Effect of water, sanitation, and hygiene interventions on enteropathogen detection in the environment: an individual-participant data meta-analysis

**Coauthor responses**

**Major manuscript changes to note**

* We have addressed in this document all emailed coauthor comments as well as major comments from the manuscript document.
  + Smaller comments added onto the manuscript are responded to in the attached tracked-changes version of the manuscript.
* We had previously combined all MapSan study environmental sampling, including unpublished results, and labeled them under Holcomb et al. 2020. But Drew Capone has published results from analyzing pathogens in soil, and has an analysis of pathogens on flies in preparation, so we have split the soil and fly samples into their respective studies (Capone et al. 2021 and Capone et al. 2022 in prep.).
  + We combined fly types (kitchen versus latrine) in Capone et al in prep based on feedback on sampling schemes where latrine entrance flies were only analyzed when they didn’t catch any flies at the kitchen area.
* Based on multiple co-author comments on the lack of specificity of general MST markers like GenBac, we dropped general MST markers from the analysis.
  + Previously, we were excluding three very high prevalence pathogens or MST markers from the aggregate outcome to make the analysis possible, but that biased estimates away from the null. We now include all pathogens/MST markers, especially because GenBac, which has been excluded for being a general MST marker, was the highest prevalence markers that was primarily causing the analysis issue of close to 100% prevalence in some aggregate outcomes.
* Slight changes in adjusted results due to updated covariate cleaning that occurred in completing the second manuscript (associations between environmental measures and child health outcomes).
* We dropped “vote counting” of the number of significant results as well as non-significant but protective results on based on Cochrane guidance to avoid this practice.

# Jacky Knee

This is something I should have mentioned in a review of the analysis plan way back when (so feel free to ignore) but might it make sense for human MST markers to be the primary outcome? Just thinking that most of the interventions in the included study would have limited impact on animal faecal contamination (aside from the scoop provided with WASH-B), plus lumping them together sort of negates the rationale for using them in the first place.

Because we pre-registered it in the analysis plan, we should keep any MST marker as a primary outcome. We had pre-registered that as the primary outcome because we did not know a priori that we would catch only sanitation trials. Water treatment and handwashing can reduce animal markers, so we analyzed human and animal MST markers together (pre-specified) as well as separately. But we’ve tried to better highlight in the results sections the human MST results and how they are the outcome expected to be affected more by interventions.

# Amy Pickering

Editors/reviewers might ask why this is a separate paper- I could see them thinking at a big picture level, there is limited new insight from combining five studies that had no/little impact of WASH interventions on env pathogens to conclude there is no/little impact of WASH interventions on env pathogens.

That is a concern of ours as well, and we separated partly due to the volume of results and the difference between evaluating randomized interventions and the observational analysis nested within studies. We plan on submitting the manuscripts together and would be open to combining the results if the journal editor asks for it or if the split manuscripts get rejected.

Not sure it makes sense to combine human, animal, and general into a composite. Would suggest at least removing general from the composite.

We have removed general MST markers from the whole analysis based on comments from yourself and Ali

# Steve Luby

Indeed, given the heterogeneity of study settings the rationale for pooling is suspect. It implies that there is an underlying universal phenomenon that we are measuring. We would expect different interventions to have different impacts in different contexts depending on number of local variables.

We agree with the limitations in pooling across diverse study settings, and have added to the limitation section the assumption that pooling is reasonable, but we believe that reporting pooled estimates alongside the study-specific estimates provides a useful summary measure across studies if effects are consistent. We use random effects models so that we aren’t assuming there is a universal phenomenon that we are measuring, but instead there is a distribution of intervention effects. The heterogeneity in intervention effects was insignificant when there were >4 estimates to pool (largely because of null or near null effects), so pooling isn’t obscuring large differences in individual study effects.

Although pooling all of this does increase statistical power, it does not seem to me to be intellectually coherent to combine very different interventions and different media.

We have added a discussion of this to the limitations, as we agree with the tradeoffs you both discuss. We also discuss pathogen-specific effects in specific types of environmental samples below. With 8 sample types and 38 pathogen/MST targets, the aggregated measures allow us to distill the complexity while the pathogen-specific and sample-specific estimates then provide the nuance.

