PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	Do insomnia and/or sleep disturbances predict the onset, relapse or worsening of depression in community and clinical samples of children and youth? Protocol for a systematic review and meta-analysis.
AUTHORS	marino, cecilia; Andrade, Brendan; Aitken, Madison; Bonato, Sarah; Haltigan, John; Wang, Wei; Szatmari, Peter

VERSION 1 – REVIEW

REVIEWER	Jared M. Saletin PhD Department of Psychiatry and Human Behavior Alpert Medical School of Brown University Providence, RI, USA
REVIEW RETURNED	02-Dec-2019

GENERAL COMMENTS	I appreciate the opportunity to review this proposed meta-analytic protocol to address a major issue: the overlap of sleep disturbance and depression like phenomena in children and adolescents. The authors propose to conduct a systematic meta-analysis of the effect of insomnia and sleep disturbance in predicting depression in people ages 5- 24. While this registered protocol sticks to strong meta-analysis guidelines, has a strong plan for data extraction and bias appraisal, is registered and aligned with PRISMA, it's well purposed intention is paired with some methodological components that give me pause: * The authors require a slew of summary statistics are required for studies to be included, does this limit the amount of evidence available? * Dissertations and "grey data" will be included — these have yet to be peer reviewed and could limit the quality of evidence provided. It is not justified why these are included. * The authors seek to examine a wide age-range, how will the authors harmonize the definition of "insomnia" (or depression, for that matter) across these groups? I would recommend at minimum a hierarchical approach grouping child, adolescent, and adult participants, and perhaps the meta-analytic estimates being combined at a higher level. This will require careful thought. * By having a wide capture of sleep differences (and to that end, mixing "insomnia" and "sleep disturbances") , it's unclear and unlikely that a single phenotype will emerge. In fact, a meta-analyses could be examined for almost every sub-type of sleep disturbance (or insomnia profile) mentioned. * The authors will accept variable reporting methods for both insomnia and depression. A focus on clinical report, or valid

	questionnaires only — rather than on self-report or caretaker report, would help to harmonize the quality evidence they may expect. * While excluding anxiety without depression is clear, no clear exclusion for mania is given. * Mixing dimensional and diagnostic frames to study depression may conflate results. * The authors focus on DSM for both sleep and depression, however I wonder if ICSD would be better for the insomnia exposure? * Both introduction and discussion mention puberty as a pivot point for rates of depression betting dissociated between males and females — yet no analytic strategy for stratifying or testing trajectories by sex is provided, this would be a meaningful addition.

REVIEWER	Eleanor McGlinchey Fairleigh Dickinson University, New Jersey, USA Columbia University Irving Medical Center, New York, USA
REVIEW RETURNED	21-Jan-2020

GENERAL COMMENTS	Overall: This is a well-written and timely protocol for a proposed systematic review and meta-analysis on the topic of sleep disturbance as a predictor of later depression in children and adolescents. Lovato & Gradisar last published a meta-analysis on this topic in 2014 and there has been a great deal of literature on this topic since that time. Moreover, the present analysis proposes to expand the population to a larger age range and will include both clinical and community samples. This study will provide a valuable update and encourage additional research in this area. Hence, the comments below are provided in an effort to clarify and further strengthen this exciting and much needed work. Specific comments follow below: 1. The indicated age range for inclusive studies is quite large, 5-24 years old. The authors indicate they will analyze results based on subgroups, including “age group”. It would be helpful to add what these proposed age groups will be particularly as this relates to differences in developmental risk for sleep disturbances (e.g., night terrors in a 5yo vs 24yo) and developmental risk for depression (e.g., depression pre-puberty vs post-puberty). Related to the large age range, can the authors clarify whether studies will be included if the study population is 24 years old at the beginning of the data collection? In other words, would a study be included where baseline sleep disturbance is measured at 24 years old and depression symptoms are measured at an older age? If this is the case, can the authors also comment in the introduction or elsewhere on the evidence for including youth over age 18 or 21 under the “youth” categorization? Most literature classifies “youth” as being younger than age 18. 2. It would be helpful if the authors could include an example/template search strategy (including specific terms to be used) particularly in how the authors will search for “sleep disturbances” and “dimensional construct of depression”. For example, in the description of the definition of sleep disturbances (page 8), the first definition is given as “increased/short/disrupted sleep duration” - Will this be increased relative to what is typical for the population age group? Or short relative to the age group? Or disrupted as in fragmented sleep duration? These terms should be

	clarified. Additionally, the authors indicate that disorders such as narcolepsy and sleep apnea “may also be included”. This list includes sleep disorders that appear to be more medically based than insomnia (although I understand many arguments can be made regarding the differences or lack of). Can the authors provide some rationale for including these sleep disorders as these may skew the results given that some of these disorders may have other impacts on depression (e.g., diagnosis of narcolepsy as a child may predict depression not only due to sleep disturbance but also due to being diagnosed with a chronic illness). Similarly, can the authors clarify what they mean by a “dimensional construct of depression”? Does this mean a full depression scale or could this also refer to a single item on a large epidemiological survey study? 3. In the section titled “Search strategy” (page 9), it is stated that “keywords terms relating to sleep disorders, depression/depressive disorder and children/adolescents/young adults…” An example/template search strategy would be helpful here as these keyword terms are quite broad. Additionally this is the first time that “young adults” are mentioned. 4. Discussion section on page 11 – please add a reference to the statement (and the ones that follow) that begins “Another relevant maturational change concerns the sleep-onset time,…” 5. There were no specific sensitivity analyses described although they were alluded to (e.g., risk of bias score). It would be helpful to explicitly state the sensitivity versus subgroup analyses that will be examined.

REVIEWER	Fiona Warren University of Exeter
REVIEW RETURNED	21-Apr-2020

GENERAL COMMENTS	This is a protocol for a systematic review and meta-analysis to investigate the relationship between poor sleep and depression in children and young adults. The subject area is interesting, and I believe this review will be of interest to clinicians and researchers in this area, and may have the potential to generate impact, as this is a potentially treatable risk factor for depression in children/young adults. At the higher level, therefore, I feel this systematic review and meta-analysis is potentially valuable; most of my comments relate to technical issues of the design and conduct, in many cases simply requiring more detail as to how the different steps in the study will be performed. Title 1. “In clinical and community-based samples, do insomnia and/or sleep disturbances compared to typical sleep predict child and youth depression?” Is it incidence of depression ie new onset of depression after suffering sleep disorders (ie sleep disorders predict new depression), is it simultaneous onset of depression occurring at the same time as sleep disorders, or is it severity of depression that occurs simultaneously with sleep disorders? I appreciate words are limited in the title, but if it is specifically incident depression

	then this could be added. Introduction 1. P4 lines 13-15. “Depression affects 2-4% of children and 11% of youth between 13 and 18 years of age,”. What is the definition of “child” used in this review? What is the youngest age at which depression can be reliably diagnosed? If the definition of “youth” ie 13-18 years used consistently throughout the review? 2. P1 lines 13-15. “and peaks at 20 years of age”. Is it the incidence of depression that peaks at 20 years, or is it the overall prevalence? 3. P4 lines 59-90, P5 lines 3-5. “It is currently not known whether treating sleep problems in affected or unaffected… would lead to a lower incidence of depression than it would be if sleep problems were not treated.” This sentence is not clear. Does the “affected or unaffected” refer to sleep problems (why would you treat them if they were unaffected?) or depression (if they are already affected with depression when being treated for sleep problems then they cannot have new incidence depression). 4. P5 lines 7-9. “Including clinical and non-clinical individuals,”. Not clear what is mean by clinical and non-clinical in this context. Does it refer to people with/without depression or people with/without sleep disorders? 5. P5 lines 18-24. I would refer to pooled estimates of the odds ratios rather than `meta-analytic odds ratios’ – the latter is not inaccurate but pooled estimate is the more usual terminology. The language is a little ambiguous here: for example I would use the word review for the overall review paper, study for an individual primary study (ie do not use the word “study” to mean a primary study and review in different contexts, especially in two adjacent sentences – in the second sentence the only way you know that the ORs refer to the meta-analysis and not the individual studies is because “meta-analytic” is stated), and pooled estimates for the actual results of an individual meta-analysis, bearing in mind that an individual review paper may include more than one pooled estimate. So something like: Across Z systematic reviews [REFS], pooled estimates for the odds ratio for depression comparing insomnia vs non-insomnia range between XX (95% CI) [REF] and YY (95% CI) [REF]. 6. P5 lines 22-25. “type of ascertainment (clinical: 2.05; non-clinical: 2.34),”. Does this mean ascertainment of depression or insomnia? Not sure what clinical and non-clinical means in this context? Is clinical referring to a formal diagnosis by a clinician whereas non-clinical refers to patient self report (or possibly report by parents for children)? 7. P5 lines 24-25. Need to clarify distinction between “elderly” and “adults” ie the elderly are included within adults? 8. P5 line 26. I would avoid using the term “subjects”. Maybe use participants, children, young adults etc as appropriate. 9. P5 line 28. “and a metaanalytic estimate, i.e. odds ratio= 2.0, was provided by only one of the three reviews”. Report the

