Associations between enteropathogens detected in the environment and child growth and enteric infections: 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)

## Summary

**Background:** Water, sanitation and hygiene (WASH) improvements aim to reduce childhood enteric infections, subsequent diarrheal disease and growth faltering by reducing child exposure to fecal-orally transmitted pathogens in the environment. Until recently, trials of WASH interventions have primarily focused on documenting health outcomes such as caregiver-reported diarrhea without quantifying intermediate outcomes along the causal chain, such as detection of pathogens in environmental samples and in human biological specimens. Without measuring these causal intermediates, trials are limited to a “black box” understanding, where underlying mechanisms of interventions are unknown and investigators can only speculate about reasons for intervention success or failure. Inspecting the causal chain is especially important given the small or null effects on child diarrhea and growth achieved in recent well-conducted trials of WASH interventions. Additionally, studies that have collected environmental samples to date have primarily focused on drinking water (and, to a smaller extent, hands and food) while other pathways such as soil and surfaces in the domestic environment have received less attention. Furthermore, most studies have relied on measuring indicator organisms in the environment as a proxy for pathogens; these indicators can originate from non-fecal sources, cannot differentiate between human vs. animal fecal contamination, and correlate poorly with the presence of pathogen. Recent advances in DNA-based diagnostics now allow detection of a range of enteropathogens in human biological specimens and environmental samples, as well as distinction between human vs. animal fecal sources through molecular source tracking (MST) markers. We conducted an individual participant data (IPD) meta-analysis to pool data from WASH trials that have measured pathogens and/or MST markers in the environment to explore associations between pathogen presence in the environment and the gut, and any subsequent diarrheal infections and child growth sequelae,

**Methods:** We conducted a systematic review and individual participant data meta-analysis to estimate enteropathogens and microbial source tracking (MST) markers in environmental samples. We used covariate-adjusted regression models with robust standard errors to estimate associations between. We estimated associations between aggregate and specific pathogens or MST markers and child diarrheal disease, height-for-age Z-scores (HAZ), weight-for-age Z-scores, weight-for-height Z-scores, and stunting, wasting, and underweight statuses. We used all growth outcomes measured after environmental sampling, but only used observations of diarrheal disease taken within four months of the date of environmental sampling. We also estimated associations stratified by child age, child sex, season, and animal presence in the household.

**Findings:** We identified and received data from 9 sub-studies from five randomized or quasi-experimental studies. There was a small but consistent association between households with any pathogen contamination and children with lower HAZ (Adjusted mean difference pooled across studies: -0.08 (95% CI: -0.15, -0.01)), but there was no association with MST markers. The association between environmental pathogen contamination and HAZ was stronger for boys than girls. Associations were inconsistent for other growth outcomes. There was no consistent association between aggregate or individual enteropathogens or MST markers and caregiver reported diarrheal disease, though there was a significant increase in diarrheal disease risk in households with any pathogen presence in any sample during wet seasons when pooled across studies. Additionally specific pathogens in the environment were consistently associated with an increased risk of child infections with the same pathogen.

**Interpretation:** No other studies have examined the association between enteropathogens and MST markers in environmental samples and child health outcomes in multiple low-resource settings and WASH contexts. While we found no association between environmental contamination and diarrheal disease, children had a higher risk of pathogen infections if the pathogen was found in environmental samples, highlighting the limitation of all-cause diarrheal disease as an outcome to assess the impact of WASH interventions to reduce environmental pathogen contaminations. MST markers did not have an association with child growth, but pathogen presence in the environment was associated with lower linear growth, especially among boys, though associations with specific pathogens were inconsistent. These results may have been affected by unmeasured confounding as well as the variable lengths of time between environmental sampling and child health ascertainment across the different studies.

## Methods

We examined associations between prevalence of enteropathogens and MST markers in the environment and child health outcomes, including all-cause diarrheal disease, child growth, and enteropathogen-specific infections. The primary outcomes for all exposures were caregiver recall of diarrheal disease and child height-for-age Z-scores. For specific enteropathogen presences in the environment, primary outcomes also included the corresponding enteropathogen detection in child stool. Secondary outcomes include z-scores for weight-for-age (WAZ) and weight-for-length (WLZ) and prevalence of stunting, wasting and underweight. For the growth outcomes outcomes, we considered all environmental samples collected over the child’s lifetime prior to the anthropometry measurement. For the diarrheal disease and enteropathogen-specific infection outcomes, we will only consider environmental samples collected up to four months before the measurement of the health outcome.