# Jade Benjamin-Chung

**An advantage of measuring pathogens instead of indicators is that it could elucidate impacts on specific environmental pathways, which vary from pathogen to pathogen, and which interventions may intervene on differently. But by emphasizing the results pooled across all sample types and pathogens, you aren’t able to speak to that. This seems like it should at least be mentioned as a limitation.**

We have added a discussion of this to the limitations, as we agree with the tradeoffs you both discuss. We also discuss pathogen-specific effects in specific types of environmental samples below. With 8 sample types and 38 pathogen/MST targets, the aggregated measures allow us to distill the complexity while the pathogen-specific and sample-specific estimates then provide the nuance

# Kara Nelson

I find the results quite powerful, especially given all the sub-analyses and the different ways you present the results in the supplemental tables and figures (which I think address Steve’s concerns about the conceptual basis for pooling across so many factors, bc the results are also presented with many different ways of pooling).  I added a few minor comments to one of the middle versions, attached.  From my perspective on pathogen measurement and the engineering side, I’ll add a few other thoughts here:

* One type of sub-group analysis that you didn’t do is looking at the time after intervention, to explore whether there is more effect later (either due to higher compliance, or the RNA/DNA signal or the helminth eggs degrading in the environment).  Is there a reason that isn’t possible to do?

The primary issue with a time-after intervention subgroup analysis is that, within studies, there is limited variation in the time between intervention delivery and sampling, and across studies there is too much variation in types of samples and microbial targets to assess if differences are due to the time after the intervention. Additionally, we didn’t include this subgroup analysis in the analysis plan.

* Are there any other take-aways about the utility of specific pathogen targets measured?  I’m referring to the data in Table 2 and Table S4 - emphasizing the abundance and prevalence data rather than the ratios.  Most targets had very low prevalence, but not all.  It’s a bit hard for me to “see” any patterns with the information in Table format, so one of my suggestions is whether additional graphical analysis might help.

We explored the prevalence through heatmaps (draft version below), but we didn’t see a pattern across the sample types and targets beyond what is discussed in the results. But we are open to co-author input as to if any interesting patterns in the prevalence are apparent, or if this figure should be included in the supplementary material.

A picture containing chart

Description automatically generated

* I don’t think the manuscript mentions anything about compliance -  which seems highly relevant to the final recommendation on “alternative sanitation modalities”, and begs the question that perhaps a reason larger reductions in pathogens and diarrhea are not observed is that some household members, notably children, aren’t using the latrines.  The fact that there are such consistent small or null effects in terms of pathogen reductions, across so many different pathogen targets and sample types and contexts confirms that poop is still very present. The results of this IPD meta-analysis make it clear to me that pathogens detected in intervention hh can’t be explained away by non-human sources, or flooding, or vectors.

Great point; we had a short section in the results on high intervention uptake that focused on the physical infrastructure. We added to the final paragraph of the discussion a discussion of the compliance from WASH Benefits structured observations and how that may limit the impact of the interventions, and how methods of social change/education may be needed for interventions to be successful.

# Ali Boehm

Hi Ayse and Andrew - thank you for sending this and I’ve put some comments in here.

- I dont think you should include genbac in your study. its not an MST marker and is more like E. coli. i dont think genbac is going anywhere...

We have removed GenBac as well as the other general MST marker (BacUni) from the analysis.

- There is an issue with comparing QPCR data between studies. Some of these issues are spelled out in a recent paper I co-authored:

<https://pubs.acs.org/doi/abs/10.1021/acs.est.1c01767>

I dont think you actually combined data from across studies, right? But you might want to mention some of the technical issues associated with combining data across sites. we also touched on that issue in the Boehm et al paper

* **A. B. Boehm**, L. C. Van De Werfhorst, J. F. Griffith, P. A. Holden, J. A. Jay, O. C. Shanks, D. Wang, S. B. Weisberg. 2013. Performance of forty-one microbial source tracking methods: A twenty-seven lab evaluation study. *Water Research*, 47, 6812-6828.

Correct, we didn’t combine any of the abundance data, both because of issues in standardization but also because we estimate intervention effects separately within each study, and then there were too few target-specific abundance estimates to pool with meta-analysis models. We added citations to these references to the methods section as additional motivation to keep the abundance data separate.

- in your recommendation of things to be standardized, a standardized way of reporting data and results will also be helpful to mention.

Good suggestion, we added that recommendation to the discussion.