	95% CI with the OR. It is stated that this meta-analysis included only 3 studies, how many individual participants did this include? Can we say any more about the original studies ie age range of participants, how was insomnia defined, how was depression reported as an outcome measure (ie discussing any potential clinical heterogeneity across studies)? 10. P5 lines 38-41. Repeats P5 lines 5-9. 11. P5. Issues around controlling for baseline depression. Assuming the premise of the meta-analysis is to investigate insomnia/sleep disorders as a predictor of depression, surely this would be performed in children/young adults who did not already have depression? If there is a mixture of participants with/without baseline depression (in the same primary study) it seems to make sense to control for it, although a sensitivity analysis could be performed to include those studies that did not control for baseline depression. It really needs to be made clear whether the intended sample includes participants who have a previous history of depression (and therefore insomnia is predicting future episodes) and/or participants who have no previous history of depression (for whom insomnia is predicting new onset depression). If the primary meta-analysis includes both groups, then I'd suggest a sensitivity analysis with only new onset depression (if there are primary studies with only new onset depression). The primary meta-analysis needs to be clearly defined (eg investigating the relationship between depression in children and YAs, including those with a history of depression and no history of depression, and sleep disorders including all primary studies reporting this analysis) and any sensitivity analyses (excluding studies that do not control for baseline depression; including only studies of new onset depression). P7 lines 3 to 30 does this to some extent but could be more clearly spelt out. 12. P6 line 55. Please explain more clearly what is meant by clinical samples? Is this children/YAs who have presented with sleep disorders or who have diagnosed depression at baseline? I assume community based samples includes children/YAs who have not presented clinically with sleep disturbances or depression previously? Methods 13. P9 first para. Mean age is used to determine eligibility of the paper and should lie between 5-24 years. I'd be a bit concerned that a mean age of 24 could include some participants who were substantively older than 24 and cannot be considered as `youth'. Would it not make more sense to select only studies where the cohort was defined as being under 18 (or under 21 as a maximum)? My understanding is that the paper aims to focus on children/young adults and 24 seems too old to me to be in this age bracket? 14. P9 Search strategy: could not find full search strategy for at least one database, eg MEDLINE (item 10 of PRISMA checklist). 15. P10, lines 15-18. "For the selection of studies based on

	title&abstract screening, [potentially] eligible studies will be selected by two [independent] reviewers until at least 80% of agreement is achieved with the remainder being selected by one reviewer." Does 80% agreement mean a kappa score of 0.8 (kappa measures inter-rater reliability, not agreement which is something different)? At this stage I would suggest screening out obvious non-eligible studies and obtaining full text if the reviewers do not agree. "For the selection of studies based on full-text screening, reviewers will perform the selection in duplicate. Any uncertainty regarding selection will be resolved by consensus and if necessary, it will be reviewed by a third reviewer". Simpler to say: Full text papers will be obtained for all references thought potentially eligible by a single reviewer based on title/abstract; all full text papers will be reviewed by two independent reviewers, with disagreements referred to a third reviewer. [I would suggest review by a senior author for any disagreements at this stage.] 16. P10 Data extraction. Need to state exactly which study characteristics (eg year of publication, sample size, country) will be extracted. For participant characteristics I assume you mean participant characteristics described at the study level, eg included age range, mean age, percentage by gender? I would make more distinction between study descriptors and outcome data that will be incorporated into the meta-analysis. 17. P10 line 39. "including effect sizes, odds ratio, relative risk or hazard ratio". Not sure if this is saying that the effect sizes that will be extracted include OR, RR, HR or whether effect size is another metric that could be extracted. I don't like the term effect size as it is often used to mean different things in different contexts without being defined, so I would suggest removing the term "effect size" unless you have a very specific meaning in mind, in which case it needs to be defined. 18. P10 line 40/41. I agree that in most cases the between group comparison metric ie OR, RR etc will be reported with 95% CI but some flexibility is needed here in case the outcome is reported eg as a log OR with standard error, or as the raw numbers per cell ie insomnia/no insomnia, depressed/non depressed. 19. P10 line 43. "Discrepancies between the two authors will be resolved by discussion". Does this mean discussion between the same two authors or will a third author be called in to resolve disputes? Also applies to risk of bias section. I would recommend adjudication by a senior author for any disagreements. 20. P11 line 22-24. "Data will be synthesized by a table reporting study type, participant characteristics, exposure and outcome measures, and a second table on the risk of bias assessment." I would call this the description of primary studies rather than a data synthesis. 21. What software will be used to perform the meta-analysis, including generating the I-squared statistics? 22. P11 Lines 24-26. "The data will be categorised into each type of exposure, i.e. insomnia and sleep disturbances." Are insomnia and other sleep disorders being considered separately or will there be an overall meta-analysis including

	both combined? 10. P11 Lines 32-34. I agree that it is important to report the I-squared statistic but the decision as to whether or not to perform a meta-analysis should not be based solely on the I-squared statistic. It is important to look at the clinical characteristics of each study, eg population definition, definitions of insomnia, depression etc, and then decide whether the studies are sufficiently similar or too different clinically to be validly combined. If the studies are to be combined, the decision to use fixed or random effects should also be based on clinical characteristics of the studies and not just the I-squared value (lines 36-40). 11. P11 lines 41-48. "Meta-analyses will be performed in subgroups stratified by key variables if there are ≥ 3 studies. Such variables include risk of bias score, type of ascertainment, type of sleep disturbance, measurement of outcome, presence of baseline comorbidity, sample size, age group, gender. Subgroups will be defined depending on the amount and type of information retrieved." I feel that different ideas are being conflated here. For example for study level characteristics such as the risk of bias score, sample size, included age range, etc, that presumably means a sensitivity analysis excluding studies with certain characteristics eg high risk of bias. Year of publication might be a relevant study level characteristic due to trends over time, eg increased screen use, that may have an effect on sleep. For ppt level characteristics such as presence of baseline comorbidity, age, gender etc, does that mean taking different estimates from the same study that adjust/don't adjust for comorbidity, or using estimates from different studies where some adjust for comorbidity and others don't? 12. Is timepoint of outcome measurements being considered, eg how long after start of follow-up (at which time point I assume insomnia status is ascertained) is depression ascertained? This could be a major point of clinical heterogeneity, eg the time difference from baseline to outcome measurement could vary from weeks to years, and the time difference in relation to the outcome could have a different influence depending on baseline age, eg sleep disturbances could take longer to have an impact on depression (assuming a causative effect) at different ages. I feel this requires some thought as to how it will be addressed eg through a sensitivity analysis. 13. Related to the above, would you consider trying to obtain individual participant level data, which might be feasible if most of the studies are fairly recent? IPD would allow more flexibility to explore subgroups such as different ages, genders, comorbidity etc, although the additional time resource required for an IPD meta-analysis should not be underestimated. 14. P13 lines 13-17. "While this link is well established and quantified in adult and elderly populations, a comprehensive and updated summary is missing in children and youth". This has already been stated twice in the introduction. PRISMA Checklist

	Item 10: could not find full draft of search strategy. Items 12/13: Needs more detail on extracted data in Data extraction section (ie actual variables rather than vague categories such as `participant characteristics').

