January 12, 2025

Dear Dr. Bell,

Thank you very much to you and the reviewers for your consideration of our manuscript, titled, “Disruption to Test Scores after Hurricanes in the United States.” We appreciate the reviewers’ insightful comments and feedback. You will find each of their comments included, in turn, below, followed by our bolded author responses.

Reviewer 1

This research is an important step to identify educational attainment after storms. The limitations you provide impact the robustness of the article and explanation of impact of storms on test scores. The results are not unexpected: Educational scores are highly variable by state, racial/ethnic, and sociodemographics. Further research into how return-to-school policies in place allow for better access to education and their long-term impact on test scores.

**Thank you very much for acknowledging the importance of this research. We agree with you that the results are not unexpected given geographic and sociodemographic variability in the United States and that further research is needed to enhance school system and student recovery in the aftermath of hurricanes and other disaster events.**

Reviewer 2

It is not clear how confounders were selected and they seem to only be factors

that would be associated with the outcome and not the exposure of interest.

Adjusting for factors that are only associated with the outcome may attenuate

observed associations so I recommend the authors reconsider their adjustment

variables. This paper may be helpful with thinking about that.

Schisterman, Enrique F.a; Cole, Stephen R.b; Platt, Robert W.c. Overadjustment

Bias and Unnecessary Adjustment in Epidemiologic Studies. Epidemiology

20(4):p 488-495, July 2009. | DOI: 10.1097/EDE.0b013e3181a819a1

**Thank you for this comment. We agree with you that selecting for these variables may have attenuated the association between hurricane exposure and standardized test scores. Considering this, we did a sensitivity analysis where we restricted the model to only the exposure and outcome of interest.**

**We have revised the Sensitivity Analysis section within the Methods of the manuscript to read as follows:**

**“Given that the addition of county-level and grade cohort-level covariates may have attenuated the association between hurricane-force tropical cyclone exposure and standardized test scores (Schisterman et al., 2009), we also conducted a sensitivity analysis where we removed all covariates from both the main model, as well as that restricted to counties with a single hurricane exposure over the study period.”**

**We also added a Sensitivity Analysis section within the Results describing the sole association we observed:**

**“Most of the sensitivity analyses we conducted showed similar results to the main model with some exceptions. In the main, state-specific model excluding all covariates, we observed a negative association between counties exposed to hurricane-force tropical cyclone exposure and RLA scores in Texas (β = -0.11; 95% CrI: -0.22, -0.01; PP[β<0] = 98%).”**

It would be very useful to incorporate any available information on school

closures into this analysis. Grouping together all counties that have high wind

speeds as being exposed may smooth over associations you wish to observe

because the impacts may be much worse for counties that experienced actual

school closures during the time period as a result of the hurricanes. With this, I

think the authors should also explain more why wind speed is the measure they

decided to use as a proxy for hurricane exposure. Wind speed is only one small

part of a hurricane’s impacts and it is no surprise that by using that as the

exposure, the majority of the counties included in the analysis are mostly directly

along the coast because hurricanes lose speed as they are on land longer.

However, it is important to note that in examples like Hurricane Florence, larger

impacts were observed further inland in North Carolina due to extreme flooding

that occurred as a result of the hurricane. These counties that were impacted

with things like school and hospital closures further inland due to hurricane-

related flooding would be included in your control group in this analysis when you

use wind speed as the exposure determinant.

**Thank you for this very insightful feedback. We agree with you that information on school closures would have enhanced our analysis, but we unfortunately did not have access to this data. We have added language about this in the Limitations:**

**“Sixth, we did not have any data available to us on school closures. Grouping together all counties that experienced high wind speeds irrespective of whether their schools closed may have attenuated the association of interest. Hurricane impacts on educational outcomes may have been far worse in those counties that experienced actual school closures in the aftermath of hurricanes during our study period.”**

**We also appreciate your comment about using wind speed as a proxy for hurricane exposure. In our analysis, wind speed is an indication of hurricane *occurrence* rather than an exposure variable itself. We recognize that hurricanes bring along precipitation and storm surge in addition to high winds. With that said, most hurricane related flooding occurs with high wind speeds, and storm surges inflicting the most damage are largely coastal and aligned with wind speed. We have therefore added the following sentence to our description of this variable in the Methods: “Research has shown that high wind speeds during hurricanes are strongly correlated with flooding, high storm surge, and structural damage, especially in coastal areas (Chavas et al., 2013; Murnane and Elsner, 2012; Musinguzi and Akbar, 2021).”**

Similarly, I think the authors should consider doing additional analyses examining

the impacts of repeated exposures as some of these may be areas that

experience more than one hurricane in the time period studied.

