**Response letter Data analysis project**

Dear Professor Fregni, dear editors,

First of all, we want to thank you and your colleagues for taking the effort to assess our work in this great detail. We put thought in all comments and gladly incorporated them into our manuscript. Below, you will find a systematic response to both reviewers. Changes made due to comments from reviewer 1 are marked red in the manuscript, changes made due to comments from reviewer 2 are marked blue. Additionally, we incorporated changes made after our discussions with the faculty during the 5-day immersion course in Brazil, these changes are marked in green, in case they were not mentioned by reviewer 1 or 2 yet.

Additionally, we would like to make changes to the author list. The correct order can be found below as well as in the manuscript files.

In case of any questions, please contact us anytime. Otherwise, we will curiously await your re-evaluation of our manuscript.

Thank you in advance and kind regards,

Group 12

*the BiasBusters*

**Reviewer 1**

1. **Title:** *The manuscript's title reads more like a research question than a report of findings. While the content appropriately presents a secondary data analysis, the title's interrogative format ("Does physical activity influence...?") should be reformulated to reflect the investigative nature of the completed study. I suggest the authors revise it to a declarative statement accurately representing their analytical work.*

Thank you for the comments, we decided on the title ‘The Association of Physical Activity and

Allergies in Children: Insights from the NHANES 2005-2006 Database'

2. Introduction: The authors present a solid foundation for the study; however, some parts need clarification. First, the statement that "prevalence of allergies have become increasingly higher in recent decades" could be enhanced by specifying whether this trend is observed globally or is specific to certain regions. Given the study is based on survey data, this is an important point to clarify. In addition, the author's statement where they say "burden on the child and their parents." could be improved by elaborating on the types of burdens experienced, such as emotional, financial, or quality-of-life impacts, which would better illustrate the implications of these diseases.

To address your valuable and correct input, we tried to work out global differences in the incidence of allergies more comprehensively and included these adjustments in our revised manuscript. Also, we emphasize the burden of URTI even stronger and offer the reader an insight about the most significant points that impact children and their relatives.

3. **Methods/ Hypothesis:**

***a)*** *The hypothesis section is presented clearly but could benefit from expanding the methodology rationale, particularly in terms of why logistic regression was chosen for this analysis. Stating that the study will "adjust for known confounders" is important. However, the specific confounders intended for adjustment could be briefly mentioned to highlight the methodological rigor and thoroughness. Furthermore, detailing the types of allergies considered in the analysis and the criteria for "frequent and intense PA" would make the hypothesis and study design more transparent to readers.*

The study uses logistic regression because it has the capacity to model binary outcomes, facilitating the relationships between physical activity and allergies while controlling confounders such as age, gender, ethnicity and socioeconomic status.
The physical activity is categorized on frequency per week. Allergy types were not determined separately, the category was answered by the participant with yes or no, depending on whether they have been told by a doctor they have allergies or not.
Logistic regression produces odds ratios, which provide an intuitive interpretation of the relationship between predictors and the likelihood of the outcome, which we want to achieve for the interpretation of our analysis.

*b) The methods section includes some points that require explanation. For instance, the authors omit explicitly stating why a cross-sectional design was chosen for this study. This is important, particularly since understanding the causation between physical activity and allergies would require longitudinal analysis. Also, the authors chose the age range of 2-11 for their population. However, any considerations regarding potential developmental differences across this wide age range could be discussed to clarify if and how age-related variations in physical activity or immune system maturity were controlled.*

This study used a cross-sectional design due to the nature of the available data from the NHANES 2005-2006 survey. The NHANES database provides a rich source of data on health and lifestyle factors, including physical activity and allergies. While we acknowledge that establishing causation between PA and allergies would require longitudinal data. This study aimed to perform an exploratory analysis of associations and correlation to generate hypotheses rather than determine causality.