# Tom Classen

1.  Did many of your eligible papers also analyze samples for FIB or other proxies?  Your advocacy for assessing pathogens rather than proxies might be stronger if you could show the results are actually different when assessing the former rather than simply the latter.  I understand that is not your focus here, but your dataset might be uniquely well designed to make this comparison.

Most of the included intervention studies also analyzed FIB but not necessarily in the same studies measuring pathogens or MST markers, and the FIB was not shared with us. We agree that the results would be strengthened by comparing the results with FIB as an outcome, but unfortunately that is beyond the scope of the current analysis. That is a good idea as a follow-on paper if we can get the FIB data from all data contributors.

2. I don’t think the advantages of IPDs over conventional SRs is well understood, and the paper only addresses that briefly in Methods (line 4 under Data Collection and Analysis) and in the Discussion.  I’d suggest touching on that in the Introduction to build the rationale for this approach.

We have expanded on the advantages of IPD meta-analyses within the methods and added a sentence in the introduction to give the advantages and rationale for the approach.

3. Either in the Introduction or in the Discussion, I would mention Fred’s scoping paper (Environ. Sci. Technol. 2020) on the challenges of assessing faecal contamination in the environment.  I think we should acknowledge here the raises important shortcomings about the sampling approaches and the assumptions about dose that many of our evaluations included in this paper make about how to assess intervention effects.

Great suggestion, we have added that source and a sentence to the discussion on how single time point samples in the environment may be poor proxies for the actual ingestion of pathogens.

4.  Someplace in the limitations, we should also emphasize that none of these studies was actually designed primarily to assess intervention effects on pathogens in the environment.

Great point, we have added that to the limitations.

5. I understand the shortcomings of proxies vs pathogens, but I’m not quite so willing to dismiss the usefulness of assessing proxies.  In separate IPDs with many more observations than presented here, Goddard (Lancet PH 2020) and Hodge (EHP 2016) both showed a pretty strong association between proxies and health outcomes (DD and HAZ).  It’s also a much cheaper, accessible and field-ready method.   We need to acknowledge this in the Discussion, and perhaps not be too quick to announce the death of proxies just yet.  As you know, they still are prominent in water quality testing worldwide, as recommended by the GDWQ.

Based on this and other suggestions, we have rewritten the discussion to better discuss the comparisons and use-cased for FIB vs. pathogens, and now advocate for using both in different circumstances due to the low cost of FIB, especially in light of the limited association between pathogens and diarrheal disease in our second paper on this dataset.

# Ben Arnold

Hi Ayse and Andrew

Really great synthesis and clear paper!  I also agree with Tom and Steve's points. I have added a few additional small comments for your consideration.

Echoing Tom's remarks around FIB but channeling our work in a different field around seroepidemiology (e.g., with Njenga): I have found when proposing new biomarkers for surveillance or endpoints in trials, that the scientific community typically wants to see the head-to-head comparison of the new markers with the current standard measures. I suspect that is beyond the scope of this work, but it is something you can anticipate people being pretty curious about.

 Thanks Ben, including FIB bacteria as another outcome is beyond the scope of this work and isn’t currently included in all studies in the data shared with us, but is a good idea for a potential follow-on paper.

 I know space is limited, but might be helpful to understand the rationale for this composite outcome and how it could/should be interpreted. I can imagine upsides and downsides, but I’m sure you all have discussed and thought very carefully about it!

We have clarified here that the composite markers and assessed separately.

While we are also cautious about the composite indicator because of the greater uncertainty in what an effect on a composite indicator should be interpreted as, we chose to focus on them as primary outcomes because on the mostly different set of individual targets measured in studies as well as the sparsity of some individual outcomes. To be able to pool across studies and better leverage the IPD nature of the analysis, composite measures were needed, but we hope we also sufficiently highlighted intervention effects (or lack thereof) on individual outcomes.

We’ve added a longer discussion of the limitation of interpreting effects on composite outcomes to the Discussion (and cited the paper you shared).

# Laura Kwong

Excellent write-up and clear graphics. I don't have anything to add to the helpful comments that have already been made, except that you could mention the limitation of environmental sampling - we are only able to sample relatively small quantities from single locations at single points in time, which does not reflect a child's exposure; it is possible that we could see reductions in pathogen concentrations if we were able to sample larger quantities of (composite) environmental samples.

Thanks,

Laura

Good suggestion, we added that limitation to the discussion.