information from our original submission. The revised manuscript now states the dates of coverage for each database that will be searched (see p. 1, line 240-241).

Reviewer 1

1. Reviewer 1 felt that our assertion that “intervention studies can demonstrate a causal relationship between sleep and mental health outcomes” may be incorrect, as the same problems inferring causality can be encountered with intervention studies as in longitudinal research (e.g., that a third, unmeasured variable accounts for the effect of the intervention on outcomes). Reviewer 1 suggested that mediation analysis might help to elucidate causality, but was concerned that we did not indicate whether or not we were planning to identify and extract this information from the primary studies and conduct mediation analyses.

Reviewer 1 is correct that intervention research is also susceptible to the third variable problem. That is, although (well-designed) experimental studies evaluating the effects of an intervention (relative to a comparison condition) on outcomes are able to attribute any change in outcomes to the intervention, this does not necessarily mean that the proposed target of the intervention (e.g., improvements in sleep quality) is responsible for the change in outcomes. Instead, a third – unmeasured – variable (e.g., a therapeutic relationship) could be responsible for the change in outcomes. Therefore, although intervention studies can demonstrate a causal relationship between sleep and mental health outcomes, such a conclusion is only valid if the studies measure changes in sleep quality in the wake of the intervention and demonstrate that these changes account for the effects of the intervention on outcomes (e.g., changes in mental health).

Given the importance of this issue, one of the inclusion criteria for studies to be included in the proposed review is that they report a statistically significant improvement on a measure of sleep quality among participants in the experimental group as compared to those in the comparison group (see p. 10, line 214 of the revised manuscript). This will ensure that we only examine the effect of interventions that improve sleep as intended. In addition, the information on the effect of the interventions on sleep quality, will enable us to follow Reviewer 1’s suggestion and conduct a mediation analysis (see p. 16, line 350 of the revised manuscript) to investigate whether changes in sleep quality mediate any effects of the intervention on (changes in) mental health outcomes (thereby, providing evidence for a causal relationship between sleep and mental health outcomes). Indeed, the review team has prior experience using a similar mediation method in a meta-analysis investigating the impact of (changes in) intention on changes in behaviour(3).

2. Reviewer 1’s primary concern with the proposed review was the comprehensive nature of its scope, a concern which was predicated on the inclusion of studies focusing on (i) all sleep disorders and (ii) a wide range of (potentially quite different) interventions. Reviewer 1 was concerned that the broad scope of the proposed review could lead to a heterogeneous group of studies and effect sizes. Reviewer 1 correctly noted that we planned to deal with this heterogeneity using moderator analyses (that compare effect sizes between studies focusing on, for example, different samples and between different interventions), but stated that it was not clear how we planned to proceed with these analyses.

The nature and severity of problems sleeping and the nature of the interventions offered to

participants to improve sleep are not the focal point of the proposed review (our primary aim is simply to examine whether changes in sleep quality are associated with changes in mental health outcomes). Therefore, we do not plan to restrict inclusion in the proposed review to studies focusing on specific sleep disorders or to exclude particular types of intervention from the review. Indeed, by including a broad range of sleep related search terms our aim is to identify as many interventions targeting sleep problems as possible. Many other reviews have investigated the impact of a range of interventions on a variety of sleep problems (4-10) and these reviews are better able to conclude as to what types of intervention work best for specific sleep problems.
Instead, the proposed review aims to tackle a slightly different question about the effect of improving sleep (generally) on mental health and quality of life. For example, we want to identify the size of the effect on mental health when sleep improves, and on what outcomes reflecting to mental health and quality of life improving sleep has the largest and most consistent effects (e.g., can improving sleep confer a benefit for psychosis, or is it better suited to affective experiences?). Consequently, we aim to include as many studies as possible that report a (statistically significant) change in sleep quality as a function of an intervention and examine the effects of this change on outcomes related to mental health and quality of life. Restricting the review to primary studies focusing on, for example, those with clinically defined sleep problems or receiving a particular intervention would limit the evidence base on which we could draw conclusions about this central issue, as well as limit the extent to which our findings can be generalised.

Having said this, we are interested in knowing if a particular types of sleep intervention (e.g., those based on the principles of CBT, self-help interventions vs. those delivered face-to-face, medication etc.) have a greater impact on outcomes than others, and if the effects of such interventions differ as a function of the targeted sample. As such, we will conduct moderation analyses that compare the (sample-weighted average) effect sizes between studies focusing on, for example, different samples and between different interventions.

