# PEER REVIEW HISTORY

BMJ Paediatrics Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

## ARTICLE DETAILS

| TITLE (PROVISIONAL) | Parent-reported early sleep problems and internalizing, externalizing, and dysregulation symptoms in toddlers |
|---|---|
| AUTHORS | Morales Muñoz, Isabel; Lemola, Sakari; Saarenpaa-Heikkila, Outi; Kylliainen, Anneli; Polkki, Pirjo; Paunio, Tiina; Broome, Matthew; Paavonen, Juulia |

## VERSION 1 – REVIEW

| REVIEWER | Reviewer name: Sarah J Nevitt<br>Institution and Country: University of Liverpool United Kingdom<br>Competing interests: I have no competing interests |
|---|---|
| REVIEW RETURNED | 05-Dec-2019 |

| GENERAL COMMENTS | I have conducted a statistical review of the manuscript "Early sleep problems and internalizing, externalizing, and dysregulation symptoms in toddlerhood"<br><br>The authors assess the association of sleep related variables (time asleep and sleep quality) of infants to emotional and behavioural symptoms in toddlers aged 24 months.<br>I have one major comment regarding the analysis approach and interpretation of the results as I am confused regarding the time points and how the multiple time points are included within analysis.<br><br>Have I understood correctly that the outcomes (BITSEA subscales) are measured only at 24 months?<br><br>And that the sleep variables measured at 3, 8, 12 and 24 months are included within separate models? i.e. BITSEA subscale at 24 months with sleep variables at 3 months + covariates, BITSEA subscale at 24 months with sleep variables at 8 months + covariates, BITSEA subscale at 24 months with sleep variables at 12 months + covariates, BITSEA subscale at 24 months with sleep variables at 24 months + covariates.<br><br>The authors make several references in the results and discussion to sleep variables at 3, 8 and 18 months being 'longitudinally related' to the outcomes. This is not a conclusion which can be drawn based on the statistical methods which the authors have used (from my understanding).<br><br>If separate models for the time points have been applied, then it can be inferred that the sleep variables at each of these time variables (independently) are related to the outcomes but to infer a longitudinal relationship (i.e. a relationship over the different time points) would require a longitudinal or repeated measures analysis (e.g. a mixed-effects linear regression).<br><br>Could the authors please clarify the methodology used and reword interpretations relating to 'longitudinal' if longitudinal methods are not used? |
|---|---|

I also have a few minor comments

1) Page 5: "The study is based on the CHILD-SLEEP birth-cohort (14)"
I see that a reference for this cohort has been given but for the purposes of this manuscript, a little more information would be helpful. Could a couple of sentences be added about this cohort?

2) Page 6: Covariates: the covariates selected were: child's age at the measurement time point of 24 months and sex, and prenatal maternal health, age and education, and gestational age at birth. "

Please give some further details – were age covariates measured continuously or in categories? What were the categories for health / education etc.?

3) Page 6: "due to the large sample size, parametric statistical testing was considered appropriate."

While I agree with this in principle, the impact of the data being skewed depends on how skewed it is. Could some summary measures of the outcome variables be provided at baseline and 24 months? e.g. mean, median, range? This can give an indication of how skewed the data actually is and therefore the potential impact on analysis

Were the variables also tested for skewness? Linear regression assumes that both the outcome variable and the independent variables are normally distributed

4) Page 7: "Patient representatives were involved in this follow-up study."
Please give further details – were patient representatives involved in the design of the study? The interpretation of the results etc.?

5) Page 7 (Results): "Sociodemographic variables are described in Table 1."
Could a sentence or two be added about these variables?

6) Table 1: I found the top half of this table a little confusing (e.g. age in weeks, with four measurements in months directly below didn't make sense at first).
I suggest that the top half of this table should be formatted like the bottom i.e. with the time points as the column headings and the various summary statistics of age, sex, weight etc. at each time point summarised.

7) Within Table 2, it would also be helpful to present the regression coefficients for the covariates at the different time points – particularly as some of these regression coefficients are mentioned within the results text

| REVIEWER | Reviewer name: Serge Brand<br>Institution and Country: Kermanshah University of Medical Sciences (KUMS), Sleep Disorders Research Center, Kermanshah, Iran<br>Competing interests: No competing interests |
| --- | --- |
| REVIEW RETURNED | 02-Jan-2020 |

| GENERAL COMMENTS | The authors presented a very well written, interesting and very informative manuscript. The reference list is very timely; the authors might consider that the prevalence rate of infantile colic (IC) is about 5%; thus, from a statistical point of view, about 47 infants with IC |
| --- | --- |

and their parents took part in the study; it further follows that this sub-sample could have biased the pattern of results (Brand et al., 2011, 2014). Please check once again ref# 14; it appears that 'Juulia' is a first name.

