Author's response to reviews

Title: Sleep complaints in adolescent depression: One year naturalistic follow-up study

Authors:

Anna S Urrila (anna.urrila@helsinki.fi)
Linnea Karlsson (linnea.karlsson@utu.fi)
Olli Kiviruusu (olli.kiviruusu@thl.fi)
Maiju Pankakoski (maiju.pankakoski@thl.fi)
Mirjami Pelkonen (mirjami.pelkonen@thl.fi)
Thea Strandholm (thea.strandholm@thl.fi)
Mauri Marttunen (mauri.marttunen@thl.fi)

Version: 2
Date: 18 June 2014

Author's response to reviews: see over
Helsinki, June 18, 2014

Dear Editor,

We would like to thank the reviewers for the thoughtful comments on our manuscript. In the revised version of the manuscript we have paid more attention particularly to providing more information on the study selection criteria, justifying the study measures, and justifying the categorization of subjects according to sleep complaints. Based on the reviewers’ valuable comments we have also made changes to the figures and added a table for clearer presentation of the results. The point-by point response to the reviewers’ comments is enclosed.

Please let me know if there are any questions concerning the revised manuscript.

On behalf of all the authors,

Sincerely,

Anna Sofia Urrila
POINT-BY-POINT RESPONSE TO REVIEWERS' COMMENTS

Manuscript title: Sleep complaints in adolescent depression: One year naturalistic follow-up study

Manuscript number: 1783153471208744

Reviewer 1 / Brandy Roane

Major Compulsory Revisions

1. More information is needed on the sample used for these analyses. The authors reported a sample size of 218 in the larger study with a sub-sample of 166 included in these analyses; however, exclusion criteria were not clear. Were these exclusions applied to the larger study or the subsample? If it was applied to the larger sample and not the sub-sample, why were 52 participants excluded from the sub-sample?

   a. If it was the result of attrition for their exclusion, data on those that did not continue should be included. What did their measures look like in comparison to those who completed? This is particularly relevant given the conclusion the authors have drawn from their data indicating that “sleep disturbances at baseline do not necessary lead to poorer clinical outcome during follow-up.”

   b. If attrition was an exclusion for participants to be included in the final larger sample, this should still be listed as a limitation in the results. In addition, it really tempers the conclusion as the data cannot speak to this without knowing if those that did not complete showed more severe baseline measures than those who did complete. It’s possible that for the participants who did complete treatment, sleep disturbance at baseline did not predict worse outcomes; however, it is also just as possible that sleep disturbance at baseline predicts completion with those who exhibit more severe sleep disturbances also are those more likely to not complete treatment.

We thank the reviewer for this important comment. In the first version of the manuscript, the description of the selection process was clearly insufficient, and a more detailed description has now been added. (lines 108-112)

The parent sample (n=218) included adolescents with any DSM-IV depressive mood disorder, while the sub-sample presented in this paper consisted of subjects with major depressive disorder (MDD). The exclusion criteria mentioned in the first manuscript version (mental retardation, age under 13 or over 19 years, or insufficient knowledge of the Finnish language) applied to the whole parent sample.
We were specifically interested in the relationship between unipolar MDD and sleep, and therefore subjects with any hypomanic or manic episodes in the past (diagnosed as bipolar disorder, n=15), subjects with depression not otherwise specified (NOS, n=13), subjects with dysthymia (n=10), subjects with adjustment disorder with depressed mood (n=2), and subjects with mood disorder due to a general medical condition (n=1) were excluded from the analyses presented in this paper. In addition, 8 subjects with a unipolar depressive mood disorder in full remission were excluded, and 3 subjects were excluded from the analyses because of missing/invalid data in baseline. (= a total of 52 excluded subjects)

The following text was added to the methods: “As we were interested specifically in the relationship between unipolar MDD and sleep, subjects with any other depressive mood disorder diagnosis than unipolar MDD (e.g. bipolar disorder, dysthymia) were excluded from the analyses (n=41). Further, subjects in full remission already in baseline (n=8) and subjects with missing/inadequate data in baseline (n=3) were excluded, leaving n=166 adolescents in the analyses presented in this paper.” (lines 108-112)

Attrition rate did not affect inclusion/exclusion in the presented subsample, as described above. Please see also response to comment 2 below for more detailed information on attrition.