We apologise if our approach to these moderation analyses was not clear in the first submission. In the revised manuscript, we have added more detail to the description of the planned moderation analysis (see p. 15, line 335).

3. Reviewer 1 requested further clarity on how sleep outcomes would be analysed and, specifically, which measure(s) of sleep we consider acceptable for; i) deciding which interventions have a 'significant impact on sleep'; and ii) computing effect size estimates.

We would like to thank Reviewer 1 for raising this issue. They are correct in saying that many outcome measures are specific to a given sleep problem (e.g., nightmares, sleep apnoea) and, thus, may not be comparable (c.f., the problem of comparing apples to oranges (11)). Therefore, we do not propose to include measures of sleep that are specific to particular disorders (e.g., occurrence of sleep apnoea), or that reflect specific symptoms (e.g., frequency of nightmares per night).

Instead, to be eligible for inclusion the primary studies need to report a measure of sleep quality that is likely to be comparable across studies. In line with both Harvey et al. (12) and Libman et al. (13) we propose to use measures that reflect; (i) sleep continuity (e.g., measures of sleep onset, sleep maintenance, number of awakenings); and (ii) daytime impact (e.g., feeling refreshed on waking and throughout the day) to assess sleep quality. We feel that this offers a balance between the primary aim of the review (i.e., to understand the impact of improving sleep on mental health and quality of life) and the need to ensure that the metrics that we use are comparable between and across studies. A more detailed discussion of this issue can be found on p. 8, line 148 of the revised manuscript.

4. Additionally, Reviewer 1 requested that we explain how we propose to deal with studies that report multiple sleep outcomes.

Thank you for suggesting that we be more explicit about how we will deal studies with multiple sleep outcomes (and apologies for not doing so originally). As described above, we will take care to ensure that all of the measures that we include in the review reflect sleep quality. Therefore, where studies report multiple measures of sleep quality (e.g., the Pittsburgh Sleep Quality Index (14) and the Sleep-50 Questionnaire (15)), the effect sizes will be computed for each outcome and meta-analyzed in their own right to form one overall effect for inclusion in the main analysis. This procedure capitalises on the available data while also ensuring the independence of effect sizes that is central to the validity of meta-analysis. We have added an example on p. 14, line 302 of the revised manuscript to illustrate this approach.

5. Reviewer 1 asked whether our inclusion criterion (namely, that only interventions that lead to an improvement in sleep quality be included in the review) reflected statistical or clinical significance of the impact of the intervention on sleep outcomes.

For a study to be eligible for inclusion in the proposed review, we require that the intervention resulted in a statistically significant improvement in sleep (as this is likely both easier to judge and more likely to be reported than whether the respective improvement was of clinical significance). We apologise if this was unclear in the original submission; however, this information is now reported on p. 10, line 214 of the revised manuscript.

6. Reviewer 1 asked whether we plan to use statistics that have been adjusted for baseline differences between groups (if reported) in order to compute effect sizes.

Again, this is an important point and we apologise if we were not sufficiently clear about our proposed approach in the original submission. Where an article has reported data that has been adjusted for baseline differences between groups, then we will use this to compute an effect size estimate. However, if this information is not reported then we will use the unadjusted statistics to compute the effect size. We propose to add subscripts to the table reporting the effect sizes derived from the primary research studies in order to identify how each effect size was computed and also to compare outcomes between studies that report adjusted vs. unadjusted statistics. We have updated the protocol to include this information in a footnote on p. 13 and in Table 2 (see p. 29, line 654 of the revised manuscript).

7. Reviewer 1 also asked whether we plan to compute effect sizes based on participants who completed the study, or based on intention to treat (ITT) analyses.

Again, this is an important point and thank you for drawing our attention to the need to specify our approach in this regard. Where reported, we will compute effect sizes using the data from ITT analyses (and now state as much in a footnote on p. 13 of the revised manuscript). We have taken this decision because effect sizes based on ITT analyses are more likely to; i) retain the balance of prognostic factors arising from the original randomisation; ii) give an unbiased estimate of treatment effects; and iii) better reflect real world applications of sleep interventions for mental health and quality of life (16 17). However, as with our approach to adjusted vs. unadjusted effect sizes, we will include the type of analysis (i.e., ITT vs. non-ITT, see Table 2, p. 29 line 6655) as a potential moderator of the effect(s) of the interventions on outcomes in order to empirically examine its influence.