For binary outcomes, we estimated prevalence ratios using modified Poisson regressions.1 For continuous outcomes (child anthropometry Z-scores), we used linear regressions to estimate adjusted 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 strata of the exposure.

## Results

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. Table 1 shows the number of health observations with corresponding environmental samples, the prevalence of diarrheal disease and measures of child growth failure, and mean child anthropometry Z-scores by study.

In general, enteropathogen and host-specific fecal markers were not associated with child diarrheal disease risk. The presence of any enteropathogen or any MST marker in any environmental samples were not associated with diarrheal disease, except any enteropathogen presence located on child hands in Boehm et al. 2016 (Figure 1). 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, and in general child hands are a major source of fecal exposure for kids, both in terms of estimated E.coli ingestion3 and associations between E. coli and diarrhea.4 There was no significant associations between specific groups of MST markers and child diarrheal disease in any sample type (Figure 4). Most associations between specific pathogens in the environment and child diarrheal disease were null, but *Giardia* in latrine soil in Capone et al. 2021 and Rotavirus on child hands in Boehm et al. 2016 were both significantly associated with approximately twice the risk of diarrheal disease in children (Figure 5). Specific MST markers in environmental samples were generally not associated with the risk of diarrheal disease but the avian MST marker GSD was significantly associated with an increased diarrheal disease risk in Mapsan latrine soil samples, and had non-significant but positive association in all three sample types from Boehm et al 2016 (Figure 6). There are inconsistent associations between the abundance of specific enteropathogens and MST markers and child diarrheal disease, with most estimates having null effects, and with significant effects occurring in both harmful and protective directions (Figures 7-8).

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 9), when diarrheal disease risk is elevated, potentially due to rainfall flushing environmental pathogens into contact with children.5,6 There was no association with MST markers in either season. 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 sex or age (Figures 11,13). 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.

The presence of any enteropathogen (but not any MST marker) in any environmental sample was significantly associated with lower height-for-age Z-scores (HAZ) when pooled across studies (Adjusted mean difference: -0.08 (95% CI: -0.15, -0.01), Figure 2). This was driven primarily by the number of slightly harmful but insignificant effects rather than by any strong effect of any enteropathogen in specific studies or sample types, highlighting the strength of IPD meta-analyses. Nevertheless, water samples with any enteropathogen presence were also significantly associated with lower mean HAZ in Boehm et al. 2016. Also, there was a borderline reduced HAZ associated with any pathogen in soil when pooled across studies. Any MST presence in water was also significantly associated with lower mean HAZ in Boehm 2016, but was associated with higher mean HAZ in latrine soil samples in Holcomb et al. 2020.

When separated out by groups of enteropathogen, enteropathogen presence in any environmental sample is no longer significantly associated with lower HAZ when pooled across studies, though sample positivity with any STH has a pooled borderline association with reduced HAZ (Figure 4). Additionally, any virus presence on child hands was significantly associated with lower mean HAZ in Boehm et al. 2016 and any protozoa in water was significantly associated with lower mean HAZ in Furhmeister et al. 2020, both from the WASH Benefits Bangladesh trial. There is also a general trend of presence of groups of MST markers in water samples being associated with lower mean HAZ (Figure 4). There were inconsistent associations between the presence of specific enteropathogens or specific MST markers and child HAZ scores, with most estimates having null effects (Figures 5-6). 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 are inconsistent associations between the abundance of specific enteropathogens and MST markers and child HAZ scores, with most estimates having null effects, and with significant effects occurring in both harmful and protective directions (Figures 7-8).

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 10). There was no association with MST markers. There was no consistent relationship between child age and the strengths associations between environmental pathogens or MST markers and HAZ (Figure 14), but there appears to be a sex difference. The presence of any pathogen in any environmental sample was associated with a larger decrease in height-for-age Z-scores among boys than among girls in most studies, though the difference was only significant in Steinbaum 2019 (Figure 12). Pooled across studies, pathogen presence was associated with twice the reduction in Z-scores 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)).

For other measures of growth, there were inconsistent associations between the presence or abundance of any enteropathogen or any MST marker and WHZ, WAZ, stunting, and underweight, with most estimates having null effects, and with significant effects occurring in both harmful and protective directions (Figures 4-8).

