AAssociations between detection of enteropathogens and microbial source tracking markers in the environment and child enteric infections and growth: an individual participant data meta-analysis

Andrew Mertens, Jack Colford, Oliver Cumming, Joe Brown, Jill Stewart, David Holcomb, Drew Capone, Jackie Knee, Tom Clasen, Heather Reese, Amy Pickering, Clair Null, Steve Luby, Jessica Grembi, Ben Arnold, Audrie Lin, Jade Benjamin-Chung, Laura Kwong, Lauren Steinbaum, Ali Boehm, Kara Nelson, Erica Fuhrmeister, Mahbubur Rahman, Sammy Njenga, Rassul Nala, Ruwan Thilakaratne, Ayse Ercumen (middle order not finalized)

## Abstract

**Background:** Fecal contamination is typically measured using fecal indicator bacteria (FIB). FIB in environmental samples have been associated with increased risk of diarrhea and reduced linear growth in children. However, FIB are imperfect predictors of enteropathogens, and few studies have assessed associations between detection of enteropathogens and microbial source tracking (MST) markers in the environment and child health outcomes.

**Methods:** We conducted an individual participant data (IPD) meta-analysis to pool data from WASH trials to assess associations between the presence of pathogens and/or MST markers in the environment and enteric infections with specific pathogens, caregiver-reported diarrhea and height-for-age Z-scores (HAZ) in children. We used covariate-adjusted regression models with robust standard errors to estimate associations and pooled results across studies. For the infection and diarrhea outcomes, we used environmental data from up to four months prior to the measurement of health endpoints; we used all available environmental data prior to the HAZ measurement.

**Findings:** We identified and received data from five randomized or quasi-experimental studies. Detection of a specific pathogen in environmental samples was consistently associated with increased risk of subsequent child infection with the same pathogen. However, there was no consistent association between detection of enteropathogens or MST markers in the environment and subsequent caregiver-reported diarrhea, except for a significant increase in diarrhea risk associated with any pathogen detection in any sample type during wet seasons. Detection of any pathogen in any sample type was consistently associated with slightly lower HAZ (adjusted pooled mean difference: -0.08 (95% CI: -0.15, -0.01)); there was no association between MST markers and HAZ.

**Interpretation:** Detection of enteropathogens in the environment was associated with increased risk of pathogen-specific infections and lower HAZ but not with caregiver-reported diarrhea, highlighting the limitation of reported diarrheal symptoms as an outcome measure. MST markers showed no consistent associations with any child health outcome. Funding: The Bill & Melinda Gates Foundation.

## Methods

We examined associations between enteropathogens and MST markers in the environment and child health outcomes, including enteropathogen-specific infections, caregiver-reported diarrheal disease and growth . We classified enteropathogens and MST markers in the environment into multiple exposure variables. Our primary exposure variables were the prevalence of any enteropathogen and any MST markers in any type of environmental sample. We also tabulated prevalence separately for each sample type (e.g., water, hands, soil, flies). Secondary exposure variables included the prevalence of specific pathogen types (any viruses, any bacteria, any protozoa, any helminths), the prevalence of MST markers from specific host types (general, human, animal), and the prevalence and abundance of individual enteropathogens and MST markers. The primary outcomes were diarrheal disease prevalence and height-for-age Z-scores (HAZ). For specific enteropathogens detected in the environment, primary outcomes also included child infection with the same pathogen ascertained by stool testing. Secondary outcomes included Z-scores for weight-for-age (WAZ) and weight-for-length (WLZ) and the prevalence of stunting, wasting and underweight. For the growth outcomes, we used data from all environmental samples collected over the child’s lifetime prior to the anthropometry measurement. For diarrheal disease and enteropathogen-specific infections, we only used environmental samples collected up to four months before the measurement of the health outcome.

For binary outcomes (prevalence of pathogen-specific infection, diarrhea, stunting, wasting, underweight), we estimated prevalence ratios associated with the different exposure variables using modified Poisson regression.1 For continuous outcomes (child anthropometry Z-scores), we used linear regression to estimate mean differences. Because of repeated sampling or clustered designs in some studies, we used the Huber Sandwich Estimator to calculate robust standard errors.2 All analyses were adjusted for potential confounders. We included child age and asset-based household wealth as adjustment covariates for all adjusted estimates. Other covariates were prescreened using likelihood ratio tests, and only variables associated with the outcome with a p-value < 0.2 were included in the model for each outcome. We included the following variables in the prescreening set if they were measured within an included study: child age, child sex, maternal age, household food security status, number of people in the household, age and education of primary caregiver in the household, asset-based household wealth, number of rooms, construction materials (walls, floor, roof), access to electricity, land ownership and if anyone in the household works in agriculture. Within each study, we only estimated associations when there were at least 5 cases of the binary outcome in the rarest stratum of the exposure.