8. Reviewer 1 asked us to clarify whether dichotomous outcomes are likely to be reported in primary studies included in the present review and, if so, how we will deal with such outcome measures.

All studies that report outcome measures that allow the computation of Hedges g will be eligible for inclusion in the review, including dichotomous outcomes. For example, Chi-square statistics resulting from the analysis of dichotomous outcomes (e.g., a measure of whether participants report that they slept well or badly on the preceding night) can be converted to an effect size r and subsequently to Hedges g (using an effect size calculator such as Lyon-Morris (18), see p. 13, line 284 of the revised manuscript).

9. Reviewer 1 suggested that the introduction section of our original manuscript could be read to suggest that the traditional view that mental health problems cause sleep disturbance has been replaced by a view that the causal relationship is in the other direction (rather than by the view that it is a bidirectional relationship).

We agree that the relationship between sleep and outcomes pertaining to mental health and well-being is likely to be bidirectional and apologise if our original discussion sounded as if we are only open to the possibility that sleep influences outcomes pertaining to mental health and well-being. Although for a number of theoretical and practical reasons this is the direction of the relationship that our proposed review examines, we have revised the introduction to be explicit that the relationship between sleep and mental health is likely to be bidirectional and that finding that sleep influences outcomes pertaining to mental health and well-being does not preclude the possibility that the latter also influences the former (see p. 4, line 61 of the revised manuscript).

10. Reviewer 1 felt that the paper ended rather abruptly, and that readers may find a discussion that reflects on the strengths and weaknesses of the proposal would be helpful.

We would like to thank Reviewer 1 for this suggestion. In response, the revised manuscript now includes a brief discussion (p. 17, line 368) that summarises the aims of the proposed review and identifies strengths (i.e., the use of the GRADE system and proposed mediation analysis) and weaknesses (i.e., a potentially heterogeneous group of studies) of the proposal.

11. Finally, Reviewer 1 noticed a few discrepancies with the reference list.

We apologise for these oversights and have checked that all of our references are reported in line with BMJ Open policies and have ensured all in-text citations are listed in the reference section.

Reviewer 2

1. Reviewer 2 asked us to clarify the age group(s) that the proposed review proposes to include and, specifically, suggested that we should include RCTs of sleep interventions in children and adolescents.

We agree that studies focusing on children and adolescents may contribute to our understanding of the effect of improving sleep on mental health and well-being. To clarify, our search terms do not preclude the inclusion of studies focusing on children and adolescents and we will include such RCTs focusing on such samples providing that they meet the proposed eligibility criteria.

We apologise if this was not clear in the original submission and have clarified this issue on p. 11, line 230 of the revised manuscript (i.e. “for example, we will not restrict the type of intervention (e.g., psychological and pharmacological), publication status, nature of the comparison condition, or sample (i.e., interventions directed toward adults, children, and adolescents will all be eligible)”. The nature of the sample will, however, be examined as a potential moderator of the impact of the interventions in order to empirically examining whether improving sleep has different outcomes for people of, for example, different ages.

2. Reviewer 2 requested that we explain why we only propose to include sleep quality as an outcome,

and not outcomes such as sleep duration and sleep scheduling.

This an important point and we have sought to provide further clarity on the nature of the outcome measures that we deem relevant throughout the revised submission (in part, in response to Reviewer 1's comments as well – see above). To reiterate, the main aim of the proposed review is to quantify the effect of improving sleep on outcomes reflecting mental health and quality of life. However, we recognise that the notion of 'improving sleep' is broad and multifaceted; and, as Reviewer 1 stated, many specific sleep difficulties are tied to mental health in unique ways, with their own specific measures (e.g., frequency of nightmares). Consequently, one challenge for the proposed review is to ensure that all of the primary studies assess a similar notion of sleep improvement.

In order to achieve this, we will require that the primary studies report a measure of sleep quality that is likely to be comparable across studies. In line with both Harvey et al. (12) and Libman et al. (13), we will use measures pertaining to; (i) sleep continuity (e.g., measures of sleep onset, sleep maintenance, number of awakenings); and (ii) daytime impact (e.g., feeling refreshed on waking and throughout the day) to assess sleep quality. We feel that this offers a balance between the primary aim of the proposed review (i.e., to understand the impact of improving sleep on mental health and quality of life) and the need to ensure that the metrics that we use are comparable between and across studies. A more detailed discussion of this can be found on p. 8, line 148 of the revised manuscript.