2. Overall, more specific information is needed on attrition. This is really not touched on until the discussion, but no data is included with regards to rates of attrition. The authors starting on line 307 discussing attrition. They noted that there “was a slight but no major difference in attrition rates in favour of adolescents with the least sleep complaints,” but what does this mean? Was it a significant difference? Did they examine statistical significance?

More detailed information is now given on attrition already in the methods and results sections as follows:

The following text is included in the methods (lines 136-138): “A total of n=24 adolescents did not attend the 1-year interview. The adolescents who did not attend the 1-year interview, did not show any statistically significant differences in clinical or sociodemographic characteristics at baseline as compared to the adolescents (n=142) who completed the 1-year interview.”

More specifically, the mean BDI-21 score (18.5 vs. 17.8; independent samples t-test n.s.), mean BAI score (26.0 vs. 23.1; independent samples t-test n.s.), mean GAF score (48.8 vs. 50.8; independent samples t-test n.s.), mean age (16.9 vs. 16.5; independent samples t-test n.s.), or gender (boy 20 % vs. 13.2 %; χ² test n.s.) at baseline were not statistically different when comparing the adolescents who dropped out to the rest of the group.

Concerning sleep symptoms, the following text is included in the results (lines 270-273): “In subgroup A (no/minor sleep complaints) n=5 (9.8 %) adolescents dropped out, while the numbers were slightly, but not significantly higher in the two other subgroups (subgroup B n=10 (14.5 %), subgroup C n=9 (20.0 %)). (χ² tests n.s.)”

3. Clearer presentation of the variables being examined is needed (i.e., clearly identify the variables and define them). At present, all but 2 of the variables are contained within one paragraph. Brevity is great, but not at the cost of clarity.
Attention has been paid to how the variables are presented and more information is given (see also response to comment 4 below) (see Methods/Assessment of psychiatric symptoms, lines 148-181). Many of the variables used are, however, very commonly used and consequently self-explanatory for most readers.

4. Similarly, rationales need to be provided for the variables. For instance, why was GAF scores used for global functioning rather than a more objective (and reliable) assessment? Why were the sleep variables excluded from the BDI? While this is a reasonable call, why was it done? What support do you have for doing it? Excluding items from an assessment can impact the integrity of the assessment, whether the assessment is a validated one or not. Thus, it would also be very beneficial to report on the reliability of the measures in your sample (e.g., listing Cronbach’s alphas for your specific sample).

The GAF scores (Axis V) were defined as part of the DSM-IV axial diagnostic procedure (American Psychiatric Association; Diagnostic and Statistical Manual of Mental Disorders, 4th edn (DSM–IV); American Psychiatric Press, 1994). We decided to use the GAF, since it is a widely known and used method for assessing overall psychosocial functioning, and it has been used in a wide range of studies also in adolescent populations, e.g. in the following studies:


A comment on the validity of GAF in adolescents has been added to the text (lines 160-162).

The sleep item was excluded from the BDI-21 because we were looking at the relationship between sleep and depression. If the sleep variable would have been left in, we would expect to see correlation between sleep symptoms severity and depression severity (“sleep correlates to sleep). The BDI-21 score without the sleep items reflects the severity of depressive symptoms other than sleep.

The Cronbach’s alpha for BDI-21 at baseline showed excellent consistency both when the sleep item was taken into account (21 items), and when the sleep item was left out (20 items) (Cronbach’s alpha = 0.912; identical in both cases). This is now mentioned in the text (lines 173-176).

5. The results section is confusing at times, particularly the interaction section. A table would greatly help with presenting and discussing the data.