Given the heterogeneity in study settings (e.g., local WASH and nutrition conditions, climate, urbanization, population density, region-specific infectious disease patterns), we reported individual study-specific estimates for all analyses. For targets where data were available from four or more studies, we tested for heterogeneity in estimates using Cochran’s Q-test.3 If there was no significant heterogeneity (p-value>0.2), we pooled estimates using fixed-effects models, otherwise we pooled estimates using random-effects models.

Additionally, we conducted subgroup analyses by animal ownership in the household, season (dry vs. wet), study setting (rural vs. urban), child age (immobile vs. crawling vs. walking pre-school-age vs. school-age) and sex. The wet season for each study was defined as the 6 months of highest average rainfall, obtained from <https://www.weather-atlas.com/>.4 We assessed effect modification by examining the p-values on the interaction terms between the treatment and the indicator variable for the subgroup in the regression models; a p-value <0.2 was considered evidence of effect modification.

As sensitivity analyses, we compared estimates adjusted for potential confounders using parametric regression models with estimates from flexible machine-learning based targeted maximum likelihood estimation models,5 we compared associations between environmental contamination and child diarrhea observed within 4 months with diarrhea observed within a month and observed at any time, and we compared covariate adjusted estimates with unadjusted estimates.

All analyses were conducted in R 4.0.4, and analysis scripts are publicly available (<https://github.com/amertens/wash-ipd>). The systematic review search strategies and the analysis plan were pre-registered on Open Science Framework (<https://osf.io/8sgzn/>). Our PRISMA checklist can be found in Supplementary Table XXX.

## Results

The number of child diarrhea observations with time-matched environmental samples ranged from 210-2034 observations and diarrhea prevalence ranged from 6.1-25.9 across studies (Table 1). The number of HAZ observations with time-matched environmental samples ranged from 202-1800 observations and mean HAZ ranged from -1.90 to -1.33 (Table 1).

#### Associations with pathogen-specific infections

Detection of a specific enteropathogens in the compound environment was associated with increased prevalence of subsequent infection with the same pathogen in children living in the compound; trends were consistent across different enteropathogens and sample types (Figure 1). C. difficile, Ascaris and Trichuris detected in courtyard soil and Shigella and Trichuris detected in flies were significantly associated with increasing prevalence of infection with the same pathogens. Pathogenic E.coli and Giardia detected in soil and/or flies also had borderline associations with increasing infection prevalence. Few studies had time-matched data on water/hand samples and child infections; these studies found no association between pathogens measured in water or on hands and child infections with the same pathogen (Figure 1).

#### Associations with diarrhea

In general, enteropathogens and host-specific fecal markers in the environment were not associated with child diarrheal disease risk. The presence of any enteropathogen in any type of environmental sample was not associated with diarrheal disease, except for significantly increased diarrhea prevalence associated with any enteropathogen detection on child hands in WASH Benefits Bangladesh (Figure 2); we note that the only pathogen investigated in this study was rotavirus.6 When broken down by groups of pathogens, bacteria on child hands7 and protozoa in soil8 were also borderline associated with increasing risk of diarrhea (Figure S1). Most associations between specific pathogens in the environment and diarrhea were null, but rotavirus on child hands6 and Giardia in latrine soil8 were both significantly associated with an approximately two-fold increase in diarrhea risk (Figure S2). Detection of pathogenic E. coli on child hands7 and Ascaris, astrovirus and C. difficile in soil8,9 were also borderline associated with increased risk (Figure S2). Examining enteropathogen abundance in the environment revealed similar trends; increasing abundance of Ascaris and rotavirus in soil6,9 and rotavirus on child hands [Boehm 2016] was associated with increasing risk of diarrhea (Figure S3).