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Jared M. Saletin PhD
Institution and Country: Department of Psychiatry and Human Behavior
Alpert Medical School of Brown University
Providence, RI, USA
Please state any competing interests or state 'None declared': None declared
Please leave your comments for the authors below

I appreciate the opportunity to review this proposed meta-analytic protocol to address a major issue: the overlap of sleep disturbance and depression like phenomena in children and adolescents. The authors propose to conduct a systematic meta-analysis of the effect of insomnia and sleep disturbance in predicting depression in people ages 5- 24.
While this registered protocol sticks to strong meta-analysis guidelines, has a strong plan for data extraction and bias appraisal, is registered and aligned with PRISMA, it's well purposed intention is paired with some methodological components that give me pause:

* The authors require a slew of summary statistics are required for studies to be included, does this limit the amount of evidence available?

Thank you for highlighting this. The list includes any possible statistics to be extracted. We re-phrased the paragraph changing 'and' into 'or':
'Outcome data will include odds ratio, relative risk, hazard ratio, regression coefficients or correlation coefficients with their respective measurements of error and p-values.'

* Dissertations and "grey data" will be included — these have yet to be peer reviewed and could limit the quality of evidence provided. It is not justified why these are included.

We agree that this decision has been little circumstantiated. We integrated the paragraph as follows: 'Grey literature, i.e. dissertations and conference proceeding/papers, will be included in the search in order to mitigate against publication bias.63 Dissertations have been noted as providing a valuable methodology and clinical source of information.64 Additionally, searching for relevant conference proceedings/papers is recommended in order to identify recent, but unpublished research.65'

* The authors seek to examine a wide age-range, how will the authors harmonize the definition of "insomnia" (or depression, for that matter) across these groups? I would recommend at minimum a hierarchical approach grouping child, adolescent, and adult participants, and perhaps the meta-analytic estimates being combined at a higher level. This will require careful thought.

We agree, thank you. We added the following paragraph in the Methods – Population Type section:

‘Mean age to determine eligibility for inclusion of the study is between 5-24 years. This age range was chosen to cover the whole age range of risk for depression in childhood and youth, and to include the period of peak incidence for depression. In our review, we identify three categories: ‘childhood’ between 5 and 9 years of age, ‘adolescence’ between 10 and 19 years of age, and ‘young adulthood’ between 20 and 24 years of age.’

And age was added as one of the relevant characteristics based on which to perform subgroup analyses:

‘Subgroup analyses will be performed related to relevant participant-level characteristics, e.g. baseline comorbidity, age at baseline, gender, duration of follow up, exposure and outcome definition and diagnostic tool, by calculating pooled estimates in subgroups of studies, e.g. studies of children versus adolescents versus young adults, and/or taking different estimates from the same study related to the subgroups of interest, e.g. male versus female.’

* By having a wide capture of sleep differences (and to that end, mixing “insomnia” and “sleep disturbances”), it’s unclear and unlikely that a single phenotype will emerge. In fact, a meta-analyses could be examined for almost every sub-type of sleep disturbance (or insomnia profile) mentioned.

Yes, we agree, thank you. This has been better addressed in the following paragraph:

‘Studies will be combined based on similarities across relevant characteristics, e.g. type of sample ascertainment, age at baseline, duration of follow-up, exposure and outcome definition and diagnostic tool, and meta-analysis will be performed if ≥ 3 studies are available.’

* The authors will accept variable reporting methods for both insomnia and depression. A focus on clinical report, or valid questionnaires only — rather than on self-report or caretaker report, would help to harmonize the quality evidence they may expect.

We agree with the reviewer that our inclusion criteria might contribute significantly to increase the heterogeneity across studies. We opted for broad inclusion criteria because we wanted to include as many studies as possible and explore the robustness of the primary hypotheses and potential source of heterogeneity by subsequent sensitivity and subgroup analyses.
This issue is now clarified in the following paragraph, Methods – Data Analysis section:

‘To test the robustness of the primary hypothesis against study-level characteristics, e.g. risk-of-bias score, sample size, year of publication, sensitivity analyses will be performed by excluding those studies with the characteristic that can potentially be a source of heterogeneity, e.g. excluding studies with a high risk-of-bias score. (…)
Subgroup analyses will be performed related to relevant participant-level characteristics, e.g. baseline comorbidity, age at baseline, gender, duration of follow up, exposure and outcome definition and diagnostic tool, by calculating pooled estimates in subgroups of studies, e.g. studies of children versus adolescents versus young adults, and/or taking different estimates from the same study related to the subgroups of interest, e.g. male versus female.’

* While excluding anxiety without depression is clear, no clear exclusion for mania is given.

Thank you for highlighting this. We added the following sentence:

‘Studies including participants with bipolar disorders will be excluded given that sleep problems’ prevalence differ significantly in bipolar versus unipolar disorders. 47 62’

* Mixing dimensional and diagnostic frames to study depression may conflate results.

We agree with the reviewer that including both dimensional and categorical constructs might contribute significantly to increase the heterogeneity across studies. We opted for broad inclusion criteria because we wanted to include as many studies as possible and we aim to explore the robustness of the primary hypotheses and potential source of heterogeneity by subsequent sensitivity and subgroup analyses, i.e. by calculating pooled estimates in subgroups of studies where constructs are dimensional versus categorical.

* The authors focus on DSM for both sleep and depression, however I wonder if ICSD would be better for the insomnia exposure?

Thank you for the suggestion, indeed ICSD is a valuable option. We would prefer to refer to the DSM as research including both depression and sleep problems focuses on this system rather than the ICSD. We hope the reviewer agrees with our view.

* Both introduction and discussion mention puberty as a pivot point for rates of depression betting dissociated between males and females — yet no analytic strategy for stratifying or testing trajectories by sex is provided, this would be a meaningful addition.

We agree, thank you. The following paragraph in the Methods – Data Analysis section has been re-phrased accordingly.

‘Subgroup analyses will be performed related to relevant participant-level characteristics, e.g. baseline comorbidity, age at baseline, gender, duration of follow up, exposure and outcome definition and diagnostic tool, by calculating pooled estimates in subgroups of studies, e.g. studies of children versus adolescents versus young adults, and/or taking different estimates from the same study related to the subgroups of interest, e.g. male versus female.’

Point-by-point replies to Reviewer 2 comments:
Reviewer: 2
Reviewer Name: Eleanor McGlinchey
Institution and Country: Fairleigh Dickinson University, New Jersey, USA
Columbia University Irving Medical Center, New York, USA
Please state any competing interests or state ‘None declared’: None declared
Please leave your comments for the authors below
Overall:
This is a well-written and timely protocol for a proposed systematic review and meta-analysis on the topic of sleep disturbance as a predictor of later depression in children and adolescents. Lovato & Gradisar last published a meta-analysis on this topic in 2014 and there has been a great deal of literature on this topic since that time. Moreover, the present analysis proposes to expand the population to a larger age range and will include both clinical and community samples. This study will provide a valuable update and encourage additional research in this area. Hence, the comments below are provided in an effort to clarify and further strengthen this exciting and much needed work. Specific comments follow below:

1. The indicated age range for inclusive studies is quite large, 5-24 years old. The authors indicate they will analyze results based on subgroups, including “age group”. It would be helpful to add what these proposed age groups will be particularly as this relates to differences in developmental risk for sleep disturbances (e.g., night terrors in a 5yo vs 24yo) and developmental risk for depression (e.g., depression pre-puberty vs post-puberty).
Related to the large age range, can the authors clarify whether studies will be included if the study population is 24 years old at the beginning of the data collection? In other words, would a study be included where baseline sleep disturbance is measured at 24 years old and depression symptoms are measured at an older age?
If this is the case, can the authors also comment in the introduction or elsewhere on the evidence for including youth over age 18 or 21 under the “youth” categorization? Most literature classifies “youth” as being younger than age 18.