3. Reviewer 2 asked why we did not include ADHD in our search terms and whether we should also seek to include studies focusing on improving sleep among those with other developmental disorders such as autism spectrum disorder.

We agree that ADHD is commonly associated with sleep problems and that RCTs of sleep interventions in children with ADHD have shown benefits to their mental health. Additionally, there is evidence to suggest improving sleep in those with autistic spectrum disorders can also confer benefits to mental health. As such, we have revised our search terms to include 'attention deficit', 'hyperactivity disorder', 'ADHD,' 'autis*' and 'asperger*'. Developmental disorders will be included as a category of mental health problems in our moderator analyses that seek to determine whether the impact of improving sleep on outcomes differs between non-clinical and clinical samples; and between clinical issues.

4. Reviewer 2 asked us to clarify whether we plan to include or exclude studies that are not published in English.

We apologise if our approach to this issue was not clear in the original submission. Our plan is to include articles that are not published in English provided that the relevant details can be extracted via the use of available translation resources (e.g., with the help of colleagues at the University of Sheffield, Google Translate, and friends). We have added this as an inclusion criteria on p. 11, line 223 of the revised manuscript.

5. Reviewer 2 suggested that we search large international trial registries such as the International Standard Randomised Controlled Trial Number (ISRCTN) in order to find primary studies.

We believe that this is a sensible idea (thank you) and, indeed, already plan to search ISRCTN. For example, on p. 9 of the original submission we stated that we will search WHO approved clinical trial databases for additional articles. It may not have been clear, but ISRCTN is one of the databases included in this list. We have made this clearer on p. 12, line 251 of the revised manuscript.

6. Reviewer 2 asked whether the subset of articles to be second coded would be randomly selected.

We can confirm that the subset of articles to be second coded will be randomly selected using a computer-generated algorithm, and have made this explicit on p. 13, line 271 of the revised manuscript.

7. Reviewer 2 requested that we clarify what we mean by 'wellbeing' and, specifically, whether this would include measures of quality of life (QoL).

We agree that the concept of wellbeing was not clearly defined in the original submission. Indeed, in considering your request for further clarity on this matter we have taken the decision to focus on measures of QoL, rather than wellbeing. Wellbeing and QoL are often (incorrectly) used interchangeably(19, 20), a pitfall that the original submission fell into. The construct of QoL is closer to the intended aim of the present review, which is to examine the impact of interventions that improve sleep on not only mental health, but also on outcomes that go beyond mental health (i.e., functional, occupational, social indicators etc.). Consequently, we will include measures of QoL that assess at least one of the five dimensions of QoL outlined by Felce and Perry (21): (i) Physical wellbeing (e.g., health and fitness); (ii) material wellbeing (e.g., financial security, possessions); (iii) social wellbeing (e.g., breadth and depth of relationships); (iv) emotional wellbeing (e.g., affect or mood, fulfilment, self-esteem); and (v) development and activity level (e.g., the possession and use of skills, work, education) A more detailed discussion of this issue can be found on p. 9, line 174 of the revised manuscript.

8. Reviewer 2 asked us to clearly state our principle summary measure as per PRISMA guidelines.

We apologise for not doing so in the original manuscript. The principle summary statistic for the proposed review will be Hedges g (and 95% confidence intervals) based on post-treatment means and standard deviations (adjusted for baseline scores, where available). Where means and standard deviations are not available, we will compute effect sizes by converting relevant summary statistics (e.g., F values from ANOVA testing the impact of an intervention on outcomes) using Lyons Morris' meta-analysis calculator (18). This information can be found on p. 13 of the revised manuscript. We now comply with all applicable statements in the PRISMA-P checklist (see Supplementary Materials 2).