Based on the reviewer’s comment, a table (Table 1) has been added to clarify the interaction section.
6. In the discussion (lines 272-275), the authors indicated that “these findings highlight the tight link between sleep complaints and depressive state and support the view of sleep complaints as state-dependent features of adolescent depression, as opposed to sleep disturbances representing persistent vulnerability traits” is an overstatement of the findings. This type of statement requires a more rigorous hypothesis testing and likely more data points. Like depression, sleep disturbances can remit and reappear over time. At the very least an additional assessment point (say another year later) would need to be included or data on how sleep changed throughout the course of treatment.

Based on the reviewer’s comment, the mentioned sentence has now been more carefully formulated to avoid overstatement: “These findings support the tight link between sleep complaints and depressive state.” (lines 308-309)

7. Similar to the above point, the authors note the limitation of the use of sleep items on a non-sleep specific measure to determine the presence of a sleep complaint; however, their conclusions seem to forget this limitation (lines 301-305). Specifically, their conclusions exclude the fact that their measure of sleep disturbance is based on a single time-point measure at baseline and at follow-up as opposed to more frequent assessment. The questions are from a self-report questionnaire that can be influenced by mood state and is plagued with retrospective recall issues. The decrease in sleep complaints may not be due to an actual decrease in symptoms, but due to an improvement in mood. The actual change cannot be sufficiently determined with the limited assessment of symptoms.

We agree with the reviewer that mood may influence how the subject perceives his/her sleep. Most sleep questions (all except nightmares) were from a semi-structured clinical interview (K-SADS-PL) by trained mental health professional (i.e. not from a self-report questionnaire). Self-report questionnaires and interviews may be problematic in assessing sleep disturbances, since they do not give us any objective values of sleep. However, they are the optimal way of getting information on the patient’s own experience of sleep. The relationship between mood and sleep in depressed adolescents needs to be studied further with a combination of subjective and objective measurements in longitudinal studies.

The use of the K-SADS-PL to assess sleep disturbances is not the ideal approach to assess sleep disturbances, but it has been successfully used in previous studies, and it may provide valuable information on some aspects of sleep.

We would also like to emphasize the naturalistic nature of this study. The K-SADS-PL is a typical way of interviewing adolescents in clinical settings at the beginning of treatment. One of the main questions we wanted to answer with this study was, does the presence and amount of baseline sleep complaints affect depression 1-year clinical outcome, i.e. are the baseline sleep complaints valuable in predicting clinical outcome.

The sometimes problematic use of self-report questionnaires and the use of the K-SADS-PL interview to assess sleep symptoms have now been discussed and acknowledged as a major limitation of the study (lines 329-333 and 380-383).

More attention has been paid to careful wording of the general conclusions. (lines 385-392)
The need for larger longitudinal studies combining both subjective and measures of sleep in depressed adolescents to clarify the link between sleep and depression in adolescents in further detail is also acknowledged in the conclusions (lines 389-391).

8. Later in the discussion (lines 314-321) the authors again seem to ignore previously reported fluctuations in sleep complaints and the very complex relationship between sleep and mood. One possible alternative conclusion given the data is that the sleep disturbance may have been driving the depression for some of the teens. The successful resolution of the sleep disturbance subsequently resulted in a reduction in depression symptoms. The treatments used, particularly the SSRIs, are prescribed for patients with insomnia. Thus, the SSRI, while prescribed to target the depression, may have actually successfully managed the sleep disturbance, and, thereby, the depression improved.

We agree with the reviewer that the resolution of the sleep disturbance may have resulted in a reduction of depression symptoms. This possibility has been added to the discussion (lines 374-376):

“The relationship between sleep and mood is complex and bidirectional. One possibility is, that at least for some of the adolescents the sleep disturbance may have been the driving force for depression, and the resolution of sleep problems may have also resulted in a reduction of depression symptoms.”

