Peer-Review comments and author responses

Reviewer 1

Dear Reviewer

We would like to thank you for your insightful suggestions and the careful reading of our manuscript. We adopted the STROBE checklist for cross-sectional studies and the STROSA checklist. We greatly appreciated the suggestions to improve the data analysis and we tried to accomplish with them. We performed the adjustments in the results section and the tables to adequate the reporting to the writing conventions. Please, find below the answers to your comments point by point.

> **-Comments:** The authors conducted a great analysis to see if there is any association between depressive symptoms and subjective daytime sleepiness using the NHANES dataset. First, I congratulate the authors for their incredible efforts in conducting this analysis. Even though previous studies have linked these two variables, to my knowledge, there are no records of population-based studies. Here are a few recommendations:

1. Abstract:

> **-Comment:** In the abstract, please use the same type of letters as the rest of the text

Response: *We appreciate the comment and are happy to inform you that it was fully taken into account on the correction.*

> -**Comment:** When reporting the Odds ratios (OR) from the model, there is no need to report the p-values or the log-likelihood. Most of the time if the odds ratios include one, it is not statistically significant; the exceptions to this rule are very specific. In the ORs, instead of providing the log-likelihood, provide the confidence interval, which gives more information about the variability of the estimate.

Response: *We appreciate your comment and are happy to inform you that the p-values and log-likelihood were removed from the abstract, leaving only the confidence intervals.*

2. Introduction

> -**Comment:** Good and well-explained. The use of the expression "oversimplify the complex mechanisms underlying the interaction between depression and sleep " is correct, but I will be more specific. Are there non-linear relationships? Dose-response (more symptoms, more sleepiness?).

Response: *Thank you for this suggestion. Indeed, research has shown a bidirectional relationship between sleep disturbances and emotional regulation, as well as non-linear relationship of depression and sleep duration, and this was highlighted in the text.*

3. Methods:

-**Comment:** This is a secondary analysis of a public dataset. There are two checklists that I recommend you use (to give a standardized structure to the manuscript: 1) The STROBE checklist for cross-sectional studies (as you are not using longitudinal data) https://www.strobe-statement.org/checklists/ 2) The STROSA checklist (https://pubme)

Response: *Thank you. We rewrote the specific sessions following STROBE and STROSA recommendations.*

-**Comment:** Most of the information required from the previous checklist is in your methods, just needs a reorganization. Using the previous checklist, there is missing information that is important to assess the quality of your statistical inferences.

Response: *We appreciate your comment and are happy to inform you that we included the previous checklist in the methods section to assess the quality of your statistical inferences using STROBE and STROSA, as recommended.*

-**Comment:** A flowchart (like the strobe flowchart) can help the reader to know how many participants from the original dataset make it to the final analysis, and what were your criteria (beyond just the exposure and outcome selection) for selecting these patients.

Response: *Thank you for your comments. We created a STROBE flowchart regarding participant's inclusion and selection.*

-**Comment:** Describe the data access or reference to an article or website of the NHANES dataset you used.

Response: *Thank you. We appreciate your suggestion and added the NHANES dataset reference.*

-**Comment:**. The PHQ-9 is a questionnaire of nine questions, where question three is "trouble falling or staying asleep, or sleeping too much". This means your exposure could be highly correlated/associated with your outcome. Is there enough data to exclude this question from the overall score? Or maybe just keep the people who answer "not at all" to this question.

-If you can exclude this question, you should report how you did it in the methods

-If there is no way to exclude this, I would still consider adding a more complex analysis to improve the statistical inference.

Response: *We appreciate the thoughtful comment and we agree sleep is one of the questions included in PHQ9 in such a way that any outcome in sleep would be intrinsically associated with depression symptoms measured by the PHQ9. However, the scale has been validated with this question and we cannot assure that the tool would preserve its validity after removing it. Besides,*

including only people who had answered "not at all" to this question would include a non-representative sample of the NHANES participants and of people with depressive symptoms, as sleep disturbances are common in depression. Considering this we decided to keep the question and the participants previously included and discuss this point as a limitation of the study.