We agree that more specification was needed.

We added a paragraph in Methods – Population type section and a new reference:

‘This age range was chosen to cover the whole age range of risk for depression in childhood and youth, and to include the period of peak incidence for depression.1 2 In our review, we identify three categories: ‘childhood’ between 5 and 9 years of age, ‘adolescence’ between 10 and 19 years of age, and ‘young adulthood’ between 20 and 24 years of age. For simplicity, the term ‘youth’ refers to individuals whose age is between 10 and 24 years.’

One reference was added:
Bundy DAP, de Silva N, Horton S, et al. Investment in child and adolescent health and development: key messages from Disease Control Priorities, 3rd Edition. Lancet 2018;391(10121):687-99. doi: 10.1016/S0140-6736(17)32417-0 [published Online First: 2017/11/21]

And in the Methods – Data Analysis section:

‘Subgroup analyses will be performed related to relevant participant-level characteristics, e.g. baseline comorbidity, age at baseline, gender, duration of follow up, exposure and outcome definition and diagnostic tool, by calculating pooled estimates in subgroups of studies, e.g. studies of children versus adolescents versus young adults, and/or taking different estimates from the same study related to the subgroups of interest, e.g. male versus female.’

Finally, a study would not be included if baseline sleep disturbance is measured at 24 years old and depression symptoms are measured at an older age.
The following sentence is in the Methods – Outcome section:

‘To be included, both exposure and outcome have to be measured at mean age ≥ 5 years and ≤ 24’

2. It would be helpful if the authors could include an example/template search strategy (including specific terms to be used) particularly in how the authors will search for “sleep disturbances” and “dimensional construct of depression”.

This is now included in the manuscript as figure 1.
‘Figure 1: Database: Ovid MEDLINE: Epub Ahead of Print, In-Process & Other Non-Indexed Citations,

Ovid MEDLINE® Daily and Ovid MEDLINE® <1946-Present>
Search Strategy:

1 exp Sleep Wake Disorders/
2 exp Sleep/
3 (apnea$ adj3 sleep$).mp.
4 (restless adj3 leg$ adj3 syndrome$).mp.
5 narcolep$.mp.
6 (night$ adj3 terror$).mp.
7 sleep$.mp.
8 dyssomni$.mp.
9 parasomni$.mp.
10 insomni$.mp.
11 exp Depressive Disorder/
12 exp Depression/
13 depress$.mp.
14 exp Child/
15 exp Adolescents/
16 exp Young Adult/
17 (child$ or juvenile$ or pubescen$ or prepubescen* or pre-pubescen* or teen$ or preteen$ or tween$ or youth$ or adoles$ or young$ adult$ or emerg$ adult$).mp.
18 exp Epidemiologic studies/
19 exp Case control studies/
20 exp Cohort studies/
21 case control.tw.
22 (cohort adj (study or studies)).tw.
23 cohort analy$.tw.
24 (follow up adj (study or studies)).tw.
25 (observational adj (study or studies)).tw.
26 longitudinal.tw.
27 retrospective.tw.
28 cross sectional.tw.
29 cross-sectional studies/
30 or/1-10
31 or/11-13
32 or/14-17
33 or/18-29
34 30 and 31 and 32 and 33
35 limit 34 to yr="1980 -Current" ’

For example, in the description of the definition of sleep disturbances (page 8), the first definition is given as “increased/short/disrupted sleep duration” - Will this be increased relative to what is typical for the population age group? Or short relative to the age group? Or disrupted as in fragmented sleep duration? These terms should be clarified.

We agree that this vague, thank you. The paragraph was re-worded as follows:

‘Sleep disturbances, defined as patterns of sleep duration and quality sleep duration which deviate from expected age-related norms, i.e. increased/short/fragmented sleep duration with or without daytime fatigue/sleepiness, nightmares/terror attacks, circadian disturbances, and/or night walking/talking.’

Additionally, the authors indicate that disorders such as narcolepsy and sleep apnea “may also be included”. This list includes sleep disorders that appear to be more medically based than insomnia (although I understand many arguments can be made regarding the differences or lack of). Can the authors provide some rationale for including these sleep disorders as these may skew the results given that some of these disorders may have other impacts on depression (e.g., diagnosis of narcolepsy as a child may predict depression not only due to sleep disturbance but also due to being diagnosed with a chronic illness).

Thank you for bringing this issue forward. We agree with the reviewer that including medically based sleep disorders might contribute significantly to increase the heterogeneity across studies and skew the results. We opted for including them because we wanted to be able to provide estimates that could be generalized as much as possible and that could be useful in different clinical settings. We plan to understand the extent to which these conditions might contribute to heterogeneity across studies by performing sensitivity analysis, i.e. excluding studies of medically related sleep problems. To address this issue, the following sentence is added to the Methods – Exposure section:

‘To increase the generalizability of the results, narcolepsy, breathing-related sleep disorders, central sleep apnea, non-rapid/rapid eye movement sleep disorder, restless legs syndrome, substance/medication-induced sleep disorder, and enuresis nocturna primaria may also be included.’

And to the Methods – Data Analysis section:

‘Likewise, a sensitivity analysis will be performed by removing studies focusing on medically based sleep disorders, i.e. narcolepsy, breathing-related sleep disorders, central sleep apnea, non-rapid/rapid eye movement sleep disorder, restless legs syndrome, substance/medication-induced sleep disorder, and enuresis nocturna primaria, to prevent the possibility that the pooled estimates could be impacted by factors related to the chronicity of these conditions.’

Similarly, can the authors clarify what they mean by a “dimensional construct of depression”? Does this mean a full depression scale or could this also refer to a single item on a large epidemiological survey study?
Thank you, we agree that this is ambiguous. We follow your suggestions and integrated the paragraphs as follows:

‘Studies will be included if they focus on either one of the following outcomes: major depressive disorder, depressive disorder not otherwise specified, dysthymic disorder, and/or dimensional constructs of depression, i.e. defined by scores from single- or multiple-item questionnaires, internalizing disorders or anxious depression.’

3. In the section titled “Search strategy” (page 9), it is stated that “keywords terms relating to sleep disorders, depression/depressive disorder and children/adolescents/young adults…” An example/template search strategy would be helpful here as these keyword terms are quite broad. Additionally this is the first time that “young adults” are mentioned.

Thank you for noting this. The search is now in the revised manuscript as figure 1.

‘Figure 1: Database: Ovid MEDLINE: Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE® Daily and Ovid MEDLINE® <1946-Present>
Search Strategy:

1 exp Sleep Wake Disorders/
2 exp Sleep/
3 (apnea$ adj3 sleep$).mp.
4 (restless adj3 leg$ adj3 syndrome$).mp.
5 narcolep$.mp.
6 (night$ adj3 terror$).mp.
7 sleep$.mp.
8 dyssomni$.mp.
9 parasomni$.mp.
10 insomni$.mp.
11 exp Depressive Disorder/
12 exp Depression/
13 depress$.mp.
14 exp Child/
15 exp Adolescents/
16 exp Young Adult/
17 (child$ or juvenile$ or pubescen$ or prepubescen* or pre-pubescen* or teen$ or preteen$ or tween$ or youth$ or adoles$ or young$ adult$ or emerg$ adult$).mp.
18 exp Epidemiologic studies/
19 exp Case control studies/
20 exp Cohort studies/
21 case control.tw.
22 (cohort adj (study or studies)).tw.
23 cohort analy$.tw.
24 (follow up adj (study or studies)).tw.
25 (observational adj (study or studies)).tw.
26 longitudinal.tw.
27 retrospective.tw.
28 cross sectional.tw.
29 cross-sectional studies/
30 or/1-10
31 or/11-13
32 or/14-17
33 or/18-29
34 30 and 31 and 32 and 33
35 limit 34 to yr="1980 -Current" ‘