Title: well-chosen
Abstract: very well written
Introduction; very well written; see above as regards infants with IC.
Methods:
The study design is nicely described; the samples and the tools are well reported. As regards the attrition and the characteristics of those mothers completing also later assessment points, please report all statistical indices: first, reporting p-values without further statistical indices is somehow pointless; second, reporting just p-values is highly discouraged (Wasserstein et al., 2019); third, it is conceivable that the large sample size conferred to 'significant' p-values; fourth, it follows that effect sizes should be reported, too.
SPSS should be: SPSS® version 25.0 (IBM Corporation, Armonk NY, USA) for Windows®/Apple Mac®
"….the associations between sleep and BITSEA,…"; please replace the acronym BITSEA with the appropriate dimension; actually, the authors associated dimensions of sleep with dimensions of social, emotional and behavioral problems, and not with an acronym. Accordingly, replace the acronym also in the text.
Results: well-reported.
Discussion; in my opinion, the authors discussed very well their data.

References
Brand, S., Furlano, R., Sidler, M., Schulz, J., Holsboer-Trachsler, E., 2011. 'Oh, baby, please don't cry!': in infants suffering from infantile colic hypothalamic-pituitary-adrenocortical axis activity is related to poor sleep and increased crying intensity. Neuropsychobiology 64(1), 15-23.
Brand, S., Furlano, R., Sidler, M., Schulz, J., Holsboer-Trachsler, E., 2014. Associations between infants' crying, sleep and cortisol secretion and mother's sleep and well-being. Neuropsychobiology 69(1), 39-51.
Wasserstein, R.L., Schirm, A.L., Lazar, N.A., 2019. Moving to a World Beyond "p < 0.05". The American Statistician 73(sup1), 1-19.

## VERSION 1 – AUTHOR RESPONSE

Reviewer # 1

Comment 1: I have conducted a statistical review of the manuscript "Early sleep problems and internalizing, externalizing, and dysregulation symptoms in toddlerhood." The authors assess the association of sleep related variables (time asleep and sleep quality) of infants to emotional and behavioural symptoms in toddlers aged 24 months. I have one major comment regarding the analysis approach and interpretation of the results as I am confused regarding the time points and how the multiple time points are included within analysis. Have I understood correctly that the outcomes (BITSEA subscales) are measured only at 24 months? And that the sleep variables measured at 3, 8, 12 and 24 months are included within separate models? i.e. BITSEA subscale at 24 months with sleep variables at 3 months + covariates, BITSEA subscale at 24 months with sleep variables at 8 months + covariates, BITSEA subscale at 24 months with sleep variables at 12 months + covariates, BITSEA subscale at 24 months with sleep variables at 24 months + covariates.

Reply 1: Thank you for this question. This was indeed the case. First, as the reviewer mentions, BITSEA subscales were only measured at 24 months, and this scale was not available at earlier stages (3, 8 and/or 18 months).

Second, and as the reviewer acknowledges, the sleep variables were measured at 3, 8, 18 and 24 months included in separate models: for example, sleep duration at 3 months, sleep-onset latency at 3 months, number of night wakings at 3months, proportion of daytime sleep at 3 months, and bedtime at 3 months were all included within the same model as independent variables, together with the covariates, and BITSEA internalizing, for example, as dependent variable. We have now clarified this in the Statistical Analyses section: "To do this, all the sleep variables from the same time point were included within the same regression model (e.g. total sleep duration , sleep-onset latency, night awakenings, proportion of daytime sleep and bedtime were all included within the same model), together with the covariates. Each time point (i.e., three, eight, 18 and 24 months) was treated in separate regression models" (page 7).

C2: The authors make several references in the results and discussion to sleep variables at 3, 8 and 18 months being 'longitudinally related' to the outcomes. This is not a conclusion which can be drawn based on the statistical methods which the authors have used (from my understanding). If separate models for the time points have been applied, then it can be inferred that the sleep variables at each of these time variables (independently) are related to the outcomes but to infer a longitudinal relationship (i.e. a relationship over the different time points) would require a longitudinal or repeated measures analysis (e.g. a mixed-effects linear regression). Could the authors please clarify the methodology used and reword interpretations relating to 'longitudinal' if longitudinal methods are not used?

R2: Thank you for this note. We used the term "longitudinal" to refer to an association over time (so as to contrast to cross-sectional findings), but we have now revised the terminology, so that we now refer to "prospective associations or relationships" (instead of longitudinal associations). In this way we would like to emphasize that this cohort is truly a follow-up sample of same children and therefore sleeping problems predict future emotional problems (although causality cannot be confirmed on the basis of an observatory study).This term has now been changed throughout the whole text.