-**Comment:** For instance, first, I will calculate the degree of correlation (the r) between the PHQ-9 and Sleepiness points. Also, look at the association (if you treat this as categorical) using a chi-square (look at the p-value if there is or is no association).

Response: *We thank the reviewer for the comments and suggestions to address this limitation. We handled PHQ9 as a categorical variable and the outcome for sleepiness is assessed in NHANES as a categorical variable. We investigated the association between these two variables using Chi-square, finding a significant association.*

-**Comment:** If these variables are correlated (e.g. r=0.8), and you feel comfortable using STATA, trying a Ridge logistic regression could be a good way to correct this. This can be a bit complex for the course objectives, as you will need to calculate the right lambda for your dataset. But if you feel curious about this, you can try it

-Why to use Ridge regression? Multicollinearity is a problem when using regression analysis not because the linear models can't represent highly correlated data, but is an optimization problem. In other words, when doing the "mathematics" the computer makes mistakes and needs another value to correct for those mistakes (the lamba), which is called a penalization for the correlation, or a regularization value. Here more info https://www.stata.com/manuals/lassoelasticnet.pdf. In page 8 you can see the lambda, which penalizes the high correlation.

Response: *We thank the reviewer for the suggestions. We considered enhance our model by Ridge regression, but we understood that this a method designed to correct linear regression models in which collinearity is a concern. However, we designed a logistic regression for our main outcome and the maximum VIF for our variables is lower than 10, which demonstrates there is not collinearity. Also, the maximum VIF for the linear regression was 1.01 for the included variables. So, we understood it would not be mandatory to penalize with a Lambda of Lagrange and preferred to discuss the limitation involved. We will try in the future to test Ridge regression to correct our results by the lambda.*

-**Comment:** If you are unable to correct this, please mention this in the limitations.

***Response*:** Thank you for your suggestions. We included the discussion about the limitation due to the intrinsic correlation of the exposure and the outcome.

-**Comment:** I see you used different ways to represent your exposure (PHQ-9). When you do this, this is called a "sensitivity analysis". Please report these changes of the exposure definition as sensitivity analysis.

Response: *We appreciate the comment and these alternative analyses are now reported as sensitivity analysis.*

-**Comment:** There are some unmeasured confounders you did not consider, like medication (some medications can make you sleepier, like dopaminergic medications). The presence of comorbidity can represent this, but if you can calculate the E-value for your estimates, that can be a great addition. E-value example: "Suppose a study finds that an exposure has an odds ratio (OR) of 3.0 for an outcome. The E-value for this OR would be calculated to show the minimum strength of confounding that would be needed to bring the OR down to 1.0 (no association). If the E-value is 2.5, it means that a confounder (That you don't have information about) would need to have an association with both the exposure and the outcome with an OR of at least 2.5 to explain away the observed association." This gives information for a future study that can calculate the effect of medication on sleepiness. Evalue calculator: https://www.evalue-calculator.com/, https://www.evalue-calculator.com/evalue/

Response: *Thank you very much for the insightful comment, which we happily took into account for the outcome analysis. Considering a non-exposed group prevalence of 20% in our sample, and the following odds ratio and confidence intervals identified in our study, we are able to use R as follows:*
Observed Odds Ratio (OR) = 4.39
Lower 95% Confidence Interval = 3.83
Upper 95% Confidence Interval = 5.04

After running the code, we got:

- *E-value for Point Estimate: 4.68*
- *E-value for Lower CI: 4.33*
- *E-value for Upper CI: 5.02*

The E-value of 4.68 for the point estimate implies that any confounder would need a very strong association (Risk Ratio ≥ 4.68) with both the exposure and the outcome to negate the observed effect. This is unlikely for many confounders, suggesting that the result is robust. Whereas for the lower and upper bound of the confidence interval the confounder would need Risk Ratios of 4.33 and 5.02. The E-value of 4.33 for the lower bound indicates that even under the most conservative scenario, a confounder would still need a strong association to fully explain away the effect.

Considering that the measured confounders had much weaker association, it´s plausible to assume that confounders with associations this strong would be unlikely. So, we can reasonably conclude that the observed effect is unlikely to be fully explained by confounding.