**Thank you for this suggestion. As mentioned in the Sensitivity Analysis section of the Methods, we already conducted a sensitivity analysis running the model in only those counties that experienced a single hurricane over the course of the study period. The results did not differ from those of the main model. However, considering your comment above, we conducted an additional sensitivity analysis where we repeated this removing all covariates.**

**We have revised the Sensitivity Analysis section of the Methods to read as follows:**

**“Given that the addition of county-level and grade cohort-level covariates may have attenuated the association between hurricane-force tropical cyclone exposure and standardized test scores (Schisterman et al., 2009), we also conducted a sensitivity analysis where we removed all covariates from both the main model, as well as that restricted to counties with a single hurricane exposure over the study period.”**

**We also reported the results we observed in the Sensitivity Analysis section within the Results:**

**“In the state-specific model restricted to those counties that experienced only one hurricane in the study period excluding all covariates, we observed a positive association between hurricane exposure and RLA scores in counties in North Carolina (β = 0.20; 95% CrI: 0.09, 0.32; PP[β>0] = 99.9% ), and a negative association in counties in Texas (β = -0.17; 95% CrI: -0.26, -0.09; PP[β<0] = 99.9%).”**

A lot of descriptive information that is useful for interpreting these results is

missing and making it difficult to interpret these results. The authors should

report baseline standardized test scores. The authors state that scores go up in

florida and down in North Carolina, but the baseline scores are not reported so it

is hard to understand what this means in comparison to each other. Additionally,

if I understand the methods correctly, the authors assume that the baseline

scores are the same across all states, but the baseline scores are not reported

so this cannot be confirmed.

**Thank you so much for pointing this out. We report the distribution of standardized test scores at baseline in 2009 and at the end of the study period 2018 for all the states included in our analysis in Supplementary Table 1. We also made this clearer to the reader at the beginning of the Results, writing, “The distributions of standardized test scores for each state included in our analysis at baseline in 2009 and the conclusion of the study period in 2018 are shown in Table S1.”**

In the figure 2A the authors report associations that seem to be associations with

demographic groups and test scores which is not the objective of this research

study. I do not understand why these are reported as these should be covariates,

not the exposure of interest. I recommend refencing this paper on the table 2

fallacy.

Daniel Westreich, Sander Greenland, The Table 2 Fallacy: Presenting and

Interpreting Confounder and Modifier Coefficients, *American Journal of*

*Epidemiology*, Volume 177, Issue 4, 15 February 2013, Pages 292–

298, <https://doi.org/10.1093/aje/kws412>

**We have moved the figures showing the covariates to the Supplement.**

The authors should provide more background literature on why this outcome

matters for health. Especially given the comments made in the limitation section,

it is known that standardized test scores do not do the best job giving a fair

assessment of student’s performance and it is unclear what the impacts are on

health of children. More background literature on what this means for health

would be useful if this paper is to be published in a health journal.

**Thank you for this suggestion. We have added the following language to the Introduction to emphasize the importance of academic achievement for concurrent and future health behaviors and outcomes:**

**“Academic achievement during childhood and adolescence has been shown to shown to be strongly associated with concurrent health risk behaviors including violence, tobacco use, alcohol and drug use, unsafe sex practice, sedentariness, and poor eating (Bradley and Greene, 2013; Skalamera and Hummer, 2016). Educational attainment is also a strong predictor of mental health (Jareebi and Alqassim, 2024), self-rated health (Elovainio et al., 2016; Lê-Scherban et al., 2014), risk of chronic disease (Choi et al., 2011; Magnani et al., 2024), and overall mortality and life expectancy (Kaplan et al., 2014; Puka et al., 2022).”**