Thank you for noting that ‘young adults’ was out of context. Now, this paragraph has been added in the Methods – Population Type sections:

‘In our review, we identify three categories: ‘childhood’ between 5 and 9 years of age, ‘adolescence’ between 10 and 19 years of age, and ‘young adulthood’ between 20 and 24 years of age.61 For simplicity, the term ‘youth’ refers to individuals whose age is between 10 and 24 years. ‘

4. Discussion section on page 11 – please add a reference to the statement (and the ones that follow) that begins “Another relevant maturational change concerns the sleep-onset time,…”

Done, thank you.
‘Hummer DL, Lee TM. Daily timing of the adolescent sleep phase: Insights from a cross-species comparison. Neurosci Biobehav Rev 2016;70:171-81. doi: 10.1016/j.neubiorev.2016.07.023

[published Online First: 2016/10/22]'

5. There were no specific sensitivity analyses described although they were alluded to (e.g., risk of bias score). It would be helpful to explicitly state the sensitivity versus subgroup analyses that will be examined.

Thank you for highlighting this. The following paragraph has been added in Method – Data Analysis:

'To test the robustness of the primary hypothesis against study-level characteristics, e.g. risk-of-bias score, sample size, year of publication, sensitivity analyses will be performed by excluding those studies with the characteristic that can potentially be a source of heterogeneity, e.g. excluding studies with a high risk-of-bias score. Likewise, a sensitivity analysis will be performed by removing studies focusing on medically based sleep disorders, i.e. narcolepsy, breathing-related sleep disorders, central sleep apnea, non-rapid/rapid eye movement sleep disorder, restless legs syndrome, substance/medication-induced sleep disorder, and enuresis nocturna primaria, to prevent the possibility that the pooled estimates could be impacted by factors related to the chronicity of these conditions.
Subgroup analyses will be performed related to relevant participant-level characteristics, e.g. baseline comorbidity, age at baseline, gender, duration of follow up, exposure and outcome definition and diagnostic tool, by calculating pooled estimates in subgroups of studies, e.g. studies of children versus adolescents versus young adults, and/or taking different estimates from the same study related to the subgroups of interest, e.g. male versus female.'

Point-by-point replies to Reviewer 3 comments:
Reviewer: 3
Reviewer Name: Fiona Warren
Institution and Country: University of Exeter
Please state any competing interests or state 'None declared': None declared
Marino et al.: In clinical and community-based samples, do insomnia and/or sleep disturbances compared to typical sleep predict child and youth depression? Protocol for a systematic review and meta-analysis
This is a protocol for a systematic review and meta-analysis to investigate the relationship between poor sleep and depression in children and young adults. The subject area is interesting, and I believe this review will be of interest to clinicians and researchers in this area, and may have the potential to generate impact, as this is a potentially treatable risk factor for depression in children/young adults. At the higher level, therefore, I feel this systematic review and meta-analysis is potentially valuable; most of my comments relate to technical issues of the design and conduct, in many cases simply requiring more detail as to how the different steps in the study will be performed.

Title 1. "In clinical and community-based samples, do insomnia and/or sleep disturbances compared to typical sleep predict child and youth depression?"
Is it incidence of depression ie new onset of depression after suffering sleep disorders (ie sleep disorders predict new depression), is it simultaneous onset of depression occurring at the same time as sleep disorders, or is it severity of depression that occurs simultaneously with sleep disorders? I appreciate words are limited in the title, but if it is specifically incident depression then this could be added.

Thank you for highlighting this, we agree that this needs to be clarified. However, this is not easy to address because the outcome in the included studies will be both a continuous and a categorical measure. Furthermore, included studies will be both controlling and not controlling for baseline depression.
Therefore, we propose the following change in the title

'Do insomnia and/or sleep disturbances predict the onset, relapse or worsening of depression in community and clinical samples of children and youth? Protocol for a systematic review and meta-analysis'.

We hope that this choice meets the reviewer request.

Introduction
1. P4 lines 13-15. "Depression affects 2-4% of children and 11% of youth between 13 and 18 years of age,". What is the definition of "child" used in this review? What is the youngest age at which depression can be reliably diagnosed? Is the definition of "youth" ie 13-18 years used consistently throughout the review?

In our review, definitions are 'child' 5-10 y, 'adolescent' 11-18 y, 'young adult' > 19-24 y. We added the following paragraph in the 'Methods-Population type' section and its relative reference:

'Mean age to determine eligibility for inclusion of the study is between 5-24 years. This age range was chosen to cover the whole age range of risk for depression in childhood and youth, and to include the period of peak incidence for depression.1 2 In our review, we identify three categories: 'childhood' between 5 and 9 years of age, 'adolescence' between 10 and 19 years of age, and 'young adulthood' between 20 and 24 years of age.61 For simplicity, the term 'youth' refers to individuals whose age is between 10 and 24 years.'

Depression can be reliably diagnosed in this age range (see ref. "2": Park RJ, Goodyer IM. Clinical guidelines for depressive disorders in childhood and adolescence. Eur Child Adolesc Psychiatry 2000;9(3):147-61).

We agree that there must be congruence between our definition for child/youth/adult and the reported estimates for depression. However, we have not found estimates for these exact age ranges, although we might be wrong. We opted to report prevalence estimates of depression from the 2016 National Survey of Children's Health for the age-ranges 6-11 and 12-17 years as the best compromise. The text was revised accordingly as follows:

'A recent epidemiological study analyzing data from the 2016 National Survey of Children's Health reports prevalence estimates of depression of 1.7% and 6.1% in 6-11 and 12-17 years age-ranges, respectively.' This study was added to the reference list (https://doi.org/10.1016/j.jpeds.2018.09.021).

We hope that the reviewer agrees with our choice.

2. P1 lines 13-15. "and peaks at 20 years of age". Is it the incidence of depression that peaks at 20 years, or is it the overall prevalence?

Incidence. The sentence is revised accordingly, thank you.

3. P4 lines 59-90, P5 lines 3-5. “It is currently not known whether treating sleep problems in affected or unaffected… would lead to a lower incidence of depression than it would be if sleep problems were not treated.” This sentence is not clear. Does the “affected or unaffected” refer to sleep problems (why would you treat them if they were unaffected?) or depression (if they are already affected with depression when being treated for sleep problems then they cannot have new incidence depression).

We agree, thank you. We re-phrased as follows:

‘It is currently not known whether treating sleep problems in depressed or at risk for depression children (e.g. children with history of abuse, with high emotional dysregulation, offspring of depressed parents) would lead to a reduction or lower incidence of depression, respectively, than it would be if sleep problems were not treated.

4. P5 lines 7-9. “Including clinical and non-clinical individuals,”. Not clear what is meant by clinical and non-clinical in this context. Does it refer to people with/without depression or people with/without sleep disorders?

We agree that this is not clear. We rephrased as follows:

‘including depressed and non-depressed individuals.’

5. P5 lines 18-24. I would refer to pooled estimates of the odds ratios rather than `meta-analytic odds ratios’ – the latter is not inaccurate but pooled estimate is the more usual terminology.

Corrected, thank you.

The language is a little ambiguous here: for example I would use the word review for the overall review paper, study for an individual primary study (ie do not use the word “study” to mean a primary study and review in different contexts, especially in two adjacent sentences – in the second sentence the only way you know that the ORs refer to the meta-analysis and not the individual studies is because “meta-analytic” is stated), and pooled estimates for the actual results of an individual meta-analysis, bearing in mind that an individual review paper may include more than one pooled estimate. So something like: Across Z systematic reviews [REFS], pooled estimates for the odds ratio for depression 2 comparing insomnia vs non-insomnia range between XX (95% CI) [REF] and YY (95% CI) [REF].