References

1. Kelley GA, Kelley KS, Callahan LF. Community-deliverable exercise and anxiety in adults with arthritis and other rheumatic diseases: A protocol for a systematic review and meta-analysis of randomised controlled trials. BMJ Open 2017;7(3):e014957-e57. doi: 10.1136/bmjopen-2016-014957
2. Hilbert A, Petroff D, Herpertz S, et al. Meta-analysis of the effectiveness of psychological and medical treatments for binge-eating disorder (metabed): Study protocol. BMJ Open 2017;7(3) doi: 10.1136/bmjopen-2016-013655
3. Webb TL, Sheeran P. Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. Psychological Bulletin 2006;132(2):249.
4. Morin CM, Culbert JP, Schwartz SM. Nonpharmacological interventions for insomnia. American Journal of Psychiatry 1994;151(8):1172.
5. Kanji S, Mera A, Hutton B, et al. Pharmacological interventions to improve sleep in hospitalised adults: A systematic review. BMJ Open 2016;6(7):e012108.
6. Hansen K, Höfling V, Kröner-Borowik T, et al. Efficacy of psychological interventions aiming to reduce chronic nightmares: A meta-analysis. Clinical Psychology Review 2013;33(1):146-55.
7. Aiello KD, Caughey WG, Nelluri B, et al. Effect of exercise training on sleep apnea: A systematic review and meta-analysis. Respiratory Medicine 2016;116:85-92.
8. Smith MT, Perlis ML, Park A, et al. Comparative meta-analysis of pharmacotherapy and behavior

therapy for persistent insomnia. American Journal of Psychiatry 2002;159(1):5-11.
9. Araghi MH, Chen Y-F, Jagielski A, et al. Effectiveness of lifestyle interventions on obstructive sleep apnea (osa): Systematic review and meta-analysis. Sleep 2013;36(10):1553-62. doi: 10.5665/sleep.3056
10. Blake MJ, Sheeber LB, Youssef GJ, et al. Systematic review and meta-analysis of adolescent cognitive-behavioral sleep interventions. Clinical Child and Family Psychology Review 2017 doi: 10.1007/s10567-017-0234-5
11. Barone JE. Comparing apples and oranges: A randomised prospective study. BMJ 2000;321(7276):1569-70.
12. Harvey AG, Stinson K, Whitaker KL, et al. The subjective meaning of sleep quality: A comparison of individuals with and without insomnia. Sleep 2008;31(3):383.
13. Libman E, Fichten C, Creti L, et al. Refreshing sleep and sleep continuity determine perceived sleep quality. Sleep Disorders 2016;2016
14. Buysse DJ, Reynolds CF, Monk TH, et al. The pittsburgh sleep quality index: A new instrument for psychiatric practice and research. Psychiatry Research 1989;28(2):193-213.
15. Spoormaker VI, Verbeek I, van den Bout J, et al. Initial validation of the sleep-50 questionnaire. Behavioral Sleep Medicine 2005;3(4):227-46.
16. Gupta SK. Intention-to-treat concept: A review. Perspectives in Clinical Research 2011;2(3):109.
17. Heritier SR, Gebski VJ, Keech AC. Inclusion of patients in clinical trial analysis: The intention-to-treat principle. Medical Journal of Australia 2003;179(8):438-40.
18. Lyons LC, Morris WA. Lyons-morris meta-analysis calculator 2017 [Available from: http://www.lyonsmorris.com/ma1/ accessed 15th May 2017.
19. Pinto S, Fumincelli L, Mazzo A, et al. Comfort, well-being and quality of life: Discussion of the differences and similarities among the concepts. Porto Biomedical Journal 2016
20. Dodge R, Daly AP, Huyton J, et al. The challenge of defining wellbeing. International Journal of Wellbeing 2012;2(3)
21. Felce D, Perry J. Quality of life: Its definition and measurement. Research in Developmental Disabilities 1995;16(1):51-74.

VERSION 2 – REVIEW

REVIEWER	John Cape University College London, UK
REVIEW RETURNED	28-May-2017

GENERAL COMMENTS	The authors' thorough responses and changes in response to review comments on their original submission have addressed most of my previous concerns. The lack of pre-specification as to the specific moderators they are planning to analyse remains a specific issue they have not addressed. Their table of 27 moderators as I understand it sets out the types of moderator they will explore and is not a specific or exhaustive list of the moderators they will analyse. For example, their response to the reviewer comment about scope and moderator analyses states "we are interested in knowing if particular types of sleep intervention (e.g., those based on the principles of CBT, self-help interventions vs. those delivered face-to-face, medication etc.) have a greater impact on outcomes than others, and if the effects of such interventions differ as a function of the targeted sample". This both adds specificity and numbers to the "intervention characteristics" moderators in their table and also suggests that moderators will be analysed in sample/disorder X intervention subsets. Given there could be hundreds of ways of specifying the 27 moderators and of moderator X sample/disorder subsets, greater