There was no significant associations between the presence of any MST marker or specific groups of MST markers (human, animal, general) and child diarrheal disease in any sample type (Figure 2, Figure S1 ), except for a borderline increase in diarrhea associated with detection of animal fecal markers.6 Among specific markers, detection of the avian marker GFD in soil was significantly associated with an over two-fold increase in diarrhea risk in Mozambique,10 and the same marker in stored water and on child hands was borderline associated with increased diarrhea in Bangladesh (Figure S4). Other general, human and animal markers were not associated with diarrhea.

#### Associations with child growth

The presence of any enteropathogen in any environmental sample was significantly associated with lower HAZ when pooled across studies (adjusted mean difference: -0.08 (95% CI: -0.15, -0.01), Figure 3). This was driven primarily by the number of slightly harmful but insignificant effects rather than by any strong effect in specific studies, with the exception of water samples with any enteropathogen presence being significantly associated with lower mean HAZ in Bangladesh6. When broken down by groups of enteropathogens, presence of viruses in stored water and protozoa on child hands was significantly associated with a reduction in HAZ on the order of 0.5 in Bangladesh6,7 and presence of bacteria in drinking water was borderline associated with a reduction in HAZ on the order of 0.15 in Bangladesh and India6,11 (Figure S1). Also, there was a borderline reduced HAZ associated with any pathogen in soil when pooled across studies. Individual pathogens whose detection was associated with reduced HAZ were Ascaris in soil and flies, E. histolytica in soil, Giardia on child hands and rotavirus in water (Figure S2). There were inconsistent associations between the abundance of specific enteropathogens and child HAZ, with most estimates having null effects, and with significant effects occurring in both harmful and protective directions (Figure S3). For other measures of growth, there were inconsistent associations between the presence or abundance of enteropathogens and WAZ, WHZ, stunting, underweight and wasting, with most estimates having null effects, and with significant effects occurring in both harmful and protective directions (Figures S1-S3) .

The presence of any MST in any environmental sample was not associated with HAZ when pooled across studies (Figure 3). Any MST presence in water was significantly associated with lower mean HAZ in Bangladesh6 , but any MST presence in soil was associated with higher mean HAZ in Mozambique.10 Among groups of MST markers, animal and general fecal markers in stored water were associated with lower HAZ in Bangladesh6 and human MST markers in water were borderline associated with lower HAZ in Mozambique10 but for most combinations of markers and sample types, the associations were null (Figure S1). There were inconsistent associations between the presence of specific MST markers and HAZ, with most estimates having null effects (Figure S4). Of the statistically significant associations, half of the sample-specific estimates were associated with increased linear growth and half were associated with decreased linear growth. There were inconsistent associations between the abundance of specific MST markers and child, with most estimates having null effects, and with significant effects occurring in both harmful and protective directions (Figure S5). For other measures of growth, there were inconsistent associations between the presence or abundance of any MST marker and WAZ, WHZ, stunting, underweight and stunting, with most estimates having null effects, and with significant effects occurring in both harmful and protective directions (Figures S1, S4, S5).

#### Subgroup analyses

Pooled across studies, there was a significant increase in child diarrheal disease risk in compounds with any sample with any enteropathogen detected when the child diarrheal disease occurred during the wet season (Figure S6). Diarrheal disease was too sparse to estimate differences in associations between households with and without animals, and there was no difference in associations when analyses were stratified by child age or sex (Figures S8,S10). There was no association with between MST markers and diarrhea in either season.

Pooled across studies, there was a significant decrease in child HAZ in compounds with any sample with any enteropathogen detected when the child lives in a compound with no animals, but not when animals were in the compound (Figure S7) . There was no consistent effect of child age on associations between environmental pathogens and HAZ (Figure S11) . Pooled across studies, pathogen presence was associated with twice the reduction in HAZ in boys (adjusted mean difference: -0.08 (95% CI: -0.16, -0.01)) than in girls (adjusted mean difference: -0.18 (95% CI: -0.30, -0.06)). The decrease in HAZ associated with the presence of any pathogen in any environmental sample was higher among boys than in most individual studies, though the difference was only significant in Kenya9 (Figure S9).

There was no consistent effect of child age on associations between MST markers and HAZ (Figure S11). There were also no significant differences in pooled estimates between the one urban study (Holcomb et al. 2020) and the four rural studies.

#### Sensitivity analyses

Lastly, the choice to use a maximum of four months as a time window between sampling and diarrheal disease did not obscure major trends in the associations between any enteropathogen or any MST marker in environment and diarrheal disease, as results were similar when using all diarrhea observations or when only using diarrhea observations within a month of the environmental sampling date. Most associations were null regardless of the time window.

