

Peer-Review

Fang, Joyce. 2026. "Association between Flooding and Opioid Overdose Deaths in Appalachia." *Journal of High School Science* 10 (3): 197–214. <https://doi.org/10.64336/001c.165257>.

1. Instead of your method in the current paper, use a difference-in-differences (DiD) or matching approach to better control for pre-existing trends and confounders. Instead of just looking at post-flood deaths, DiD compares how much the treatment group changed relative to the control group. This is stronger than your current approach, because it accounts for differences that existed before the flood. Matching is a method to pair "treatment" counties with similar "control" counties based on characteristics like: Pre-flood opioid death rates, Poverty, Unemployment, Education, Demographics. Common techniques: propensity score matching or nearest neighbor matching. This reduces confounding because the control group is now more similar to the treatment group. Both methods help control for pre-existing trends and confounders, making your estimates more credible. They strengthen your argument that any observed increase in opioid deaths is associated with flooding, rather than other differences between counties. This will probably change your entire manuscript because it will probably change the results.
2. Provide effect sizes in meaningful terms (e.g., additional deaths per 100,000 population). Simplify tables and highlight key results. Include confidence intervals for all coefficients.
3. Only one flood year is inadequate for comparison and may introduce selection bias. Report at least 3 flood years with individual statistics.
4. Reword statements to emphasize association, not causation. Replace "strong correlation" with "significant association" and clarify limitations throughout the text. Focus on plausible mechanisms as hypotheses rather than established causal pathways. Discuss limitations in more detail and suggest avenues for future research. Condense literature review to avoid redundancy. Use precise scientific language, especially when describing statistical findings.

Point-by-point Response

Comment 1. Instead of your method in the current paper, use a difference-in-differences (DiD) or matching approach to better control for pre-existing trends and confounders. Instead of just looking at post-flood deaths, DiD compares how much the treatment group changed relative to the control group. This is stronger than your current approach, because it accounts for differences that existed before the flood. Matching is a method to pair "treatment" counties with similar "control" counties based on characteristics like: Pre-flood opioid death rates, Poverty, Unemployment, Education, Demographics. Common techniques: propensity score matching or nearest neighbor matching. This reduces confounding because the control group is now more similar to the treatment group. Both methods help control for pre-existing trends and confounders, making your estimates more credible. They strengthen your argument that any observed increase in opioid deaths is associated with flooding, rather than other differences between counties. This will probably change your entire manuscript because it will probably change the results.

Response:

Thank you for this valuable suggestion. In response, we significantly redesigned the analytical framework and implemented matching-based analyses to improve comparability between flooded and non-flooded counties. Counties experiencing severe flooding in 2011 were matched to non-flooded counties using pre-flood socioeconomic characteristics, including poverty rate, unemployment rate, educational attainment, and median household income. We evaluated multiple matching approaches, including 1:1, 1:2, and 1:3 nearest-neighbor matching, caliper matching, and full matching.

Matching quality was assessed by comparing the maximum standardized mean difference (Max SMD). As shown in Table 2, matching substantially improved covariate balance relative to the unmatched sample. Because no single matching approach is universally optimal, we evaluated the

association between flooding and opioid overdose deaths across multiple specifications. Results were highly consistent across matching approaches and are summarized in Supplementary Table S1. Following matching, fully adjusted regression models were estimated using matched samples with HC1 heteroskedasticity-robust standard errors. These results are presented in Table 3. In addition, the outcome variable was redefined as the change in average annual opioid overdose deaths between the three years before flooding (2008–2010) and the three years after flooding (2012–2014). By comparing pre- and post-flood changes between treated and control counties, the analysis is conceptually similar to a Difference-in-Differences framework. We have clarified this in both the Methods and Introduction sections.

Comment 2:

Provide effect sizes in meaningful terms (e.g., additional deaths per 100,000 population). Simplify tables and highlight key results. Include confidence intervals for all coefficients.

a. *Provide effect sizes in meaningful terms (e.g., additional deaths per 100,000 population).*

Response:

We agree. To improve interpretability, we added a public-health interpretation of the estimated effect size in the Results section. Specifically, we report that the average treated county had a population of approximately 66,000 residents in the Results section. Therefore, the estimated increase of approximately 1.8 opioid overdose deaths per county corresponds to roughly 2.7 additional overdose deaths per 100,000 population. This provides a more meaningful interpretation of the observed association.

b. *Simplify tables and highlight key results.*

Response:

We substantially revised and reorganized all tables to improve readability and emphasize the primary findings.