The effects of SSRI (and other) medication on adolescents’ sleep remain, however, controversial. An additional paragraph on the possible impact of medication on sleep has now been added to the discussion (lines 352-366):

“Depressed patients with sleep disturbances may respond differently to both pharmacological and other treatment of depression compared to those without sleep problems [11,50-52]. The more detailed findings remain, however, mixed, and research evidence is lacking. Antidepressant drugs generally tend to ameliorate the sleep impairments in depression (e.g. they inhibit REM sleep) [53], and especially patients with reduced REM sleep latency may respond favourably to antidepressants [54]. In a preliminary study, however, fluoxetine had a negative impact on sleep in depressed children and adolescents [55]. Depressed adolescents reporting insomnia have been observed to be less likely to respond to antidepressant treatment than those without insomnia [52]. In addition, depressed adolescents receiving medication for sleep have been found to respond less likely to depression treatment than those without sleep medication [51]. The low response rate of adolescents treated with sleep medication has been particularly associated with the use of trazodone (NB trazodone was not in use among our study sample), pointing towards drug-specific effects. [51] Further, the wide range of sleep disturbances (e.g. insomnia vs. hypersomnia) may interact with various depression treatments differently. Although the subgroups in our sample did not differ in terms of medication used, we can not rule out the possibility that individuals may have responded differently to prescribed medications.”

Minor Essential Revisions

9. Consistent naming of variables throughout the manuscript. If you are reporting on symptoms and not actual diagnosis, establish the term you will use to refer to it and stick with that term. For instance, the authors referred to anxiety symptoms as “anxiety levels,” “anxiety,” and “anxiety symptom severity.” These do not all mean the same thing.

Based on the reviewer’s comment, we have now paid attention on consistent naming of the variables.
10. A rationale is needed for why sleep disturbances was categorized the way the authors chose to. The division appeared very arbitrary. Why was a complaint of non-restorative sleep not considered a “complaint of sleep disturbance,” but instead was a “minor sleep complaint” that was grouped with “no complaint”?

We wanted to divide the total group to subgroups based on the severity of sleep complaints. As no actual sleep scale was available, an arbitrary division based on the K-SADS items was used.

The complaint of non-restorative sleep is an extremely common complaint (69% of adolescents in our sample in baseline), and it was considered as the mildest, most unspecific and most subjective of the sleep complaints assessed. All other symptoms assessed involve more objective characteristics, i.e. changes in length/timing of sleep/continuity of sleep.

In general population-based adolescent samples, non-restorative sleep or tiredness in the morning are very common sleep complaints. For example, in the Finnish part of the Health Behaviour in School-Aged Children (HBSC) Study, a WHO Collaborative Cross-National Study, over 40 % of the Finnish 15-year-olds report tiredness when waking up in the morning at least 4 times a week, pointing towards complaints of non-restorative sleep (Kronholm E et al, submitted data). This might reflect the tendency of adolescents to experience sleepiness even after a good night’s sleep.

As the division to subgroups was arbitrary and raised some concerns among the reviewers, we performed the analyses also with an alternative way to divide the group to sleep subgroups. We used the total sum of all the six K-SADS sleep item scores at baseline (including also the item on non-restorative sleep; 1-3 points per item, minimum 6 points, maximum 18 points; missing answers were substituted with the average points of all items) and the groups were formed as follows: a total sum of 6-9 points (mild sleep complaints; n=52), 10-14 points (moderate sleep complaints; n=99), and 15-18 points (major sleep complaints; n=15). By using these alternative subgroups, there were no changes in the main results: the more the adolescent had sleep complaints, the quicker was the clinical improvement. This has now been indicated in the text (lines 278-285).

11. Lines 220-221 state findings that are not supported with statistics. According to what analysis was age associated with suicidality/self-harm symptoms and sex associated with BDI scores?

The statistics have now been added to the text, and the order of presentation in the text has been changed for clearer presentation. (lines 250-252)

12. The first two figures appear somewhat redundant and could be combined.

We thank the reviewer for this suggestion. We decided to combine the two figures (figure 1 + figure 2 = figure 1; figure 3-> figure 2).
Discretionary Revisions

13. The manuscript overall reads well, but a few grammar/punctuation edits would be useful. Professional language editing is available at our institution if needed. In case language revision would be required for publication, it is possible, but it will take some additional time.