Comparison between associations estimated with generalized linear models (GLM) and machine-learning based targeted likelihood estimation models (TMLE) for the diarrhea outcome. The estimation approach chosen did not affect our conclusions about associations between environmental contamination and diarrheal disease or HAZ. Associations between adjustment covariates and the presence of different enteropathogen and MST markers in different environmental samples. Most covariates were not strongly associated with enteropathogen or MST marker presence in the environment, meaning they were not strong confounders of the relationship between environmental contamination with enteropathogen or MST markers and child infections or poor growth. Measures of household wealth generally had the strongest association with environmental contamination, though the association varied by study, sample, and microbial target. Between the low association between covariates and environmental contamination, and the generally limited differences between unadjusted, adjusted, and TMLE estimates, we believe our modeling approach adequately adjusted for measured confounding, but unmeasured confounding may bias the results. The average associations were slightly larger in magnitude after covariate adjustment. On average the covariate adjustment had small effects on the results though it was slightly greater when a larger number of covariates were used for adjustment.

## Discussion

Pathogens in the environment that had associations with increased risk of child diarrhea included rotavirus, Giardia and, to a smaller extent, pathogenic E. coli, Ascaris, astrovirus and C. difficile. These findings are consistent with multi-country case-control studies that have identified rotavirus, pathogenic E. coli and astrovirus among the pathogens with the highest attributable burden of diarrhea .12,13 Other dominant pathogens in these studies included Cryptosporidium, Shigella, Campylobacter and norovirus; we note that we did not have sufficient time-matched data to estimate associations between detection of these pathogens in the environment and child diarrhea. The avian marker GFD was the only MST marker associated with diarrhea; this is consistent with growing evidence that poultry are a major source of diarrhea transmission.14  
Soil and child hands stood out as dominant pathways of environmental diarrhea transmission. Child hands have been identified a major source of children’s fecal exposure, in terms of frequency of mouth contacts[Kwong 2016] ,15 estimated E.coli ingestion16 and associations with diarrhea.17 Similarly, soil has been shown to account for a significant portion of estimated E. coli ingestion for children18 and ingestion of soil has been associated with environmental enteric dysfunction and stunting in children.19,20 Our findings corroborate the role of child hands and soil in diarrhea transmission.

Enteropathogen presence on child hands in Boehm et al. 2016 might be significantly associated with diarrhea while other estimates are not for several reasons, Boehm et al. 2016 is the only included study where diarrhea was measured <1 week after env sampling, the only pathogen it measured was rotavirus, so aggregate enteropathogens only includes rotavirus (the most common diarrheal pathogen in young children worldwide) for this study.

These findings highlight the limitations of self-reported all-cause diarrhea as an outcome to assess the impact on environmental pathogen contamination. This is consistent with research on pathogens in recreational water, where specific enteropathogens had limited association with all-cause diarrhea.21 Additionally, the link between pathogens in the environment and in child stool samples provides a link in the causal chain between environmental contamination and lower child HAZ, though the analysis was limited by the small number of pathogens measured in both the environment and children. Small non-significant effects on HAZ became significant when pooled, highlighting the strength of IPD meta-analyses.

Due to the smaller sample size of the environmental samples within the WASH trials and quasi-randomized studies, the rarity of diarrheal disease in children, and the rarity of many of the enteropathogen in environmental samples, data sparsity affected what was possible in this analysis. Many exposure-outcome associations were not estimated due to data sparsity, and others were estimated but could only be adjusted for a subset of potential confounders. There was generally no large differences between unadjusted and adjusted estimates, possibly because the measured covariates were not strongly associated with enteropathogen or MST presence in the environment, and so there may be unmeasured confounding. We did not correct for multiple comparisons, and so some significant associations are likely type-1 errors, especially when inconsistent with results from other studies or sample types. The inconsistency in the length of time between environmental and child health measurements across different studies may have also led to larger inconsistencies in associations between studies. Future studies investigating the associations between environmental fecal contamination and child health should ascertain health outcomes soon after environmental sampling, and should both focus on enteropathogens in the environment instead of MST markers, and on enteropathogen specific infections instead of all cause diarrhea.

## References

1. Zou, G. A modified poisson regression approach to prospective studies with binary data. *American Journal of Epidemiology* **159**, 702–706 (2004).