The revised manuscript now contains:

- Table 1: Unmatched regression results
- Table 2: Covariate balance across matching specifications
- Table 3: Fully adjusted regression models for matched samples
- Table 4: Subgroup analyses across demographic populations
- Supplementary Table S1: Comparison of alternative matching approaches

This reorganization reduces redundancy while highlighting the most important findings in the main text.

c. *Include confidence intervals for all coefficients.*

Response:

We agree and have added 95% confidence intervals throughout the manuscript. Confidence intervals are now reported for all coefficients presented in Tables 1, 3, and 4.

Comment 3: Only one flood year is inadequate for comparison and may introduce selection bias. Report at least 3 flood years with individual statistics.

Response:

We appreciate this important concern.

Our study focuses on severe flooding occurring in 2011 because this year provided a sufficiently large number of severe flood events in Appalachia while allowing adequate observation periods both before and after the flooding event.

To create a cleaner comparison, counties that experienced severe flooding during years other than 2011 were excluded from the analysis. Including multiple flood years would require constructing separate treatment cohorts, redefining pre- and post-event windows, and addressing overlapping flooding exposures, which would substantially alter the study design.

We therefore retained the single-event design but expanded the Limitations section to explicitly acknowledge this limitation. We also now note that future studies should evaluate multiple flooding years and broader geographic regions to determine whether similar patterns are observed across different disaster settings.

Comment 4: Reword statements to emphasize association, not causation. Replace “strong correlation” with “significant association” and clarify limitations throughout the text. Focus on plausible mechanisms as hypotheses rather than established causal pathways. Discuss limitations in more detail and suggest avenues for future research. Condense literature review to avoid redundancy. Use precise scientific language, especially when describing statistical findings.

a. *Reword statements to emphasize association, not causation.*

Response:

We agree and carefully revised the manuscript throughout. Terms implying causality have been replaced with language emphasizing association. For example, statements describing flooding as causing increases in overdose mortality were revised to state that flooding was associated with increases in overdose mortality. Similar revisions were made throughout the Abstract, Results, Discussion, and Conclusion sections.

b. *Replace "strong correlation" with "significant association" and clarify limitations throughout the text.*

Response:

We agree and revised the manuscript accordingly.

Throughout the manuscript, references to "strong correlation" were replaced with more precise terminology such as "significant association" or "positive association."

In addition, the Limitations section was expanded substantially. We now discuss:

- The focus on a single flood year
- Limited generalizability beyond Appalachia
- The relatively small number of treated counties
- The observational nature of the study
- The possibility of residual confounding despite matching and regression adjustment

c. *Focus on plausible mechanisms as hypotheses rather than established causal pathways.*

Response:

We agree.

The Discussion section has been revised to frame all proposed mechanisms as hypotheses that may explain the observed association. Psychological stress, disruptions in healthcare access, and economic instability are now discussed as plausible pathways that warrant future investigation rather than established causal mechanisms.

d. *Discuss limitations in more detail and suggest avenues for future research.*

Response:

We appreciate this recommendation.

The Limitations section was expanded substantially. We now discuss limitations related to the single-event design, regional focus, residual confounding, and the observational nature of the study. In addition, we added a dedicated Future Research section recommending studies that:

- Evaluate multiple flooding events across different years
- Examine broader geographic regions
- Investigate longer-term outcomes
- Explore mechanisms linking flooding and overdose mortality

e. *Condense the literature review to avoid redundancy.*

Response:

We agree and clarified the Introduction.

Several descriptive passages were shortened or consolidated while preserving essential background on flooding, opioid overdose mortality, and the Appalachian context. These revisions improve readability and focus.

f. *Use precise scientific language, especially when describing statistical findings.*

Response:

We agree and revised the manuscript throughout to improve scientific precision.

Statistical findings are now described using terms such as "estimated effect," "95% confidence interval," "positive association," and "statistically significant association" rather than causal interpretations. These revisions ensure that conclusions remain consistent with the observational design of the study.

Thank you for addressing my comments. The paper has significantly improved. However, some deficiencies persist. Please address further minor comments.