We integrated your suggestions as follows:

‘Pooled estimates are consistent across reviews ranging 2.10 (95% CI= 1.86-2.38)26-2.83 (95% CI= 1.55-5.17)24, and across subgroups stratified by the most common confounders, e.g. sex [male: 1.46 (95% CI= 1.13-1.88)25; females: 1.96 (95% CI= 1.05-3.66)25], type of ascertainment [non-general population: 2.05 (95% CI= 1.53-2.74)25; general population: 2.34 (95% CI= 1.85-2.96)25], and age [individuals of age ≥ 60 years: 1.87 (95% CI= 1.47, 2.37)25-1.92 (95% CI: 1.61–2.30)23; individuals younger than 60 years of age: 2.50 (95% CI= 1.96, 3.20)25].’

6. P5 lines 22-25. “type of ascertainment (clinical: 2.05; non-clinical: 2.34),”. Does this mean ascertainment of depression or insomnia? Not sure what clinical and non-clinical means in this

context? Is clinical referring to a formal diagnosis by a clinician whereas non-clinical refers to patient self report (or possibly report by parents for children)?

We agree, labels are not clear. In Li et al. the reported statistics refer to: 1) general population, defined as a sample representative of a community or population; 2) non-general population, defined as sample drawn from a special population. Accordingly, we now specify labels as follows:

'[non-general population: 2.05 (95% CI= 1.53-2.74)25; general population: 2.34 (95% CI= 1.85-2.96)25].'

7. P5 lines 24-25. Need to clarify distinction between "elderly" and "adults" ie the elderly are included within adults?

We agree, thank you. We specified:

'age [individuals of age ≥ 60 years: 1.87 (95% CI= 1.47, 2.37)25-1.92 (95% CI: 1.61–2.30)23; individuals younger than 60 years of age: 2.50 (95% CI= 1.96, 3.20)25].

8. P5 line 26. I would avoid using the term "subjects". Maybe use participants, children, young adults etc as appropriate.

We opted for participants, thank you, and this was changed throughout the ms.

9. P5 line 28. "and a meta-analytic estimate, i.e. odds ratio= 2.0, was provided by only one of the three reviews". Report the 95% CI with the OR.

Done, thank you.

It is stated that this meta-analysis included only 3 studies, how many individual participants did this include? Can we say any more about the original studies ie age range of participants, how was insomnia defined, how was depression reported as an outcome measure (ie discussing any potential clinical heterogeneity across studies)?

Yes, sure, this is relevant. We integrated the paragraph as follows:

'This estimate was based on three studies up to 2010 and included 7404 participants, whose age ranged between 6 and 16 years.27-29 In all three studies, depression was defined according to DSM IV, and insomnia to DSM-IV-TR criteria. The test for heterogeneity for this group did not show a significant index (Q-value= 0.3; df(Q)=2; p=0.9; and I2=0.0).'

10. P5 lines 38-41. Repeats P5 lines 5-9.
The redundant paragraph P5 lines 38-41 was deleted.

11. P5. Issues around controlling for baseline depression. Assuming the premise of the meta-analysis

is to investigate insomnia/sleep disorders as a predictor of depression, surely this would be performed in children/young adults who did not already have depression? If there is a mixture of participants with/without baseline depression (in the same primary study) it seems to make sense to control for it, although a sensitivity analysis could be performed to include those studies that did not control for baseline depression. It really needs to be made clear whether the intended sample includes participants who have a previous history of depression (and therefore insomnia is predicting future episodes) and/or participants who have no previous history of depression (for whom insomnia is predicting new onset depression). If the primary meta-analysis includes both groups, then I'd suggest a sensitivity analysis with only new onset depression (if there are primary studies with only new onset depression). The primary meta-analysis needs to be clearly defined (eg investigating the relationship between depression in children and YAs, including those with a history of depression and no history of depression, and sleep disorders including all primary studies reporting this analysis) and any sensitivity analyses (excluding studies that do not control for baseline depression; including only studies of new onset depression). P7 lines 3 to 30 does this to some extent but could be more clearly spelt out.

We agree that this issue needed to be clarified. We re-wrote the paragraph as follows:

'We are interested in testing this relationship in the context of both the absence and presence of baseline depression, i.e. whether insomnia and/or sleep disturbance predict both new and relapse/worsening of depression, respectively.
Our study will add to the currently available knowledge synthesis in that:
1. It will provide an updated meta-analytic estimate of the effects of insomnia predicting child and youth depression, adding studies from 2010 onward compared to Baglioni et al.26
2. By including studies on 5-24-year-old participants from both clinical and non-clinical populations, it will provide a more comprehensive and updated meta-analytic estimate compared to Lovato and Gradisar51
3. It will include studies on participants with/without baseline depression and whether or not controlling for baseline depression
4. It will include studies of both primary and comorbid insomnia and/or sleep disturbances
As a secondary aim, we will test whether insomnia and/or sleep disturbance predict new onset and/or worsening/relapse of depression, by testing the pooled estimate in two subgroups of studies, respectively, those including participants with baseline depression or where baseline depression has been controlled for, and studies including participants without baseline depression or where baseline depression has not been controlled for.'

12. P6 line 55. Please explain more clearly what is meant by clinical samples? Is this children/YAs who have presented with sleep disorders or who have diagnosed depression at baseline? I assume community-based samples includes children/YAs who have not presented clinically with sleep disturbances or depression previously?

It refers to the source of proband ascertainment. To address clarity, the paragraph was re-phrased as follows:

'The purpose of this protocol is to outline methods for a systematic review and meta-analysis of data from prospective, longitudinal studies of children and youth ascertained from clinic or community-based samples investigating the role of insomnia and/or sleep disturbances compared to typical sleep as predictors of depression'

Methods
13. P9 first para. Mean age is used to determine eligibility of the paper and should lie between 5-24 years. I'd be a bit concerned that a mean age of 24 could include some participants who were substantively older than 24 and cannot be considered as `youth'. Would it not make more sense to select only studies where the cohort was defined as being under 18 (or under 21 as a maximum)? My understanding is that the paper aims to focus on children/young adults and 24 seems too old to me to be in this age bracket?

This choice has long been debated among co-authors. We are aware that this choice implies that young adults are included. However, our aim was to cover the age range where the incidence of depression peaks. To substantiate our choice, the paragraph was integrated as follows:

'Mean age to determine eligibility for inclusion of the study is between 5-24 years. This age range was chosen to cover the whole age range of risk for depression in childhood and youth, and to include the period of peak incidence for depression.'

We hope that the reviewer agrees with our choice and finds it well substantiated.

14. P9 Search strategy: could not find full search strategy for at least one database, eg MEDLINE (item 10 of PRISMA checklist).
Thank you for noting this. The search was added as figure 1.

Figure 1: Database: Ovid MEDLINE: Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE® Daily and Ovid MEDLINE® <1946-Present>
Search Strategy:

1 exp Sleep Wake Disorders/
2 exp Sleep/
3 (apnea$ adj3 sleep$).mp.
4 (restless adj3 leg$ adj3 syndrome$).mp.
5 narcolep$.mp.
6 (night$ adj3 terror$).mp.
7 sleep$.mp.
8 dyssomni$.mp.
9 parasomni$.mp.
10 insomni$.mp.
11 exp Depressive Disorder/
12 exp Depression/
13 depress$.mp.
14 exp Child/
15 exp Adolescents/
16 exp Young Adult/
17 (child$ or juvenile$ or pubescen$ or prepubescen* or pre-pubescen* or teen$ or preteen$ or tween$ or youth$ or adoles$ or young$ adult$ or emerg$ adult$).mp.
18 exp Epidemiologic studies/
19 exp Case control studies/
20 exp Cohort studies/
21 case control.tw.
22 (cohort adj (study or studies)).tw.
23 cohort analy$.tw.
24 (follow up adj (study or studies)).tw.
25 (observational adj (study or studies)).tw.

