Author’s response to reviews

Title: Associations Between Open Drain Flooding and Pediatric Enteric Infections in the MAL-ED Cohort in a Low-Income, Urban Neighborhood in Vellore, India

Authors:

David Berendes (dberendes@cdc.gov)
Juan Leon (juan.leon@emory.edu)
Amy Kirby (agkl@cdc.gov)
Julie Clennon (jclenno@emory.edu)
Suraja Raj (suraja.jeya.raj@emory.edu)
Habib Yakubu (hyakubu@emory.edu)
Katharine Robb (karobb@emory.edu)
Arun Kartikeyan (dr.arunks@gmail.com)
Priya Hemavathy (priya.hemavathy82@gmail.com)
Annai Gunasekaran (annaimbavellore@gmail.com)
Sheela Roy (sheelaroy1@gmail.com)
Ben Ghale (benowr@live.com)
J Senthil Kumar (senthilcmc2008@gmail.com)
Venkata Mohan (venkat@cmcvellore.ac.in)
Gagandeep Kang (gkang@cmcvellore.ac.in)
Christine Moe (clmoe@emory.edu)

Version: 2 Date: 04 Jun 2019

Author’s response to reviews:

Copying and pasting from author's response letter (included in suppl material as well):
Reviewer reports:

Haochu Li (Reviewer 2): Thank you for the opportunity to review the manuscript "Associations Between Open Drain Flooding and Pediatric Enteric Infections in the MAL-ED Cohort in a Low-Income, Urban Neighborhood in Vellore, India". Overall, most of the comments from the previous reviewer have been addressed. However, some concerns remained. My comments are the following:

1. Temperature is also a determinant of enteric infections, especially bacterial ones, and so should be included as a confounder.

We appreciate the reviewer's comment and agree that temperature is an important driver of enteric infections when examined in more general models aiming to estimate more 'upstream' effects of climate on incidence of enteric infection. However, given the environmental exposure pathway being examined—rainfall causing overflowing/flooding of open drains with feces in them, and associated exposures—it would appear that temperature would be an 'upstream' driver of rainfall (e.g. acting through rainfall) in this directed acyclic graph, and thus controlling for temperature would be duplicative of testing our main effects (e.g. detrimental to testing associations with rainfall/drain flooding) rather than accounting for residual confounding.

References are needed to support the statement "it would appear that temperature would be an 'upstream' driver of rainfall (e.g. acting through rainfall) ".

- We thank the reviewer for this comment. We were able to obtain temperature data matched to the study period and at the same scale as rainfall data. We have therefore included those data in our adjusted models and updated the results of tables 2 and 3, as well as included a description of the data in the methods (lines 131-134 and 194-196). Overall, results are consistent with results from the previous version of the manuscript (table 3).
2. Overall, the lack of consideration of other exposure routes and key risk factors makes the results very difficult to interpret. The use and description of the weather (not climate) factors is also not appropriate and impossible to interpret.

I don't think this comment has been addressed properly.

- We appreciate the reviewer’s request for additional clarification. As described at the end of our background (line 112), the goal of the study is to test the hypothesis of whether or not enteric infection was associated with reported flooding/rainfall area (which we found to be present) or reported contact with drain water (which we did not find to be present). Thus, the goal of the study was hypothesis testing, and not an exploratory analysis of all possible environmental exposure pathways. We agree that there may be other important exposure pathways/risk factors, and to account for this, we adjusted for key ones identified a priori and via previous analyses (i.e. reference 15): income, assets, mother’s education, household sanitation, type of stool collected (diarrhea v. no diarrhea), and now temperature as well. We feel that the goal of the paper has been clearly stated at the end of the background, with hypotheses provided, and subsequently answered directly, step by step. Table 3 sections A and B, for instance, clearly delineate results associated with the two hypotheses in order.

- To address the weather factors, we have added sections per above.
3. The household flood exposure data (based on self-reported assessments) is not useful unless linked in time to the illness event. In some cases, the flood may have occurred after the illness?

This comment is not addressed properly.

- As previously discussed and included in our limitations, we were not able to collect flood incidence data in real time. Instead, we were able to collect questions about routine drain or house flooding (e.g. “do the drains near your household ever flood?”) and combined these data with cumulative monthly rainfall data (which are matched to the month of stool collection). Thus, models assess the interaction of flooding variables and rainfall on odds of enteric infection to see whether the relationship between a child living in a household that reports routine drain flooding and the child’s odds of enteric infection is modified by the amount of rainfall in a month—a significant interaction therefore indicates that the child’s odds of infection based on living in that household increases with rainfall, which suggests that rainfall-induced flooding affected the child’s odds of infection in that month. We have added this explanation to the end of the methods (lines 238-243) to clarify.

- Additionally, we used spatial analysis to identify identified locations where reported routine flooding was clustered, which suggesting areas that may flood habitually (with increasing rainfall). Visual identification on the map (Figure 1) confirmed that these clusters were ‘downstream’ on the drainage network by elevation, which gives us greater certainty in the self-reported results. Importantly, we discuss in lines 414-421 the limitations of reconstructing flooding using this approach (and encourage direct measurement in future studies), and discuss in lines 228-236 specifically how each exposure variable was constructed and defined.

4. what is the highest frequency of contact?

- As described in line 183 of the methods, frequency of contact was answered in categories of 0 contacts per month, 1-5 contacts/month, 6-10 contacts/month, and >10 contacts/month. Therefore the highest frequency of contact category is >10 contacts per month.
5. It is tricky that the measurement of exposures took place after outcome measurement. Also, this subanalysis is based on complex data from more than one study. A much better and clearer description of data collection, preparation, management, or utilization for this analysis are needed.

• We appreciate this comment, and have added text to the data sources (lines 124-141) to describe which variables come from which datasets/studies. Additionally, we have delineated the sections of the methods by study/data source and clearly identified them in the title of each section. Finally, lines 225-244 clearly identify which variables are used in the analysis and how they are defined, including the flooding-related variables (per #3 above).

6. Household selection needs clearer explanation. Why 25 households were selected based on a hygiene survey, but the left 75 households were selected from the other?

• As described in lines 171-176, the 25 households were selected beforehand for comparison as part of a separate study, while an additional 75 were chosen randomly. The 25 households selected as part of the hygiene study and analyzed in reference 28 (I am the lead author on this) did not vary significantly in location, demographics, WASH status, or otherwise from the 75 households selected randomly.

7. additional careful edit is needed indeed.

• We thank the reviewer for this comment and have gone through the manuscript in-depth to clarify confusion, separate long sentences, and generally revise for clarity.