1.If you want to call this a true DiD, you will need to demonstrate that treated and control counties exhibited similar overdose trajectories before 2011. At minimum, include pre-2008, 2009 and 2010 annual overdose trends for both groups. Without this, readers cannot judge whether treatment and control counties were already diverging before the flood.

Alternatively, please describe the analysis as “matched pre-post change analysis” or “matched change-score regression.” and drop the DiD terminology.

2.Why is 1:1 matching worse than the unmatched sample? Table 2 reports max SMD of full sample=0.22 but max SMD for 1:1 matching is 0.36, which is worse. This deserves explanation. Readers unfamiliar with matching may interpret this as an error. Include verbiage along the lines of ““The 1:1 nearest-neighbor algorithm produced poorer balance than the unmatched sample because each treated county was restricted to a single control, resulting in several imperfect matches. Allowing multiple matched controls (1:2 and 1:3) substantially improved covariate balance by increasing the pool of comparable control counties.”

As to the reason why this inferior metric was included at all in the manuscript, please include something along the lines of ““Although 1:1 nearest-neighbor matching is a commonly used baseline, it produced poorer covariate balance than the unmatched sample in this dataset, likely reflecting the limited availability of closely comparable control counties. Consequently, subsequent analyses focused on the better-balanced 1:2, 1:3, and full matching specifications.”

3.Only one matching approach remains statistically significant. Hence, the robustness argument should be softened. Table 3 shows

1:2 matching → not significant

1:3 matching → not significant

full matching → significant.

Although all estimates remain positive, only one specification reaches conventional significance.

Therefore the manuscript should avoid implying that significance was consistently demonstrated. A more balanced statement would be “The estimated association remained positive across all matching specifications and achieved statistical significance under the full matching approach.”

4.Explain why your model was set up to calculate absolute counts instead of normalized deaths per 100,000 people (which is the norm).

5.Your model has a substantial mechanistic gap which needs to be incorporated as a limitation.

Please incorporate a paragraph with a verbiage similar to that below:

“The manuscript hypothesizes a multi-step causal pathway in which flooding leads to infrastructure disruption, reduced healthcare access, psychological and economic stress, increased opioid misuse, and ultimately increased overdose mortality. However, the present study directly measures only the relationship between flooding and overdose deaths; none of the intermediate mechanistic steps are empirically investigated or quantitatively validated. Consequently, while the proposed pathway is

biologically and socially plausible, the mechanistic explanation remains speculative within the context of this study. The existence of a statistically significant association should therefore not be interpreted as evidence supporting the entire causal chain. The proposed mechanism is mechanistically fragile because the intermediate causal links remain untested, and future work should directly investigate these intermediary processes before stronger causal inferences can be drawn.”

We sincerely thank the reviewer for the additional thoughtful and constructive comments. We have revised the manuscript to address these concerns. Point-to-point responses are provided below.

Comment 1:

If you want to call this a true DiD, you will need to demonstrate that treated and control counties exhibited similar overdose trajectories before 2011. At minimum, include pre-2008, 2009 and 2010 annual overdose trends for both groups. Without this, readers cannot judge whether treatment and control counties were already diverging before the flood.

Alternatively, please describe the analysis as “matched pre-post change analysis” or “matched change-score regression.” and drop the DiD terminology.

Response:

Thank you for this valuable suggestion. In response, we have changed the terminology and dropped any Difference-in-Differences verbiage throughout the paper, instead using the vocabulary suggested.

Comment 2:

Why is 1:1 matching worse than the unmatched sample? Table 2 reports max SMD of full sample=0.22 but max SMD for 1:1 matching is 0.36, which is worse. This deserves explanation. Readers unfamiliar with matching may interpret this as an error. Include verbiage along the lines of ““The 1:1 nearest-neighbor algorithm produced poorer balance than the unmatched sample because each treated county was restricted to a single control, resulting in several imperfect matches. Allowing multiple matched controls (1:2 and 1:3) substantially improved covariate balance by increasing the pool of comparable control counties.”

As to the reason why this inferior metric was included at all in the manuscript, please include something along the lines of ““Although 1:1 nearest-neighbor matching is a commonly used baseline, it produced poorer covariate balance than the unmatched sample in this dataset, likely reflecting the limited availability of closely comparable control counties. Consequently, subsequent analyses focused on the better-balanced 1:2, 1:3, and full matching specifications.”