Although we recognize that developmental differences in physical activity levels and immune system maturity exist within this range, the age range of 2–11 years was chosen based on the available data in the selected database. Regarding immune system differences, we considered that children within this age range are in the developmental stage in which immune system functions are actively evolving. Regarding differences in PA levels, in our multivariable analysis we stratified for 3 groups of PA levels (less than 7 times a week, 7 times a week, and more than 7 times a week) and found no significant differences in the compared results regarding association with the outcome of "ever told by a doctor they have this age range are in the developmental stage in which immune system functions are actively evolving. Regarding differences in PA levels, in our multivariable analysis we stratified for 3 groups of PA levels (less than 7 times a week, 7 times a week, and more than 7 times a week) and found no significant differences in the compared results regarding association with the outcome of "ever told by a doctor they have allergies". Additionally, previous studies on similar topics have often used broad pediatric age ranges, which enhances the comparability of our findings with existing literature.

We have clarified these points in the revised manuscript to ensure greater transparency and scientific rigor.

c) *In defining the exposure variable, more detail on how "hard" or "vigorous" physical activity was operationalized is necessary. For instance, the authors could specify which question of the NHANES were used to operationalize the exposure; whether this measure was self-reported by the child, proxy-reported by a parent, or directly observed could help readers understand potential biases in physical activity reporting. As NHANES offers multiple variables related to physical activity, justifying why specific parameters were excluded would also improve transparency. If only one physical activity metric was used, acknowledging its limitations compared to comprehensive physical activity assessment tools could be valuable for readers evaluating the robustness and applicability of the findings.(orlando)*

The variable we have chosen is "timeweekyouplayorexercise". We selected this variable, because it was the only good activity variable available that was reliably assessed in children. We now added this fact in our manuscript and it is also mentioned in our limitation section. For our age group, proxy respondents answered the questions. This is also mentioned in the text.

d) *The outcome measurement is straightforward, relying on self-reported doctor-diagnosed allergies, as reported by NHANES. The selection of covariates is justified with prior literature, as discussed in table 1, however, the authors could include a brief discussion*

*on how these specific covariates might interact with both physical activity and allergy outcomes would strengthen the rationale for their inclusion in the model. The covariates of dust weight and room humidity, in particular, are important environmental factors, yet their measurement and relevance to the analysis would benefit from additional explanation. Detailing how these factors were quantified, especially if based on self-report or environmental measures, could enhance the reader's understanding of their influence on allergy risk.*

Table 1 was re-written and oput into the text section. After profound discussion with the teachers in the 5-day immersion course, we decided to remove the covariates of dust and humidity out of the analysis due to issues with among others missing data.

Concerning the remaining covariates, Zablotsky et al (2021) showed that the percentage of allergies increases with to age. Children aged 0-5 are less likely to suffer from seasonal allergies compared to older children ages 6-11 and 12-17. They also showed that boys are more likely to have seasonal allergies than girls (20.0% vs. 17.7%). Concerning ethnicity, they concluded that Non-Hispanic Black (21.3%) and Non-Hispanic White (20.4%) children have higher rates of seasonal allergies than Hispanic (15.3%) and Non-Hispanic Asian (11.0%) children. Low household income has been recognized as a risk factor for doctor-diagnosed asthma and eczema, which are often comorbid with allergies (Kojima, 2022). This information was included in the discussion.

4. **Statistics**:

a) *The statistical analysis section would benefit from additional information on how complete case analysis addresses missing data, especially if the data is not missing completely at random. Describing any sensitivity analysis performed to assess the impact of missing data would also enhance methodological rigor. In addition, specifying the approach used for confounder selection could further reinforce the robustness of the analysis.*

Missing data was less than 1%, therefore it was handled as complete case analysis. It is not possible to know the mechanism of missing data, however we suppose it occurred at random.

Sensitivity analysis would be an excellent approach to ensure robustness of our results, however it is outside the scope of our analysis.

Confounders selection was clarified in our text: "Covariates and their reported allergic associations were selected a priori based on a comprehensive review of the literature and their known or plausible associations with both physical activity and allergy development. Covariates of which data were available in the NHANES 2005-2006 database were included in the model."

b) *The statistical approach raises several methodological considerations that warrant attention. A key concern is the potential overadjustment. The current model adjusts for variables like room humidity, dust weight, and family income, which might mediate the*

*relationship between physical activity and allergies rather than simply confound it. This adjustment strategy could inadvertently mask true associations, particularly given the observed near-null odds ratios. Another concern is the presence of a Table 2 fallacy, where the authors interpret the coefficients of confounders as causal, where they may constitute mediators in their model. This situation needs to be addressed.*

We revised our covariates, as suggested and after speaking to Prof. Wypij and we dropped some of our covariates (dust weight and room humidity). The remaining covariates shouldn't be mediators of the correlation between activity and allergies.