	clarity and specificity is needed as to all the pre-planned moderator analyses. Otherwise, as set out in my previous review comment, with the very large number of potential analyses that could be undertaken, the protocol will not serve its function of enabling other researchers to check whether the analyses eventually reported on in the study findings are those that were pre-planned rather than opportunistic and how many analyses were undertaken with potential for chance findings from multiple analyses. I note that the second reviewer recommended including children and adolescents, for whom the nature of sleep interventions and methods of sleep outcome measurement especially in younger children are different to adults, thus widening the scope and potential required moderator analyses still further. While I remain unconvinced that a wider scope is more useful in elucidating the nature of the relationship between improvements in sleep and improvement in mental health/QOL outcomes than focusing on a narrower range of sleep disorders, it is helpful that the authors have acknowledged this as a possibility ("apples and oranges") in their added Discussion section. The authors' addition of a mediation analysis across studies within the meta-analysis is interesting and not an approach I have come across. My original reviewer comment had been about possibly reporting on studies that had undertaken such a mediation analysis within the study (so within study mediation analyses not an across study mediation analysis). Not having come across this before, I am unable to comment fully on the appropriateness of statistical assumptions of this, but can see that in principle this should be possible and note that the authors state they have used this approach before in a published review. Presumably there will be an issue of whether individual studies will report the data or analyses required to calculate some of the necessary regression equations.

VERSION 2 – AUTHOR RESPONSE

Reviewer 1

1. Reviewer 1 requested more information about the potential moderators that we plan to code and evaluate in the proposed review; stating that the list that we provided in Table 2 was neither specific, nor exhaustive.

We would like to thank Reviewer 1 for the opportunity to provide further detail regarding our planned moderator analysis. As we do not know exactly what types of interventions and studies we are likely to find, we initially proposed to identify the various ways in which the primary studies might differ after conducting the literature search. This was to ensure that we did not pre-specify a coding scheme that later proved unsuitable or did not fully capture the breadth and depth of study characteristics. However, we recognise that the protocol needs "to serve its function of enabling other researchers to check whether the analyses eventually reported on in the study findings are those that were pre-planned rather than opportunistic and how many analyses were undertaken with potential for chance findings from multiple analyses".

Consequently, we have provided a more detailed description of our planned moderation analysis on p. 16 of the revised manuscript. Additionally, we have included a detailed data extraction template form (Supplementary Materials 3) which describes how each variable will be coded, including their specific categories. This list is specific; however, it is not exhaustive (and it is not intended to be), as providing an exhaustive list would prevent us from adding new variables and categories as the review develops to better reflect the literature. However, we will be clear in the published report that any analyses that were not pre-specified the present protocol should be considered exploratory.

2. Reviewer 1 was concerned that the inclusion of children and adolescent populations (as suggested by Reviewer 2) rendered the scope of the proposed review even wider.

On reflection, we agree with Reviewer 1 that broadening the review to include studies focusing on children and adolescents could compromise the validity of our findings (e.g., lead to the problem of ‘mixing apples and oranges’), as well as the extent to which they can be generalized. Although investigating the impact of interventions designed to improve sleep on mental health among children and adolescents is important, we believe that this is best achieved in a separate review, rather than combining the two populations in one review. Given the importance of such a review, we plan to make a list of all articles that are excluded because they focus on children and / or adolescents and make this available as supplementary material in the final report.

The desire for a focused review, has also led us to decide not to report the effects of improving sleep on quality of life as originally proposed (i.e., we will report the effect of improving sleep on mental health outcomes only). On reflection, we have decided that reporting two meta-analyses in one review, (i.e. the effects of improving sleep on mental health and QoL separately), would limit the space available to provide a full and considered interpretation of the findings with respect to the effect of improving sleep on mental health. Much like our position on including children and adolescents, we feel a separate review looking at the effect of improving sleep on quality of life outcomes would be able to offer a more nuanced and substantive contribution to the literature.

VERSION 3 – REVIEW

REVIEWER	John Cape University College London, UK
REVIEW RETURNED	24-Jul-2017

GENERAL COMMENTS	The authors’ further revisions and very clear and helpful comments on their marked amended version have sufficiently addressed my review comments on their first revision. The further revisions make clear that there are a very large number of potential moderators and moderator X sample analyses. Although in practice most potential cells will not have sufficient studies for analysis, there will still be issues of interpretation around any significant findings. However, with the greater clarity about this in the protocol, other researchers will be able to understand how the analyses eventually undertaken are a subset of a much larger number of potential analyses and as such the protocol will serve its purpose.