26 longitudinal.tw.
27 retrospective.tw.
28 cross sectional.tw.
29 cross-sectional studies/
30 or/1-10
31 or/11-13
32 or/14-17
33 or/18-29
34 30 and 31 and 32 and 33
35 limit 34 to yr="1980 -Current"

15. P10, lines 15-18. “For the selection of studies based on title&abstract screening, [potentially] eligible studies will be selected by two [independent] reviewers until at least 80% of agreement is achieved with the remainder being selected by one reviewer.” Does 80% agreement mean a kappa score of 0.8 (kappa measures inter-rater reliability, not agreement which is something different)? At this stage I would suggest screening out obvious noneligible studies and obtaining full text if the reviewers do not agree. “For the selection of studies based on full-text screening, reviewers will perform the selection in duplicate. Any uncertainty regarding selection will be resolved by consensus and if necessary, it will be reviewed by a third reviewer”. Simpler to say: Full text papers will be obtained for all references thought potentially eligible by a single reviewer based on title/abstract; all full text papers will be reviewed by two independent reviewers, with disagreements referred to a third reviewer. [I would suggest review by a senior author for any disagreements at this stage.]

Thank you, we agree and implemented your suggestions.
The paragraph is as follows:

‘For the selection of studies based on title&abstract screening, eligible studies will be selected by two reviewers until a kappa score of 0.8 is achieved with the remainder being selected by one reviewer. Full text papers will be obtained for all references thought potentially eligible by a single reviewer based on title&abstract; all full text papers will be reviewed by two independent reviewers, with disagreements referred to a third, senior-author reviewer.’

16. P10 Data extraction. Need to state exactly which study characteristics (eg year of publication, sample size, country) will be extracted. For participant characteristics I assume you mean participant characteristics described at the study level, eg included age range, mean age, percentage by gender? I would make more distinction between study descriptors and outcome data that will be incorporated into the meta-analysis.

We agree, thank you. The paragraph has been re-written following your indication:

‘Extracted variables will pertain to the following domains: general study information, study characteristics (i.e. year of publication, type of sample ascertainment, sample size), participant characteristics (i.e. age at baseline, age at follow up, number of follow up, interval between baseline and each follow-up, exposure and outcome definition and diagnostic tool, covariates that were adjusted in the analyses). Outcome data will include odds ratio, relative risk, hazard ratio, regression coefficients or correlation coefficients with their respective measurements of error and p-values.’

17. P10 line 39. “including effect sizes, odds ratio, relative risk or hazard ratio”. Not sure if this is saying that the effect sizes that will be extracted include OR, RR, HR or whether effect size is another

metric that could be extracted. I don't like the term effect size as it is often used to mean different things in different contexts without being defined, so I would suggest removing the term "effect size" unless you have a very specific meaning in mind, in which case it needs to be defined.
18. P10 line 40/41. I agree that in most cases the between group comparison metric ie OR, RR etc will be reported with 95% CI but some flexibility is needed here in case the outcome is reported eg as a log OR with standard error, or as the raw numbers per cell ie insomnia/no insomnia, depressed/non depressed.

To account for points 17 and 18, the outcome paragraph was re-written as follows:

'Outcome data will include odds ratio, relative risk, hazard ratio, regression coefficients or correlation coefficients with their respective measurements of error and p-values.'

19. P10 line 43. "Discrepancies between the two authors will be resolved by discussion". Does this mean discussion between the same two authors or will a third author be called in to resolve disputes? Also applies to risk of bias section. I would recommend adjudication by a senior author for any disagreements.

Agree. With a third, senior-author reviewer was added.'

20. P11 line 22-24. "Data will be synthesized by a table reporting study type, participant characteristics, exposure and outcome measures, and a second table on the risk of bias assessment." I would call this the description of primary studies rather than a data synthesis.

Agree, done.

21. What software will be used to perform the meta-analysis, including generating the I-squared statistics?

The following specification was added: 'STATA 14 (StataCorp. 2015. Stata Statistical Software: Release 14. College Station, TX: StataCorp LP).'

22. P11 Lines 24-26. "The data will be categorised into each type of exposure, i.e. insomnia and sleep disturbances." Are insomnia and other sleep disorders being considered separately or will there be an overall meta-analysis including both combined?

The meta-analysis will include studies from both groups. We agree that this sentence is confusing, it was therefore deleted.

10. P11 Lines 32-34. I agree that it is important to report the I-squared statistic but the decision as to whether or not to perform a meta-analysis should not be based solely on the I-squared statistic. It is important to look at the clinical characteristics of each study, eg population definition, definitions of insomnia, depression etc, and then decide whether the studies are sufficiently similar or too different clinically to be validly combined. If the studies are to be combined, the decision to use fixed or random effects should also be based on clinical characteristics of the studies and not just the I-squared value (lines 36-40).

Thank you for this input, we totally agree. The paragraph was re-phrased as follows:

‘Studies will be combined based on similarities across relevant characteristics, e.g. type of sample ascertainment, age at baseline, duration of follow-up, exposure and outcome definition and diagnostic tool, and meta-analysis will be performed if ≥ 3 studies are available.
Heterogeneity of pooled estimates will be measured using the I2 statistic, which represents the percentage of variation attributable to between-study heterogeneity. I2 values of 25%, 50%, and 75% will be considered as low, medium, and high heterogeneity, respectively.63 For the meta-analysis, the choice between a random-effect or fixed-effect model will be based both on the I2 statistics and the participant-level characteristics of the combined studies.’

11. P11 lines 41-48. “Meta-analyses will be performed in subgroups stratified by key variables if there are ≥ 3 studies. Such variables include risk of bias score, type of ascertainment, type of sleep disturbance, measurement of outcome, presence of baseline comorbidity, sample size, age group, gender. Subgroups will be defined depending on the amount and type of information retrieved.” I feel that different ideas are being conflated here. For example for study level characteristics such as the risk of bias score, sample size, included age range, etc, that presumably means a sensitivity analysis excluding studies with certain characteristics eg high risk of bias. Year of publication might be a relevant study level characteristic due to trends over time, eg increased screen use, that may have an effect on sleep. For ppt level characteristics such as presence of baseline comorbidity, age, gender etc, does that mean taking different estimates from the same study that adjust/don’t adjust for comorbidity, or using estimates from different studies where some adjust for comorbidity and others don’t?

We agree. The paragraph was re-written as follows:

‘To test the robustness of the primary hypothesis against study-level characteristics, e.g. risk-of-bias score, sample size, year of publication, sensitivity analyses will be performed by excluding those studies with the characteristic that can potentially be a source of heterogeneity, e.g. excluding studies with a high risk-of-bias score. (…)
Subgroup analyses will be performed related to relevant participant-level characteristics, e.g. baseline comorbidity, age at baseline, gender, duration of follow up, exposure and outcome definition and diagnostic tool, by calculating pooled estimates in subgroups of studies, e.g. studies of children versus adolescents versus young adults, and/or taking different estimates from the same study related to the subgroups of interest, e.g. male versus female.’

12. Is timepoint of outcome measurements being considered, eg how long after start of follow up (at which time point I assume insomnia status is ascertained) is depression ascertained? This could be a major point of clinical heterogeneity, eg the time difference from baseline to outcome measurement could vary from weeks to years, and the time difference in relation to the outcome could have a different influence depending on baseline age, eg sleep disturbances could take longer to have an impact on depression (assuming a causative effect) at different ages. I feel this requires some thought as to how it will be addressed eg through a sensitivity analysis.

The following sentence has been added to ‘Method – Outcome’:

‘and the time between the exposure and outcome is ≥ one month.’
And duration of follow up was added as a relevant characteristic for subgroup analyses.

13. Related to the above, would you consider trying to obtain individual participant level data, which might be feasible if most of the studies are fairly recent? IPD would allow more flexibility to explore subgroups such as different ages, genders, comorbidity etc, although the additional time resource required for an IPD meta-analysis should not be underestimated.

We considered this option. However due to time and finance constraints we opted not to collect IPD.