C3: I also have a few minor comments; 1) Page 5: "The study is based on the CHILD-SLEEP birth-cohort (14)." I see that a reference for this cohort has been given but for the purposes of this manuscript, a little more information would be helpful. Could a couple of sentences be added about this cohort?

R3: We appreciate this comment. We have now provided the following description of the CHILD-SLEEP birth cohort, as follows: "The study is based on the CHILD-SLEEP birth-cohort, comprising all Finnish-speaking pregnant women who belonged to the area of the Tampere University Hospital, in the hospital district of Pirkanmaa, in southern Finland within the recruitment window (April 2011-January 2013). The recruitment took place during a routine visit to the maternity clinics in the 32nd week of pregnancy. Maternity care in Finland is accessible and free to everybody living in Finland and widely used (over 99.7% of women). The coverage in this study was about 29% of pregnant women in the area at the time of recruitment, and thus our sample was representative concerning age and number of previous children, but participants in this study had higher educational level. Other details of the recruitment procedure have been reported previously"[16] (see page 5).

C4: 2) Page 6: Covariates: the covariates selected were: child's age at the measurement time point of 24 months and sex, and prenatal maternal health, age and education, and gestational age at birth. "Please give some further details – were age covariates measured continuously or in categories? What were the categories for health / education etc.?

R4: We acknowledge this comment from the reviewer. In the revised version, we have now provided the additional information concerning the covariates, as follows: "Based on previous research,[7] the covariates selected were: child's age (in years) at 24 months and sex (female vs. male), and prenatal maternal health (generally healthy vs. some health related problem or disability), age (in years) and the highest education level (primary, secondary or high), and gestational age at birth (in weeks). Prevalence rates and descriptive values of these covariates are reported in Table 1 and 2" (page 6).

To note, sex, maternal health and maternal education were categorical variables, while child´s age at 24 months, maternal age and gestational age were continuous.

C5: 3) Page 6: "due to the large sample size, parametric statistical testing was considered appropriate." While I agree with this in principle, the impact of the data being skewed depends on how skewed it is. Could some summary measures of the outcome variables be provided at baseline and 24 months? e.g. mean, median, range? This can give an indication of how skewed the data actually is and therefore the potential impact on analysis. Were the variables also tested for skewness? Linear regression assumes that both the outcome variable and the independent variables are normally distributed

R5: We understand this concern. Therefore, we have now provided in Table 1 the median and the minimum and maximum values (in addition to means and SD) for all the sleep variables and for the three outcomes. Further, the predictors, except for sleep duration and proportion of daytime sleep were also skewed, and this has now been added: "The three outcome variables and the explanatory variables, except for sleep duration and proportion of daytime sleep were skewed, but due to the large sample size, parametric statistical testing was considered appropriate.[20] However, due to the skeweness of the above mentioned variables , we did sensitivity analyses after normality transformations of the skewed variables. Because the main findings were virtually unchanged, we present only the original results" (page 7).

The reviewer can find the new results after conducting regression analyses with the transformed variables (predictors and outcomes) below.

Title: Linear regression models where transformed variables were used.