You are absolutely right. Our study is a cross-sectional study and we cannot say anything about causality, we can only infer correlation in our paper. We revisited the tables and corrected the interpretation of coefficients.

5. **Discussion**:

a) *The reported lack of association between physical activity and allergies deserves deeper exploration. While the analysis shows remarkably narrow confidence intervals (0.971 to 1.026), suggesting adequate statistical power, the discussion omits that these values are compatible with a positive finding and also would benefit from comparison with previous research and examining potential methodological differences that might explain the discrepancy. The unexpected protective effect of room humidity is particularly intriguing and merits further discussion, especially since it contradicts previous research suggesting that higher humidity typically exacerbates allergic symptoms. (Timo)*

With our large study population, we should indeed have enough power to assess the association between physical activity and allergies.

Regarding your comment on our discussion. In our view, our findings are in line with previous findings. We cite the papers of Eijkemans et al. (2019), Lobelo et al. (2020), and Vlaski et al. (2008) which all did not find a significant correlation.

About room humidity, there is indeed literature that provides evidence that higher humidity correlates with the development of allergies. In our model, humidity is indeed slightly protective for developing allergies. We spoke about the variable in our group and decided to exclude it from our main model. The variable has a lot of missing data and we are not sure if the values were really measured or only reported from the participants. Therefore, we will not discuss this conflicting result in our discussion.

b) *The model's modest predictive ability, indicated by an AUC of 0.6251, reflects both the complexity of allergy etiology and the study's inherent limitations. At the same time, the goodness-of-fit test suggests adequate model fit, and the low discriminatory power points to significant missing covariates in the model. The absence of key variables such as*

*genetic predisposition, family history, passive smoking exposure, and geographical factors that likely contributes to this limited predictive capacity. Each of these factors plays a well-documented role in allergy development, and their omission should be acknowledged as a possible limitation.*

You are correct that our model misses some critical factors that play a role in the development of allergies. These variables were not provided in the NHANES dataset. This limitation was mentioned in our discussion to help the reader to critically analyze our paper.

c) *A more explicit discussion of limitations, particularly addressing issues related to study design and the nature of the NHANES dataset, would strengthen the section by providing a balanced perspective on the study's findings and contextualizing them within broader research. While the study benefits from the large, representative sample in the NHANES 2005-2006 cohort, its cross-sectional design inherently limits the ability to draw causal inferences between physical activity and allergy outcomes. Since cross-sectional data only provides a snapshot of associations, it precludes conclusions about whether physical activity influences allergy risk or if preexisting allergic conditions influence activity levels. Discussing this critical limitation would reinforce the interpretative caution needed when discussing the study's findings, especially given that causation cannot be inferred from observational associations alone.*

You are absolutely right. Using a cross-sectional study design, we cannot say anything about causality. We now also included a sentence on this problem in our limitation section.

d) *Additionally, while the study acknowledges the lack of specificity in physical activity measures for young children in the NHANES dataset, a more detailed discussion on how these limitations could affect result validity would be beneficial. For instance, relying on a single metric frequency of "hard play or exercise" per week omits other important characteristics like intensity, type, and duration, which may differentially impact immune and allergic responses. Explicitly addressing this limitation would point out the challenges of using secondary data not specifically designed to measure complex exposures like physical activity.*

You are absolutely correct, there are many limitations to working with a dataset like the NHANES. We included the mentioned limitations concerning intensity, type, and duration in the discussion.

e) *Furthermore, there is minimal discussion of potential unmeasured confounders beyond the brief mention in the limitations. A more detailed account of uncollected variables that could impact allergy risk, such as family history of allergies, exposure to indoor allergens, environmental pollutants, geographic location, and early dietary influences (e.g., breastfeeding)—would provide readers with a clearer understanding of factors that*