Response:

We agree. Additional writing has been included to reflect on the poor results of the 1:1 matching and possible explanations in the results and discussion section.

The following paragraphs were added:

“Among the nearest-neighbor approaches, 1:1 matching nearest-neighbor algorithm resulted in the poorest balance (Max SMD = 0.36), even when compared to the unmatched sample. This is because each treated county was restricted to a single control, which resulted in several imperfect matches. Balance improved considerably when multiple matched controls were allowed (1:2 matching (Max SMD = 0.17) and 1:3 matching (Max SMD = 0.10)) by increasing the pool of comparable control counties. Caliper matching produced similar patterns but excluded one treated county because no suitable match was available within the specified caliper. Full matching also achieved strong covariate balance (Max SMD = 0.13) while retaining all observations through weighting.”

“Although 1:1 nearest-neighbor matching is commonly used as a baseline, it produced poorer covariate balance than the unmatched sample in this dataset. This likely reflects the limited availability of closely comparable control counties. Consequently, subsequent analyses focused on the better-balanced 1:2, 1:3, and full matching specifications.”

Comment 3:

Only one matching approach remains statistically significant. Hence, the robustness argument should be softened. Table 3 shows

1:2 matching → not significant

1:3 matching → not significant

full matching → significant.

Although all estimates remain positive, only one specification reaches conventional significance.

Therefore the manuscript should avoid implying that significance was consistently demonstrated. A more balanced statement would be “The estimated association remained positive across all matching specifications and achieved statistical significance under the full matching approach.”

Response:

We agree. The statements in the results section and discussions section have been revised to reflect a more balanced stance that avoids emphasizing consistent significance, for example, “The estimated association remained positive across all matching specifications and achieved statistical significance under the full matching approach.”

Comment 4:

Explain why your model was set up to calculate absolute counts instead of normalized deaths per 100,000 people (which is the norm).

Response:

Thank you for this important comment. We agree that deaths per 100,000 population are more commonly used in epidemiologic studies. This study, however, focused more on changes in overdose deaths, rather than overdose rates across counties. We have clarified this in the methods section and have included it in the limitations section:

“The outcome was defined as the change in the average annual number of overdose deaths before and after the flooding event.” (Methods Section)

“Another limitation of this study is that overdose deaths were analyzed as absolute counts rather than deaths per 100,000 people. Future studies using normalized death rates would strengthen comparisons across counties.” (Limitations Section)

Comment 5:

Your model has a substantial mechanistic gap which needs to be incorporated as a limitation. Please incorporate a paragraph with a verbiage similar to that below:

“The manuscript hypothesizes a multi-step causal pathway in which flooding leads to infrastructure disruption, reduced healthcare access, psychological and economic stress, increased opioid misuse, and ultimately increased overdose mortality. However, the present study directly measures only the relationship between flooding and overdose deaths; none of the intermediate mechanistic steps are empirically investigated or quantitatively validated. Consequently, while the proposed pathway is biologically and socially plausible, the mechanistic explanation remains speculative within the context of this study. The existence of a statistically significant association should therefore not be interpreted as evidence supporting the entire causal chain. The proposed mechanism is mechanistically fragile because the intermediate causal links remain untested, and future work should directly investigate these intermediary processes before stronger causal inferences can be drawn.”

Response:

We agree. The limitations section has been revised to reflect this important mechanistic gap. A very similar paragraph has been incorporated, stating, “Furthermore, the manuscript hypothesizes a multi-step causal pathway in which flooding leads to infrastructure disruption, reduced healthcare access, psychological and economic stress, increased opioid misuse, and ultimately increased overdose mortality. However, the current study directly measures only the relationship between flooding and overdose deaths, and none of the intermediate mechanistic steps are empirically investigated or quantitatively validated. As a result, while the proposed pathway is plausible

biologically and socially, the mechanistic explanation remains speculative within the context of this study. The existence of a statistically significant association should therefore not be interpreted as evidence supporting the entire causal chain, for the proposed mechanism is mechanistically fragile because the intermediate causal links remain untested. Future work should directly investigate these intermediary processes before stronger causal inferences can be drawn.”

Thank you for addressing my comments. Accepted.