| | Internalizing symptoms | | | Dysregulation symptoms | | | Externalizing symptoms | | |
|---|---|---|---|---|---|---|---|---|---|
| | β | *p* | 95% C.I. for B | β | *p* | 95% C.I. for B | β | *p* | 95% C.I. for B |
| **3months** | | | | | | | | | |
| Total sleep | -0.093 | 0.018 | -0.066 to -0.006 | -0.065 | 0.058 | -0.070 to -0.004 | -0.031 | 0.430 | -0.043 to 0.018 |
| Sleep-onset latency | 0.069 | 0.070 | -0.008 to 0.201 | 0.129 | 0.001 | 0.086 to 0.313 | 0.017 | 0.649 | -0.082 to 0.131 |
| Number night waking | 0.106 | 0.004 | 0.037 to 0.196 | 0.122 | 0.001 | 0.060 to 0.234 | 0.097 | 0.040 | 0.006 to 0.156 |
| Proportion daytime sleep | 0.025 | 0.550 | -0.058 to 0.110 | 0.044 | 0.291 | -0.042 to 0.141 | 0.005 | 0.906 | -0.081 to 0.091 |
| Bedtime | -0.079 | 0.070 | -0.787 to 0.031 | 0.003 | 0.944 | -0.430 to 0.461 | 0.078 | 0.076 | -0.040 to 0.793 |
| **8 months** | | | | | | | | | |
| Total sleep | -0.080 | 0.047 | -0.102 to 0.007 | -0.148 | 0.001 | -0.158 to -0.039 | -0.023 | 0.620 | -0.070 to 0.042 |
| Sleep-onset latency | 0.057 | 0.138 | -0.034 to 0.247 | 0.095 | 0.011 | 0.046 to 0.353 | 0.084 | 0.064 | 0.007 to 0.327 |
| Number night waking | 0.074 | 0.048 | 0.001 to 0.171 | 0.174 | <0.001 | 0.133 to 0.319 | 0.006 | 0.862 | -0.096 to 0.080 |
| Proportion daytime sleep | -0.008 | 0.87 | -0.120 to 0.102 | 0.057 | 0.244 | -0.049 to | 0.012 | 0.81 | -0.101 to |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | 5 | | | | 0.193 | | 7 | 0.128 |
| Bedtime | -0.079 | 0.101 | -1.083 to 0.097 | -0.093 | 0.046 | -1.300 to -0.012 | 0.036 | 0.444 | -0.848 to 0.372 |
| **18 months** | | | | | | | | | |
| Total sleep | -0.023 | 0.574 | -0.081 to 0.045 | -0.095 | 0.016 | -0.152 to -0.015 | 0.019 | 0.639 | -0.050 to 0.082 |
| Sleep-onset latency | 0.045 | 0.239 | -0.056 to 0.223 | 0.109 | 0.003 | 0.080 to 0.384 | 0.074 | 0.067 | 0.019 to 0.311 |
| Number night waking | 0.098 | 0.007 | 0.051 to 0.324 | 0.233 | <0.001 | 0.368 to 0.654 | 0.050 | 0.165 | -0.041 to 0.243 |
| Proportion daytime sleep | 0.033 | 0.437 | -0.063 to 0.145 | 0.097 | 0.016 | 0.026 to 0.251 | 0.017 | 0.679 | -0.131 to 0.086 |
| Bedtime | 0.014 | 0.739 | -0.537 to 0.757 | -0.052 | 0.201 | -1.162 to 0.244 | 0.019 | 0.662 | -0.826 to 0.525 |
| **24 months** | | | | | | | | | |
| Total sleep | -0.045 | 0.282 | -0.105 to 0.031 | -0.087 | 0.022 | -0.148 to -0.011 | 0.048 | 0.258 | -0.112 to 0.030 |
| Sleep-onset latency | 0.070 | 0.060 | -0.005 to 0.226 | 0.188 | <0.001 | 0.219 to 0.450 | 0.097 | 0.009 | 0.040 to 0.281 |
| Number night waking | 0.109 | 0.002 | 0.093 to 0.396 | 0.371 | <0.001 | 0.778 to 1.083 | 0.054 | 0.117 | -0.032 to 0.286 |
| Proportion daytime sleep | 0.073 | 0.080 | -0.009 to 0.153 | 0.048 | 0.202 | -0.029 to 0.135 | 0.008 | 0.842 | -0.076 to 0.094 |

| Bedtime | 0.008 | 0.859 | -0.625 to 0.749 | -0.015 | 0.703 | -0.826 to 0.557 | 0.017 | 0.684 | -0.570 to 0.868 |

C6: 4) Page 7: "Patient representatives were involved in this follow-up study." Please give further details – were patient representatives involved in the design of the study? The interpretation of the results etc.?

R6: Thanks for this suggestion. Now this sentence has been slightly modified as follows: "Patient representatives were involved in the design of this follow-up study" (page 8).

C7: 5) Page 7 (Results): "Sociodemographic variables are described in Table 1." Could a sentence or two be added about these variables?

R7: We appreciate this comment, and thus we have now provided further information in the text concerning the results in Table 1, as follows: "Briefly, 47.1% of the sample were girls, the average gestational age at birth was 40 weeks, the average maternal age during pregnancy was 31 years, 76.9% of the mothers reported having good health and 39.0% of the mothers had high educational level" (page 8).

C8: 6) Table 1: I found the top half of this table a little confusing (e.g. age in weeks, with four measurements in months directly below didn't make sense at first). I suggest that the top half of this table should be formatted like the bottom i.e. with the time points as the column headings and the various summary statistics of age, sex, weight etc. at each time point summarized.

R8: We understand this concern and have now presented the age in days as suggested by the reviewer, before sleep variables. The rest of the variables remain the same: sex, child birth weight, gestational age at birth, maternal age during pregnancy, maternal health and maternal educational level, as they refer to one specific time point only. Furthermore, the specific time point of these socio-demographic variables appears specified now. Also, and in order to make this table more comprehensive, we have now divided this table in two: Table 1 and Table 2; in Table 1 we present the socio-demographic variables and in Table 2 the sleep and socio-emotional variables.