14. The terms “subgroup A,” “subgroup B,” and “subgroup C” is confusing once you get past the designation. As indicated above, a table with data would help with presenting the findings. It would likely also help with cementing these terms throughout the manuscript.

A table has now been added, which will hopefully help with the terms.
Major Compulsory Revisions:

1. Lines 105-107: please indicate if there were any sociodemographic or clinical differences between subjects who were included in this study (n=166) and subjects from the parent study who were not (n=52). Similarly, were there any differences between subjects who attended the one-year follow-up and those who did not.

We thank the reviewer for this comment. In the first version of the manuscript, the description of the selection process was clearly insufficient, and a more detailed description has now been added. (lines 108-112, please see also response to reviewer 1, comments 1 and 2) The subjects included in this study (n=166) were diagnosed with major depressive disorder (MDD), and therefore subjects with any hypomanic or manic episodes in the past (diagnosed as bipolar disorder, n=15), subjects with depression not otherwise specified (NOS, n=13), subjects with dysthymia (n=10), subjects with adjustment disorder with depressed mood (n=2), and subjects with mood disorder due to a general medical condition (n=1) were excluded from the analyses presented in this paper. In addition, 8 subjects with a unipolar depressive mood disorder in full remission were excluded, and 3 subjects were excluded from the analyses because of missing/invalid data in baseline. (= a total of 52 excluded subjects)

This has now been clarified in the text as follows:

“As we were interested specifically in the relationship between unipolar MDD and sleep, subjects with any other depressive mood disorder diagnosis than unipolar MDD (e.g. bipolar disorder, dysthymia) were excluded from the analyses (n=41). Further, subjects in full remission already in baseline (n=8) and subjects with missing/inadequate data in baseline (n=3) were excluded, leaving n=166 adolescents in the analyses presented in this paper.” (lines 108-112)

There were no statistically significant differences between the subjects who attended the one-year follow-up and those who did not. The following text has been added:

“A total of n=24 adolescents did not attend the 1-year interview. The n=24 adolescents who did not attend the 1-year interview, did not show any statistically significant differences in clinical or sociodemographic characteristics at baseline as compared to the adolescents (n=142) who completed the 1-year interview.” (lines 136-138)

More specifically, the mean BDI-21 score (18.5 vs. 17.8; independent samples t-test n.s.), mean BAI score (26.0 vs. 23.1; independent samples t-test n.s.), mean GAF score (48.8 vs. 50.8; independent samples t-test n.s.), mean age (16.9 vs. 16.5; independent samples t-test n.s.), or gender (boy 20 % vs. 13.2 %; $\chi^2$ test n.s.) at baseline were not statistically different when comparing the adolescents who dropped out to the rest of the group.

Concerning sleep symptoms, the following text is also included in the results (lines 270-273): “In subgroup A (no/minor sleep complaints) n=5 (9.8 %) adolescents dropped out, while the numbers were slightly, but not significantly higher in the two other subgroups (subgroup B n=10 (14.5 %), subgroup C n=9 (20.0 %)). ($\chi^2$ tests n.s.)”
2. Methods: There are major concerns with the primary study measures. The authors appropriately acknowledge the limitations of using the K-SADS-PL to assess sleep symptoms, but the validity of the nightmare measure is also unclear and the primary outcomes were either collected using non-validated measures (suicidality/self-harm, psychosocial functioning) or with scales that have unclear psychometric properties in adolescents (BDI, BAI). The authors should provide reliability and validity statistics for these measures. In addition, for the measures that were obtained as part of a clinical interview, information on rater reliability (assuming different raters were involved) should be provided.

The study was not primarily designed to study sleep disturbances, but mood disorders. Thus, it did not include any objective characterization of sleep or any sleep-specific questionnaires. This is the major limitation of the study, and it has now been emphasized in the discussion (lines 329-333 and 380-382).