In contrast to the lack of associations with diarrhea, there was a general trend of positive associations between specific enteropathogens in the compound environment and an increased risk of the same enteropathogen infecting the child living in the compound across different enteropathogens and sample types (Figure 3). *Giardia* and enteropathogenic *E. coli* were two enteropathogens without significant associations between environmental presence and child infection, but associations were borderline significant for the two Capone et al. studies. Associations were significant or near significant for *Shigella*, *Ascaris*, and *Trichuris* presence in soil and fly samples and infections across multiple studies. *C. difficile* was only measured in latrine soil in Capone et al. 2021, but had the strongest association with infections in the children among specific enteropathogens. These findings highlight the limitations of self-reported all-cause diarrhea as an outcome to assess the impact on environmental pathogen contamination. 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.

We conducted several sensitivity analyses. 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. There were also no significant differences in pooled estimates between the one urban study (Holcomb et al. 2020) and the four rural studies.

# Primary figures

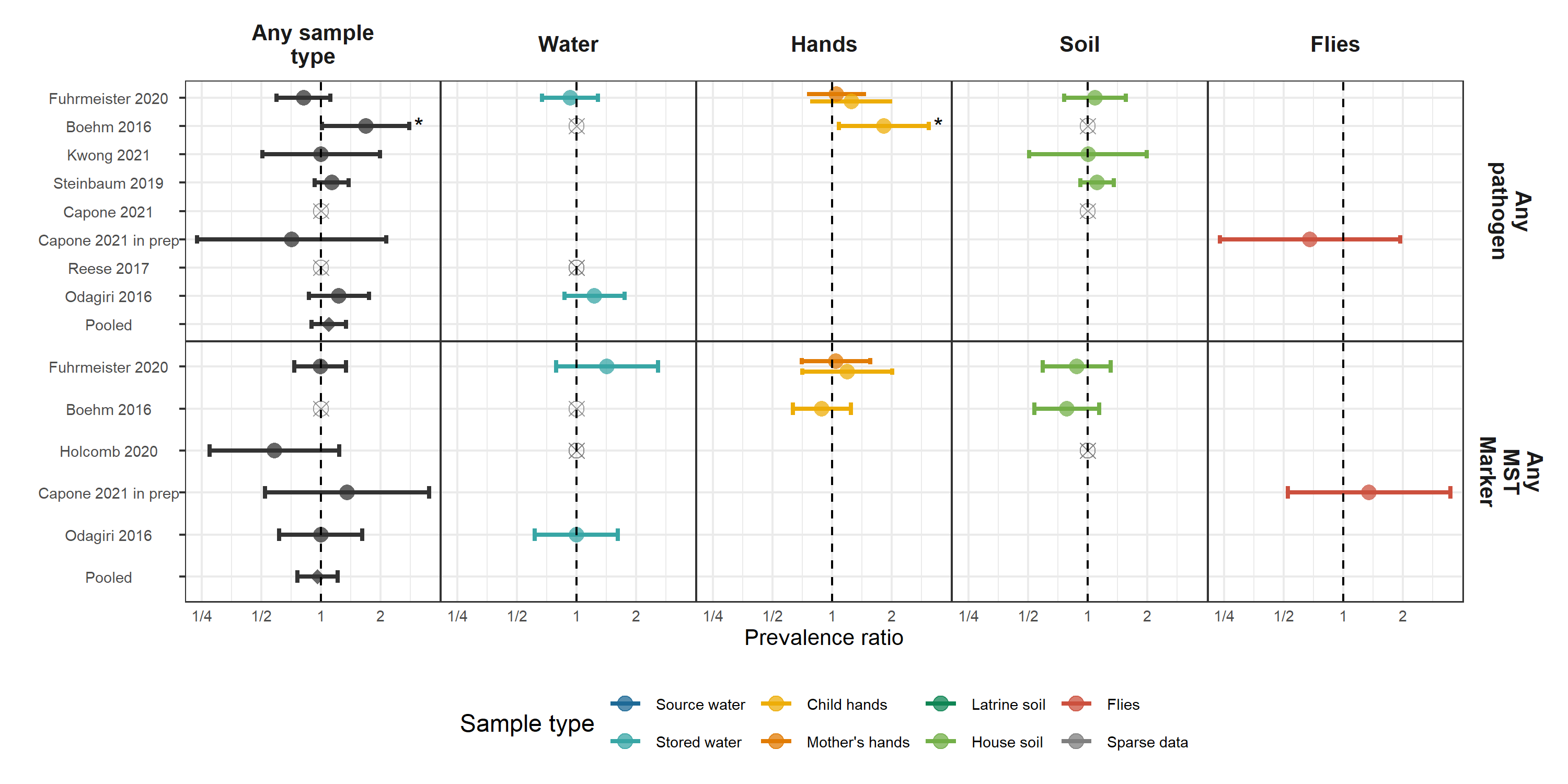


Figure 1: Forest plots of associations between child diarrheal disease and the prevalence of any enteropathogen or any MST markers in different types of environmental samples. Pooled estimates are presented when there are four or more study-specific estimates for a specific sample type and target combination and are denoted with diamond-shaped points. Grey crossed points denote data that were too sparse to estimate a prevalence ratio (i.e., <10 positive or negative observations). Samples of the same type from different locations (source vs. stored water, flies in kitchen vs. latrine, soil from courtyard vs. latrine) or different individuals (child vs. mother’s hands) are plotted separately. Asterisks above estimates denote statistical significance (\*= P-value < 0.05, \*\*= P-value < 0.01, \*\*\*= P-value < 0.001). All estimates are adjusted for potential confounders.