-**Comment:** Finally, keep in mind your question is an association question. These types of questions are knowledge-driven. The selection of the covariates is based on your knowledge and not using p-values. It will be good to represent your model using a Direct Acyclic graph (you can use Daggity or PowerPoint). In the association question, you

don't need to report the AUC or BIC. These measures are used when constructing prediction models, which is a different methodology than when doing models for association questions.

Response: We thank the reviewer for the suggestion and we fully agree that we are addressing an association question that has no predictive power. A Direct Acyclic graph was incorporated as figure 2 to represent this association.

-**Comment:** You can keep the VIF and AUC or BIC if you add in the methods, as a sensitivity analysis, that you will construct the model in a data-driven way (forward selection approach).

Response: *We appreciate your comment and decided to keep the results about VIF and AUC adding the information in the methods as part of the sensitivity analysis.*

4. Results:

-**Comment:** Looks good. Don't need to report the AUC, BIC, F-value, or R-square. Keep the CI to two decimals.

Response: *Thank you for the suggestion, however we decided to include this information in the methods section, because we understand that these analyses are also important.*

-**Comment:** Reporting the equation is good (it is a really good practice), but don't call it "prediction equation" as your model is not predictive, but explanatory (association).

Response*: Thank you very much for your comment. We are happy to inform you that the adjustment was performed accordingly.*

-**Comment:** You can write the equation using the equation tools in word, or paste it as an image after using Latex/overleaf. The way it is presented can be better. There are conventions of how to present a regression equation that you will need to meet if you want to do this.

Response: *We greatly appreciate your suggestion. We performed the adjustments to write the equation according to the convention.*

5. Discussion:

-**Comment:**. The discussion looks good. No comments.

Response: *Thank you very much for your kind words and positive feedback.*

6. Tables and figures

-**Comment:** In table one, report the p-values of the association between the rows and the column (exposure). Also include your outcome in the row. That can help to see how many of those with sleepiness have depressive symptoms. This is important due the question three of the PHQ-9.

***Response**: Thank you for the suggestion. We performed the required adjustments in Table 1 including the information about p-value and sleepiness (outcome).*

-**Comment:** The tables of the regression analysis, there is a convention to only report the OR and the CI. If you want to report the coefficients, you will need to choose only one of them (the ORs are more interpretable). You can have the crude (univariate) with the CI in one column, followed but the adjusted estimates with the CI in the next one. There is no need to present the Log likelikood or the coeffiecient as it is redundant information.

***Response**: We appreciate your comment. We performed the adjustments in the tables to report only the OR and the CI.*

-**Comment:** The calculation of the CI comes from the Standard error, so no need to report SE, the Z or the P>|z.

Response: We performed the adjustments in the tables to remove SE, the Z and the P>|z.

Reviewer 2

Dear Reviewer

We would like to thank you for the thoughtful suggestions and the careful reading of our manuscript. We adopted the STROBE checklist for cross-sectional studies and the STROSA checklist as well as we revised the manuscript to adhere to publishing guidelines. The manuscript was submitted to English revision and the necessary changes were performed to improve clarity. Please, find below the answers to your comments point by point.

-**Comment:**

Dear authors,

Your manuscript on depressive symptoms and daytime sleepiness is an important work in this field. Below you may see my comments.

1. Structure Guidelines

-**Comment:** Change "Filliations" to "Affiliations"

***Response**: We appreciate the comment, we correct the "Affiliations" term.*

-**Comment:** Try to adhere to the journal's publishing guidelines. Please use the same font for all parts of your manuscript (including the abstract).

Response: *The font of the whole manuscript was reviewed to guarantee that the journal's publishing guidelines were followed.*

-**Comment:** If you have a native English-speaking person in your group, try to have them read and review the manuscript.

Response: *Thank you for this suggestion. We understand the importance of ensuring the manuscript is written clearly and accurately. To address this, we have thoroughly reviewed the manuscript for language clarity and precision. Additionally, we sought feedback from a native English-speaking colleague to have a better-quality text. We hope the revised manuscript meets the language standards expected by the journal.*

-**Comment:** Whenever you use ">", "=" or any other mathematical signs, add a space between the numbers and the sign.