*may affect the robustness of the findings. Discussing how these unmeasured variables could influence results would enhance the section's transparency and clarify why these missing confounders limit the study's generalizability and explanatory power.*

We recognize the importance of missing several missing parameters, in the database, and elaborated on this issue in the discussion.

f) *Additionally, a brief mention of the potential for over-adjustment, as seen in the "Table 2 fallacy" where adjustments for covariates that may be on the causal pathway could obscure genuine associations between physical activity and allergies, would add depth to the limitations section. Given the study is based on the construction of a regression model, Recognizing the possibility of over-adjustment, especially with environmental covariates like room humidity and dust weight, would explain the methodological care needed when interpreting the null association observed between physical activity and allergy diagnosis.*

To address the risk of over-adjusting, we revised our model with help pof the faculty during the 5-day immersion course. We decided on eliminating the variables on dust and humidity, decreasing the over-adjusting.

g) *These observations present valuable opportunities to enhance the manuscript's contribution. Addressing these points would strengthen the work to meet publication standards. The suggested refinements would allow the manuscript to make a more robust contribution to the field.*

You are absolutely correct on the point of improving our limitations and our discussion section in general. We gladly accepted your suggestions and were happy to increase the quality of our work.

**Reviewer 2**

1. **Title:** Since this is an observational study, I recommend revising the use of the word "influence." Perhaps "associate" would be more accurate.

Thank you for your suggestion, we combined it with the title suggestion from reviewer 1.

2. **Abstract**

I. ***Background****: What is the rationale behind exploring physical activity (PA)? Is it related to inflammation, or another mechanism? Including a brief explanation could make your idea more compelling and clear.*

II. ***Methods****: How do you evaluate PA frequency in a child as young as 2 years old? For adults, there are established questionnaires, but how is this data collected for such young participants?*

III. ***Methods**: What type of allergy diagnosis are you referring to? For instance, is it food allergy or another type?*

IV. ***Results**: You mention vigorous PA; what about other intensity levels? Ensure the methodology clarifies how PA data is collected and categorized.*

V. ***Results**: The statement "median of vigorous PA 7 times per week (every day)" seems unusually high for children aged 2–11. Could this be an overestimation or misinterpretation?*

VI. ***Results**: The statement on age, race, etc., influencing the association between PA and allergy is unclear. Are these factors modifying the association or independently associated? Please clarify (this is not your objective)*

VII. ***Conclusion**: The stated conclusion includes other characteristics (demographic and environmental factors), but the primary objective focuses on PA. Consider revising to align with your stated aim.*

After adjusting our manuscript based on the comments from the reviewers we incorporated your comments into our newly formulated abstract

*3.* **Introduction***:*

**a)** *The increase in allergic conditions also poses a considerable socioeconomic burden on healthcare worldwide, with up to 6% of all annual doctor consultations attributed to allergies (House of Parliament United Kingdom, 2014) → Since your data is from NHANES, consider including a reference from the USA for consistency and relevance.*

We included a reference on the reasons for doctor consultations in the USA by St. Sauver et al (2013), an analysis performed by the Mayo clinic into the American population.

**b)** *I understand you are focusing on allergies as a broad category rather than a specific cause. What common mechanism do all allergies share? The inflammatory pathway? So, the hypothesis is that physical activity might have an anti-inflammatory component, right? This biological plausibility should be clarified to strengthen your argument.*

Allergic conditions, including asthma, rhinitis, and dermatitis, share a common underlying mechanism: the immune system's overactivation in response to environmental or endogenous triggers, leading to inflammation. This involves the dysregulation of pathways such as the Th2-mediated response, overproduction of IgE antibodies, and release of inflammatory cytokines like IL-4, IL-5, and IL-13. These inflammatory pathways contribute to the hypersensitivity and clinical manifestations of allergies. A recent systematic review and meta-analysis (Eichenberger et al.) has shown evidence that PA reduces systemic inflammatory markers and enhances immune regulation thus ameliorating symptoms in conditions like asthma, which shares inflammatory mechanisms with other allergies.