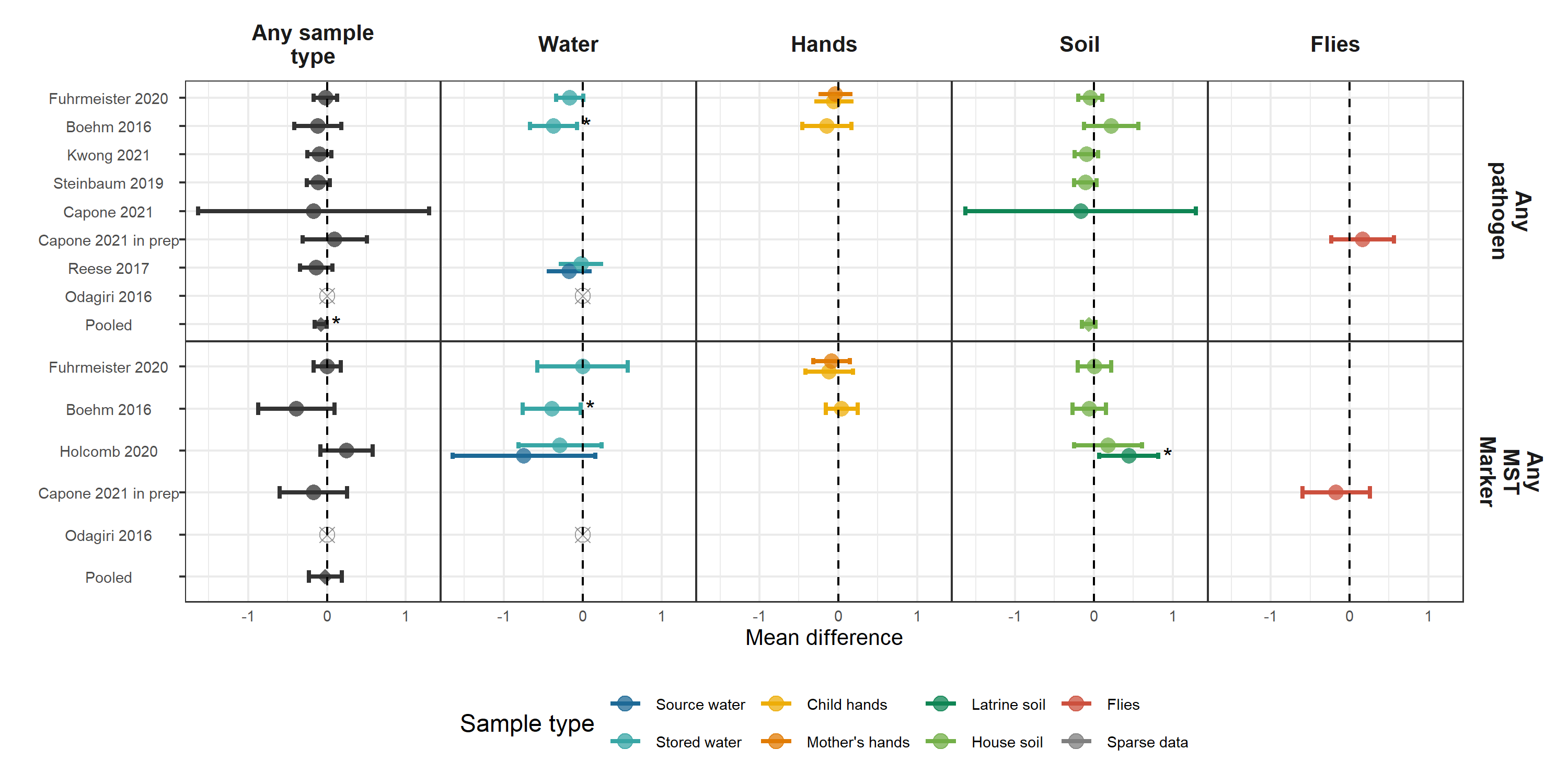


Figure 2: Forest plots of associations between child HAZ and the prevalence of any enteropathogen or any MST markers in different types of environmental samples. Pooled estimates are presented when there are four or more study-specific estimates for a specific sample type and target combination and are denoted with diamond-shaped points. Grey