Response: *We appreciate the comment; the mathematical signs were corrected considering the orientation.*

2. Introduction

-**Comment:** - The last paragraph is a bit redundant. Try to combine the information more, e.g.: "Therefore, this study aims to investigate the relationship between depressive symptoms and excessive daytime sleepiness with data from the National Health and Nutritional Examination Survey (NHANES) 2017-2018. We hypothesize that subjective daytime sleepiness is strongly associated with the presence of depressive symptoms."

Response*: Thank you for pointing this out and for providing a suggested revision. We agree that the last paragraph could be more concise to improve clarity and reduce redundancy. Following your recommendation, we have revised the paragraph in the new version of the manuscript.*

3. Methods

-**Comment:** - In the first paragraph and first sentence add "the" before "NHANES dataset".

Response: *Thank you for pointing this out. We have revised the sentence in the first paragraph to include 'the' before 'NHANES dataset,' as suggested. The updated sentence now reads: 'We analyzed the NHANES dataset to evaluate*

- The following considerations were answered below:

-**Comment:** -Please combine the paragraphs more. In the first section before "statistical analysis", you have 6 paragraphs, sometimes one paragraph is just one sentence.

- Rephrase this sentence "Depressive symptoms are measured by the Patient Health Questionnaire-9 (PHQ-9), a screening tool for severity of depressive symptoms (Kroenke

et al., 2001), chosen as the main exposure". It is unclear. Better to start with "The main outcome of the study is depressive symptoms as measured by ...".

Rephrase this sentence "The sleepiness outcome was the question “how often you feel overly sleepy during the day”.". It would be better understandable if you write for example "To evaluate the daytime sleepiness, the subjects were asked " how often do you feel overly sleepy during the day".

Response: *Thank you for your thoughtful comments. We have revised this section to address the points you raised:*

1. We combined the paragraphs in the section before "Statistical analysis" to reduce fragmentation and improve readability.
2. The sentence on depressive symptoms has been rephrased for clarity and now begins as suggested: "The main outcome of the study is depressive symptoms as measured by ..."
3. The sentence on sleepiness has been revised to improve understanding, as per your recommendation.

-**Comment:** Please also combine the "Statistical analysis paragraph". There are too many subparagraphs, which makes it hard to read. Please write "The statistical significance level was set at a p-value of <0.05". It is always important to write full sentences.

Response: Thank you for pointing this out. We agree that the statistical analysis section could be streamlined to improve readability. To address this, we have revised the section to reduce fragmentation and combine related details into a cohesive paragraph.

-**Comment:** Well done choosing your statistical methods for analyzing this dataset.

Response: *Thank you very much for your kind words and positive feedback.*

4. Results

-**Comment:** The headline "descriptive analysis" doesn't provide much information about what follows in the paragraph. Consider using "demographic data" or "Participants" or something similar which provides a description of your paragraph.

Response: *We appreciate the comment, and we’ve adapted the headline as “Participant Demographics” to guarantee a better description of the paragraph.*

5. Tables

-**Comment:** the common abbreviation for Interquartile Range is IQR and not IQ.

Response**:** Thank you for the comment; we corrected the abbreviation for Interquartile Range.

-**Comment:** Please add explanations to your abbreviation in the legends below your tables.

Response: *Thank you for the comment; we included the explanations to the abbreviations in the legends.*

-**Comment:** Table 1: Add the (% female) at the front next to Gender and not in the column with "no depressive symptoms" and "depressive symptoms".

Response: *Thank you for the comment; we corrected the term's position.*

-**Comment:** Add below marital status: married, separated and single in separate rows and then just add the numbers in the "no depressive symptoms" and "depressive symptoms" columns.

Response: *Thank you for the comment – All changes has been made in the tables.*

6. Figures

-**Comment:** Please give a clearer description of the contents of your figure.

Response: *Thank you for your valuable feedback. We have revised the description of the figure to make its contents clearer.*

7. Discussion and Conclusion

`-**Comment:** Great analysis and summary of your work.

Response: *Thank you very much for your kind words and positive feedback.*