Based on the reviewers’ comments (please see also response to reviewer 1 comments 3-4), the methods section on assessment of psychiatric symptoms has been reformulated and more detailed description of the variables is now provided (lines 146-181).

The general health questionnaire (GHQ) is not an ideal or validated means of assessing nightmares, but it was the only available means to get any information on nightmares in our sample. We were specifically interested in nightmares, since nightmares have been linked to suicidality in previous studies, e.g.:


Because of its limitations, however, the nightmare measure was not used for more extensive analyses (e.g. in forming the sleep subgroups).

The GAF scores (Axis V) were defined as part of the DSM-IV axial diagnostic procedure (American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (4th edn) (DSM–IV), American Psychiatric Press, 1994). GAF is a widely known and used method for assessing overall psychosocial functioning. It has been used in a wide range of studies also in adolescent populations, e.g. in the following studies:


The Beck Anxiety Inventory (BAI) has shown acceptable psychometric properties in adolescents:


The same applies to the Beck Depression Inventory (BDI-21):


Concerning the clinical interview, all the research diagnoses were confirmed in a diagnostic meeting. Interrater reliability was assessed using 15 randomly selected videotaped interviews, and it was good for mood disorder diagnoses. (Karlsson L et al. 2007) Inter-rater reliability is now commented in the text (lines 151-154).

3. Lines 158-160: It is not entirely clear how remission and recurrence were determined. Were these categorizations based on clinical interview or K-SADS-PL score? If the latter, please indicate what score(s) was/were used. If the former, please indicate how reliability of these categorizations was determined.

As mentioned on lines 150-154, remission was defined according to the DSM-IV criteria as no significant signs or symptoms of MDD present during at least the past two months. Recurrence was defined as remission from the baseline MDD episode and onset of a new MDD episode. These categorizations were based on the clinical K-SADS-PL interview score, in which items scored 3 were considered as significant. The following references have been added:

4. Lines 213-219: No rationale is presented for the subgroups defined according to sleep complaints. For example, why is non-restorative sleep included in the no/minor sleep complaints subgroup, particularly since this was the most common residual sleep complaint? Why were nightmares excluded from the subgroup definitions? Is it possible that the more rapid improvement in outcome measures for those with sleep disturbances at baseline is because non-restorative sleep was included as a no/minor sleep complaint? A more convincing rationale for the subgroup determinations is needed.

We wanted to divide the total group to subgroups based on the severity of sleep complaints. As no actual sleep scale was available, an arbitrary division based on the K-SADS items was used.

Nightmares were not measured as part of the K-SADS-PL interview, but a self-report questionnaire (GHQ), which was considered an additional measure. The results on nightmares were reported mainly because previous studies have shown important correlations between suicidality and nightmares (see also response to comment 2).

The complaint of non-restorative sleep is an extremely common complaint (69% of adolescents in our sample in baseline), and it was considered as the mildest, most unspecific and most subjective of the sleep complaints assessed. All other symptoms assessed involve more objective characteristics, i.e. changes in length/timing of sleep/continuity of sleep.

In general population-based adolescent samples, non-restorative sleep or tiredness in the morning are very common sleep complaints. For example, in the Finnish part of the Health Behaviour in School-Aged Children (HBSC) Study, a WHO Collaborative Cross-National Study, over 40% of the Finnish 15-year-olds report tiredness when waking up in the morning at least 4 times a week, pointing towards complaints of non-restorative sleep (Kronholm E et al, submitted data). This might reflect the tendency of adolescents to experience sleepiness even after a good night’s sleep.

As the division to subgroups was arbitrary and raised some concerns among the reviewers, we performed the analyses also with an alternative way to divide the group to sleep subgroups. We used the total sum of all the six K-SADS sleep item scores at baseline (including also the item on non-restorative sleep; 1-3 points per item, minimum 6 points, maximum 18 points; missing answers were substituted with the average points of all items) and the groups were formed as follows: a total sum of 6-9 points (mild sleep complaints; n=52), 10-14 points (moderate sleep complaints; n=99), and 15-18 points (major sleep complaints; n=15). By using these alternative subgroups, there were no changes in the main results: the more the adolescent had sleep complaints, the quicker was the clinical improvement. This has now been indicated in the text (lines 278-285).
5. Lines 241-248: No information is presented on any sleep-related treatment that might have been received by subjects during the one-year follow-up period. Is it possible that the reason that the trajectory of improvement was better for subjects with sleep disturbances at baseline was related to treatment for these sleep disturbances during follow-up?