C9: 7) Within Table 2, it would also be helpful to present the regression coefficients for the covariates at the different time points – particularly as some of these regression coefficients are mentioned within the results text.

R9: We appreciate this suggestion and have now included this information in Table 2.

Reviewer # 2

Comment 1: The authors presented a very well written, interesting and very informative manuscript. The reference list is very timely; the authors might consider that the prevalence rate of infantile colic (IC) is about 5%; thus, from a statistical point of view, about 47 infants with IC and their parents took part in the study; it further follows that this sub-sample could have biased the pattern of results (Brand et al., 2011, 2014). Please check once again ref# 14; it appears that 'Juulia' is a first name.

Reply 1: We appreciate this comment from the reviewer. The point concerning the prevalence rate of infantile colic is interesting; thanks for this suggestion. In the revised version, we have now conducted a new analysis excluding the 39 cases (36 at three months and 3 new cases at eight months) that we found in our database. We only had information about IC at three and eight months, and not at later ages. This is now mentioned in the statistical analyses ("Further, taking into account that the prevalence rate of infantile colic (IC) is about 5%,[22,23] and that we found 39 cases in our total sample (i.e. 4.2%: 36 cases at three months and 3 new cases at eight months), we conducted the same linear regression analyses after excluding all cases with IC" page 7) and results ("After excluding the 39 cases with IC, we observed that the main findings remained, which suggests that the results are not due to infantile colic (data not shown)" page 9). However, as the main findings remained the same after excluding the cases with infantile colic, we decided to keep the original analysis.

And concerning ref # 14, this mistake has now been corrected (in the current version, ref #16). Thanks for noticing this.

C2: Methods. The study design is nicely described; the samples and the tools are well reported. As regards the attrition and the characteristics of those mothers completing also later assessment points, please report all statistical indices: first, reporting p-values without further statistical indices is somehow pointless; second, reporting just p-values is highly discouraged (Wasserstein et al., 2019); third, it is conceivable that the large sample size conferred to 'significant' p-values; fourth, it follows that effect sizes should be reported, too.

R2: We appreciate this comment and all the information provided here by the reviewer. Consequently, we have now provided further detailed information concerning the significance values regards the attrition as follows: "In the drop-out analysis, we found that mothers participating at 24 months were older than those who did not [($F(1,1412)=14.766$, $p<0.001$, $\eta2=0.010$)], had higher educational levels [$X^2(2,1425)=22.811$, ($p<0.001$, $\varphi=0.127$)], and lower levels of anxiety [$F(1,1422)=10.015$, ($p=0.002$, $\eta2=0.007$)] and depression [$F(1.1421)=11.237$, ($p=0.001$; $\eta2=0.008$)]" (page 5).

C3: SPSS should be: SPSS® version 25.0 (IBM Corporation, Armonk NY, USA) for Windows®/Apple Mac®

R3: Thanks for noticing this. We have now updated this information as recommended by the reviewer (see page 7).

C4: "….the associations between sleep and BITSEA,…"; please replace the acronym BITSEA with the appropriate dimension; actually, the authors associated dimensions of sleep with dimensions of social, emotional and behavioral problems, and not with an acronym. Accordingly, replace the acronym also in the text.

R4: We understand this concern. Therefore, we have now changed this sentence "to study sex differences in the associations between sleep and BITSEA" with "to study sex differences in the associations between sleep and emotional and behavioural problems" (page 7). In the rest of the text, when we use the term BITSEA, this specifically refers to the questionnaire per se, and thus it is appropriate to use this acronym in those specific parts (Statistical Analyses and Results sections).

C5: References: Brand, S., Furlano, R., Sidler, M., Schulz, J., Holsboer-Trachsler, E., 2011. 'Oh, baby, please don't cry!': in infants suffering from infantile colic hypothalamic-pituitary-adrenocortical axis activity is related to poor sleep and increased crying intensity. Neuropsychobiology 64(1), 15-23;

Brand, S., Furlano, R., Sidler, M., Schulz, J., Holsboer-Trachsler, E., 2014. Associations between infants' crying, sleep and cortisol secretion and mother's sleep and well-being. Neuropsychobiology 69(1), 39-51; Wasserstein, R.L., Schirm, A.L., Lazar, N.A., 2019. Moving to a World Beyond "p < 0.05". The American Statistician 73(sup1), 1-19.

R5: Thanks for these useful references; the two references by Brand have now been included in the text.