**c)** *I am a bit unclear about your hypothesis. While moderate physical activity might reduce inflammation, excessive activity can increase oxidative stress and inflammation. Could higher levels of physical activity exacerbate allergic symptoms? Should there be an upper limit?*

We recognize the importance of your remark. Unfortunately, the NHANES database did not provide further information regarding the intensity, duration and modality. Empirically, there will be an upper limit the useful exercise, but we could not investigate this knowledge gap in the current analysis.

*4.* **Methods***:*

**a)** *Is your physical activity exposure dichotomous (e.g., "Yes" or "No" for playing hard)? If so, make this explicit. If not, clarify the nature and categorization of the variable.*

The exposure is not dichotomous, it concerns a continuous variable, because the proxy reports a number of times the child played or exercised 'hard' per week. We added this information into the manuscript.

**b)** *Include a clear reference for how you defined your exposure and outcomes. I understand this comes from NHANES, but is there a specific source (e.g., a web page, manual, or publication) you can cite?*

The link to the database used was included in the methods section. Our assigned research question was 'Does exercise influence allergy in children?', a question for which we had to use the NHANES_2005-2006 dataset. We then assessed the dataset for parameters that included exercise in children. Unfortunately, the majority of exercise parameters specifically excluded children, which is why the only viable option was to use the parameter described 'number of times per week play or exercise hard'. Consequently, we assessed the database for outcome parameters concerning allergies. Closest to a reliable allergy diagnosis, we found the parameter 'ever been told by a doctor they have allergies. This concerns a dichotomous parameter with the options of 'yes' and 'no'.

**c)** *Are "dust weight" and "room humidity" included as covariates in your analysis? If yes, specify this in the methods section and provide reasoning for their inclusion.*

We decided to incorporate dust weight and humidity in our analysis, due to indications in the literature, that bith influence the development of allergies. Unfortunately, during our 5-day immersion course, we performed ongoing analyses on our dataset, which showed significant issues, among others due to missing data, in both categories. Upon these findings, we decided to exclude both categories form our current analysis.

**d)** *I did not understand your table 1. What do you want the reader to understand with it? Would it be the definition of the variables, whether they are categorical or continuous? Or would it be the rationale for choosing these variables?*

Building on our answer of the previous question, we decided to exclude dust weight and humidity from our current analysis. The remaining covariates were included in the text under the Methods section, eliminating the former table 1.

*5.* **Results***:*

***a)*** *Is the reported median frequency of PA (7 with IQR 4–7) accurate? For vigorous physical activity every day, this seems unexpectedly high. Double-check this value for consistency.*

We understand your concern of the high numbers of physical exercise. However, this also includes the times the child plays intensively at home, which for 2 to 11 year old children is realistic from our point of view. For example, if the child plays soccer one time per day, this would already amount to 7 times per week. To be absolutely sure, we also re-confirmed the numbers, which show to be correct.

***b)*** *Your eligibility criteria included children with complete data on PA and allergy, correct? If so, clarify that missing data refers only to covariates and not the primary exposure or outcome.*

Yes, the eligibility criteria for our study included only children with complete data on both physical activity and allergy status. To clarify, the missing data mentioned refers solely to covariates and does not pertain to the primary exposure or the outcome. For missing data below 5%, a complete-case analysis approach was employed.

***c)*** *Why present associations of other variables with allergy if that wasn't your primary objective? It might be more insightful to evaluate whether these variables modify the association between PA and allergy, aligning better with your study aims.*

We recognize that our analysis and its description takes away focus from our primary research question. We therefore accepted your suggestion and removed it.

**d)** *The final paragraph seems more like a discussion. Consider revising and moving interpretive comments to the appropriate section.*

We appreciate your suggestion to enhance the clarity and structure of our paper.In the revised version we removed the interpretive comment from the results section and moved it to the discussion.

**e)** *Why running a ROC curve? include it in methods*

Thank you for your comment regarding the applicability of the ROC curve. After carefully reviewing the methodology we decided to not include it in the main paper and keep our detailed analysis in the methods section.