There were no statistically significant differences in treatments received by the subgroups. All psychotropic medications (also those not primarily targeted on sleep) may affect adolescents’ sleep, and the effects may be complex.

An additional paragraph on the possible impact of medication on sleep has now been added to the discussion (lines 352-366):

“Depressed patients with sleep disturbances may respond differently to both pharmacological and other treatment of depression compared to those without sleep problems [11,48-50]. The more detailed findings remain, however, mixed, and research evidence is lacking. Antidepressant drugs generally tend to ameliorate the sleep impairments in depression (e.g. they inhibit REM sleep) [51], and especially patients with reduced REM sleep latency may respond favourably to antidepressants [52]. In a preliminary study, however, fluoxetine had a negative impact on sleep in depressed children and adolescents (Armitage et al., 1997). Depressed adolescents reporting insomnia have been observed to be less likely to respond to antidepressant treatment than those without insomnia [50]. In addition, depressed adolescents receiving medication for sleep have been found to respond less likely to depression treatment than those without sleep medication [49]. The low response rate of adolescents treated with sleep medication has been particularly associated with the use of trazodone (NB trazodone was not in use among our study sample), pointing towards drug-specific effects. [49] Further, the wide range of sleep disturbances (e.g. insomnia vs. hypersomnia) may interact with various depression treatments differently. In our sample, the subgroups did not differ in terms of medication used, but we can not rule out the possibility that individuals would have responded differently to prescribed medications. “

6. Lines 301-305: The authors offer only limited discussion to explain why sleep disturbances at baseline would relate to a more favorable discussion. How do the authors reconcile these findings with studies from the adult depression literature indicating that baseline sleep disturbances adversely affect clinical course?

Our findings were unexpected based on previous adult depression literature. The relationship between sleep and depression may be different in different developmental stages and age groups. The inconsistency in findings may be linked with e.g. maturational factors and factors related to the stage of illness. The discussion has been extended and different possible explanations have been discussed. (lines 337-341) The publication bias (in favor of positive findings) may also be one potential explanation, and the topic clearly needs to be studied more extensively in the future.

Minor Essential Revisions:

1. Lines 115-116: the severity categorization of depression adds up to 101% of the sample. Please reconcile.

The percentages have now been corrected. The percentage of moderate depression was erroneously 21 % (20.48 %, erroneously rounded to 21%) -> corrected now to 20 %.
2. Lines 132-134: Please use consistent terminology. Here, the authors use the terms “recovered and relapse” to describe the status of one group, but later (lines 158-160) they report only definitions of “remission and recurrence.”

In the new version of the manuscript, more attention has been paid to consistent terminology.

3. Line 234: Figure 3 is related to the BDI not psychosocial findings.

The erroneous reference to figure 3 has now been omitted.

4. Lines 258-260: The authors indicate that there were baseline differences in depression severity among the three sleep subgroups. Was this factor controlled for in the GEE analysis?

Differences in depression severity at baseline were not controlled in the presented GEE analyses. We have, however, performed additional analyses controlling for the baseline values of the outcomes. All reported results concerning the effects of sleep complaints on clinical outcomes (main effects and interactions) remained significant after adjusting for the baseline values. This is now mentioned in the text (lines 264-266).

Discretionary Revisions:

1. Figure 3: The Loess curves and data points are difficult to distinguish. Please consider using color or some other way to more easily differentiate the three groups and trajectory lines.

Based on the reviewer’s comment colors have now been added to figure 3 (i.e. new figure 2) to make the data points and Loess curves more easily distinguishable.