2. Freedman, D. A. On The So-Called ‘Huber Sandwich Estimator’ and ‘Robust Standard Errors’. *The American Statistician* **60**, 299–302 (2006).

3. Cochran, W. G. The Combination of Estimates from Different Experiments. *Biometrics* **10**, 101–129 (1954).

4. d.o.o, Y. M. G. Weather Atlas | Weather forecast and Climate information for cities all over the Globe. *Weather Atlas*.

5. Schuler, M. S. & Rose, S. Targeted Maximum Likelihood Estimation for Causal Inference in Observational Studies. *American Journal of Epidemiology* **185**, 65–73 (2017).

6. Boehm, A. B. *et al.* Occurrence of Host-Associated Fecal Markers on Child Hands, Household Soil, and Drinking Water in Rural Bangladeshi Households. *Environmental Science & Technology Letters* **3**, 393–398 (2016).

7. Fuhrmeister, E. R. *et al.* Effect of Sanitation Improvements on Pathogens and Microbial Source Tracking Markers in the Rural Bangladeshi Household Environment. *Environmental Science & Technology* **54**, 4316–4326 (2020).

8. Capone, D. *et al.* Impact of an Urban Sanitation Intervention on Enteric Pathogen Detection in Soils. *Environmental Science & Technology* **55**, 9989–10000 (2021).

9. Steinbaum, L. *et al.* Effect of a sanitation intervention on soil-transmitted helminth prevalence and concentration in household soil: A cluster-randomized controlled trial and risk factor analysis. *PLoS neglected tropical diseases* **13**, e0007180 (2019).

10. Holcomb, D. A. *et al.* Human fecal contamination of water, soil, and surfaces in households sharing poor-quality sanitation facilities in Maputo, Mozambique. *International Journal of Hygiene and Environmental Health* **226**, 113496 (2020).

11. Reese, H. Effectiveness of a combined sanitation and household-level piped water intervention on infrastructure coverage, availability and use, environmental fecal contamination, and child health in rural Odisha, India: A matched cohort study. (Emory University, 2017).

12. Kotloff, K. L. *et al.* Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): A prospective, case-control study. *The Lancet* **382**, 209–222 (2013).

13. Platts-Mills, J. A. *et al.* Pathogen-specific burdens of community diarrhoea in developing countries: A multisite birth cohort study (MAL-ED). *The Lancet Global Health* **3**, e564–e575 (2015).

14. Ercumen, A. *et al.* Poultry Ownership Associated with Increased Risk of Child Diarrhea: Cross-Sectional Evidence from Uganda. *The American Journal of Tropical Medicine and Hygiene* **102**, 526–533 (2020).

15. Kwong, L. H. *et al.* Hand- and Object-Mouthing of Rural Bangladeshi Children 3-18 Months Old. *International Journal of Environmental Research and Public Health* **13**, E563 (2016).

16. Mattioli, M. C. M., Davis, J. & Boehm, A. B. Hand-to-mouth contacts result in greater ingestion of feces than dietary water consumption in Tanzania: A quantitative fecal exposure assessment model. *Environmental Science & Technology* **49**, 1912–1920 (2015).

17. Pickering, A. J. *et al.* Fecal Indicator Bacteria along Multiple Environmental Transmission Pathways (Water, Hands, Food, Soil, Flies) and Subsequent Child Diarrhea in Rural Bangladesh. *Environmental Science & Technology* **52**, 7928 (2018).

18. Kwong, L. H. *et al.* Effect of sanitation improvements on soil-transmitted helminth eggs in courtyard soil from rural Bangladesh: Evidence from a cluster-randomized controlled trial. *bioRxiv* 2020.09.29.318097 (2020) doi:[10.1101/2020.09.29.318097](https://doi.org/10.1101/2020.09.29.318097).

19. George, C. M. *et al.* Geophagy Is Associated with Environmental Enteropathy and Stunting in Children in Rural Bangladesh. *The American Journal of Tropical Medicine and Hygiene* **92**, 1117–1124 (2015).

20. Morita, T. *et al.* Mouthing of Soil Contaminated Objects is Associated with Environmental Enteropathy in Young Children. *Tropical medicine & international health: TM & IH* **22**, 670–678 (2017).

21. Griffith, J. F. *et al.* Epidemiologic evaluation of multiple alternate microbial water quality monitoring indicators at three California beaches. *Water Research* **94**, 371–381 (2016).