**f)** *If you have a variable for the "possible cause of the allergy," conducting a subgroup analysis could add depth to your findings and enhance their relevance.*

Thank you for your valid comment on this issue. We conducted a supplementary analysis with variables that could be related to allergies (humidity and dust) and kept it available at the appendix.

6. **Discussion**:

**a)** *"In parallel, children have been engaging in less physical activities than in the previous century." Your data shows children exercising vigorously daily, which contradicts the general literature. Address this discrepancy directly—what explains this difference? Is it related to sampling, measurement, or other factors?*

Thank you for your comment. The following paragraph was modified addressing this specific discrepancy between our dataset and existing literature.

**b)** *Expand on the biological plausibility behind your hypothesis. If previous studies show no association, emphasize why your study provides new insights (why are you doing another study if we already know the answer for your question?). Did you include*

*unique variables or methodologies, or do you focus on a specific subgroup different from the other papers?*

Although previous studies have not found significant associations, we believe our research provides new insights for several reasons. Regular physical activity may modulate the immune system, enhancing immune function and promoting responses that reduce allergic reactions. In addition, if it takes place outdoors, it may increase environmental exposure, helping children to develop tolerance to different allergens. We combined this remakr with your question on the mechanism of allergies from the Introduction session and incorporated it into the manuscript.

**c)** *Elaborate on the second paragraph. Make the connections between ideas more cohesive. When describing other studies, specify similarities and differences regarding population, exposure, and outcomes to contextualize your findings better.*

We combined all above mentioned comments and remarks from you and reviewer 1, and elaborated and arranged the discussion newly. We hope, it is to your taste.

7. **Conclusion**:

**a)** *The statement "limitations in the available data underscore the need for further research to assess whether different intensities of PA, such as low to moderate exercise, might influence allergy risk differently" is better suited for the discussion section. Expand on it there.*

We re-wrote the Conclusion section and added the limitations into the Discussion section.

**b)** *The sentence about intense PA exacerbating asthma symptoms does not pertain to your study since it wasn't analyzed. Remove this and ensure your conclusion focuses on your actual findings and their implications. These refinements can make your manuscript more cohesive and aligned with your study objectives.*

We fully agree with your suggestion and converted it into the manuscript.

5-day Immersion course revision

Comments from faculty:

During the 5-day immersion course in Brazil, we had our session with Ben Illigens and other faculty to evaluate our analysis. We summarized his comments in the table below.

| | |
|---|---|
| Hypothesis = statement, not include words like 'may' | We corrected our hypothesis and marked it in the Introduction for you to see. |

| | |
|---|---|
| Visual plot of correlation of age vs exercise | We included a visual representation of the correlation between age and exercise. |
| Make figure titles clearer, not just put information in footnotes. Also missing information in figures → should be self-explanatory, e.g. fig 3 not clear | We improved all table and figure titles, specifically focused on making the tables and figures self-explanatory. |
| New division of x axis in figure: 7, 14 and 21 times per week? | We categorized the data and provided a clearer visual representation of it. |
| Bias due to misclassification with confounders → differential misclassification → make sure there is not for all PROMs, if there is diff misclass: report them! | We re-worked the classification of data and reported on their Limitations. |
| Limitations: kills our hypothesis, provide more data —-> rephrase the final comments to not look so negative | We interpreted the comment together with the comments from reviewer 1 and 2 on the limitation and re-worked the complete Limitations section. Changes can be found in the red and blue markings of both reviewers. |
| Try report something surprising/novel | The research question as well as the database were provided to us and were not up for discussion or change. Therefore, we attempted to explore research gaps that are surprising and/or novel for further research in the future. |
| Do more exploratory analyses with variables? | After reviewing our variables again, categorization of the exercise variable was performed and made at three different levels (less than 7 times per week, 7 times per week and more than 7 times per week). This aided us to interpret better our findings.<br>From this we noted that with an increase in exercise frequency, the odds of being diagnosed with an allergy, decrease. However, this association was non-significant. Details into this are found within the statistical analysis section (marked in blue). |
| Compare to a different dataset? | Unfortunately, due to time constraints and scope of this activity, we were unable to |

| | |
|---|---|
| | compare our results against a different database. Thank you for the very interesting suggestion, which we will keep in mind for the future. |