14. P13 lines 13-17. “While this link is well established and quantified in adult and elderly populations, a comprehensive and updated summary is missing in children and youth”. This has already been stated twice in the introduction.

This was deleted.

PRISMA Checklist
Item 10: could not find full draft of search strategy.
This is now in the revised version as figure 1.

‘Figure 1: Database: Ovid MEDLINE: Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE® Daily and Ovid MEDLINE® <1946-Present>
Search Strategy:

1 exp Sleep Wake Disorders/
2 exp Sleep/
3 (apnea$ adj3 sleep$).mp.
4 (restless adj3 leg$ adj3 syndrome$).mp.
5 narcolep$.mp.
6 (night$ adj3 terror$).mp.
7 sleep$.mp.
8 dyssomni$.mp.
9 parasomni$.mp.
10 insomni$.mp.
11 exp Depressive Disorder/
12 exp Depression/
13 depress$.mp.
14 exp Child/
15 exp Adolescents/
16 exp Young Adult/
17 (child$ or juvenile$ or pubescen$ or prepubescen* or pre-pubescen* or teen$ or preteen$ or tween$ or youth$ or adoles$ or young$ adult$ or emerg$ adult$).mp.
18 exp Epidemiologic studies/
19 exp Case control studies/
20 exp Cohort studies/
21 case control.tw.
22 (cohort adj (study or studies)).tw.
23 cohort analy$.tw.
24 (follow up adj (study or studies)).tw.
25 (observational adj (study or studies)).tw.
26 longitudinal.tw.
27 retrospective.tw.
28 cross sectional.tw.

29 cross-sectional studies/
30 or/1-10
31 or/11-13
32 or/14-17
33 or/18-29
34 30 and 31 and 32 and 33
35 limit 34 to yr="1980 -Current" ‘

Items 12/13: Needs more detail on extracted data in Data extraction section (ie actual variables rather than vague categories such as `participant characteristics’).

This is now detailed in the ‘Methods – Data extraction’ section:

‘Extracted variables will pertain to the following domains: general study information, study characteristics (i.e. year of publication, type of sample ascertainment, sample size), participant characteristics (i.e. age at baseline, age at follow up, number of follow up, interval between baseline and each follow-up, exposure and outcome definition and diagnostic tool, covariates that were adjusted in the analyses). Outcome data will include odds ratio, relative risk, hazard ratio, regression coefficients or correlation coefficients with their respective measurements of error and p-values.’

VERSION 2 – REVIEW

REVIEWER	Eleanor McGlinchey Fairleigh Dickinson University, New Jersey, USA Columbia University Irving Medical Center, New York, USA
REVIEW RETURNED	29-May-2020

GENERAL COMMENTS	This is a highly responsive revision to my original reviewer comments. I think the already important contribution is further strengthened by this revision. I also appreciated the supplemental search strategy. I think that the age range for the adolescent subgroup is quite large (10-19), but this could be explored as a variable of interest once results have been obtained. I have no additional revisions requested and look forward to seeing this work published.

REVIEWER	Fiona Warren University of Exeter, UK
REVIEW RETURNED	08-Jun-2020

GENERAL COMMENTS	I have commented further on some of the authors' responses. Overall the manuscript is much clearer in terms of the aims and methods of the study. I would be happy for the manuscript to be published in BMJ Open with minor revisions, and wish the authors success in carrying out this review. 6. P5 lines 22-25. “type of ascertainment (clinical: 2.05; non-clinical: 2.34),”. Does this mean ascertainment of depression or insomnia? Not sure what clinical and non-clinical means in this context? Is

	clinical referring to a formal diagnosis by a clinician whereas non-clinical refers to patient self report (or possibly report by parents for children)? Authors' response We agree, labels are not clear. In Li et al. the reported statistics refer to: 1) general population, defined as a sample representative of a community or population; 2) non-general population, defined as sample drawn from a special population. Accordingly, we now specify labels as follows: '[non-general population: 2.05 (95% CI= 1.53-2.74)25; general population: 2.34 (95% CI= 1.85-2.96)25].' My response: I would suggest including the definitions of general and non-general populations as used in the reference. 23 P11 Lines 32-34. I agree that it is important to report the I-squared statistic but the decision as to whether or not to perform a meta-analysis should not be based solely on the I-squared statistic. It is important to look at the clinical characteristics of each study, eg population definition, definitions of insomnia, depression etc, and then decide whether the studies are sufficiently similar or too different clinically to be validly combined. If the studies are to be combined, the decision to use fixed or random effects should also be based on clinical characteristics of the studies and not just the I-squared value (lines 36-40) Authors' response: Thank you for this input, we totally agree. The paragraph was re-phrased as follows: 'Studies will be combined based on similarities across relevant characteristics, e.g. type of sample ascertainment, age at baseline, duration of follow-up, exposure and outcome definition and diagnostic tool, and meta-analysis will be performed if ≥ 3 studies are available.Heterogeneity of pooled estimates will be measured using the I2 statistic, which represents the percentage of variation attributable to between-study heterogeneity. I2 values of 25%, 50%, and 75% will be considered as low, medium, and high heterogeneity, respectively.63 For the meta-analysis, the choice between a random-effect or fixed-effect model will be based both on the I2 statistics and the participant-level characteristics of the combined studies.' My response: This is much improved – I would go slightly further in terms of being more general and say "For the meta-analysis, the choice between a random-effect or fixed-effect model will be based both on the I-squared statistics and the participant and study characteristics of the combined studies.' – thus taking into study level characteristics such as definition of outcome etc. Also applies to Abstract. 24 P11 lines 41-48. "Meta-analyses will be performed in subgroups stratified by key variables if there are ≥ 3 studies. Such variables include risk of bias score, type of ascertainment, type of sleep disturbance, measurement of outcome, presence of baseline comorbidity, sample size, age group, gender. Subgroups will be defined depending on the amount and type of information retrieved." I feel that different ideas are being conflated here. For example for study level characteristics such as the risk of bias score, sample size, included age range, etc, that presumably means a sensitivity analysis excluding studies with certain characteristics eg high risk of bias. Year of publication might be a relevant study level characteristic due to trends over time, eg increased screen use, that may have an effect on sleep. For ppt level characteristics such as presence of baseline comorbidity, age, gender etc, does that mean taking different estimates from the same study that adjust/don't

	adjust for comorbidity, or using estimates from different studies where some adjust for comorbidity and others don't? Authors' response: We agree. The paragraph was re-written as follows: 'To test the robustness of the primary hypothesis against study-level characteristics, e.g. risk-of-bias score, sample size, year of publication, sensitivity analyses will be performed by excluding those studies with the characteristic that can potentially be a source of heterogeneity, e.g. excluding studies with a high risk-of-bias score. (…) Subgroup analyses will be performed related to relevant participant-level characteristics, e.g. baseline comorbidity, age at baseline, gender, duration of follow up, exposure and outcome definition and diagnostic tool, by calculating pooled estimates in subgroups of studies, e.g. studies of children versus adolescents versus young adults, and/or taking different estimates from the same study related to the subgroups of interest, e.g. male versus female.' My response: Again this is much improved, but I am still concerned that there is not a clear distinction between participant level characteristics (age, gender etc) and study level characteristics. I would think that length of follow-up would be a study level characteristic (at least in theory, although I acknowledge that in reality there may be some loss to follow-up prior to the formal end of the study so this may be a participant level characteristic), also exposure/outcome definition, diagnostic tool would be study level characteristics. I feel that more detail is required to disentangle participant and study level characteristics.

VERSION 2 – AUTHOR RESPONSE

Thank you for your letter dated July 8, 2020 regarding our paper 'Do insomnia and/or sleep disturbances predict the onset, relapse or worsening of depression in community and clinical samples of children and youth? Protocol for a systematic review and meta-analysis.' (manuscript number: bmjopen-2019-034606.R1) in which you encourage resubmission to BMJ Open after minor revisions. We have scrupulously applied the revisions suggested by Reviewer 3.