crossed points denote data that were too sparse to estimate a mean difference. Samples of the same type from different locations (source vs. stored water, flies in kitchen vs. latrine, soil from courtyard vs. latrine) or different individuals (child vs. mother’s hands) are plotted separately. All estimates are adjusted for potential confounders.

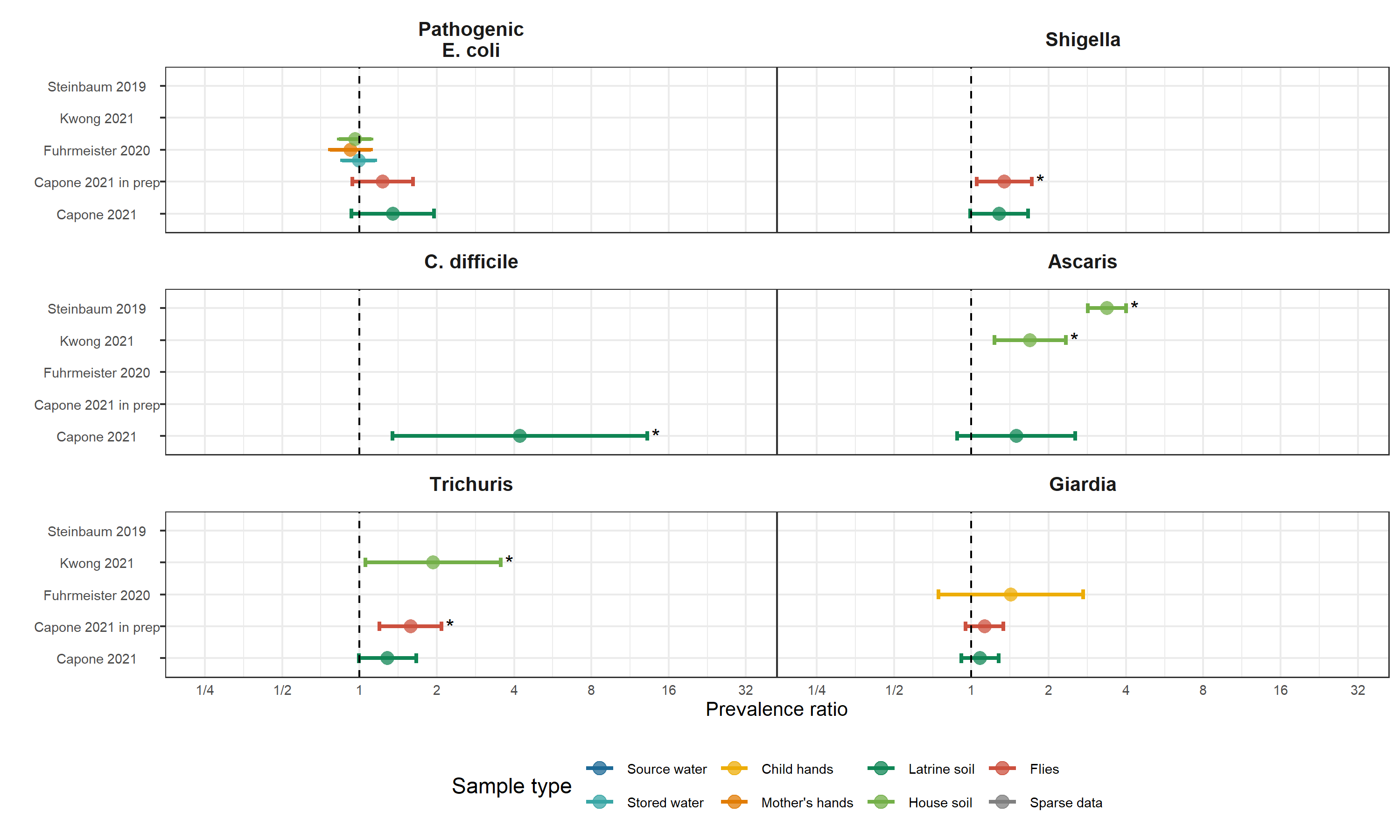
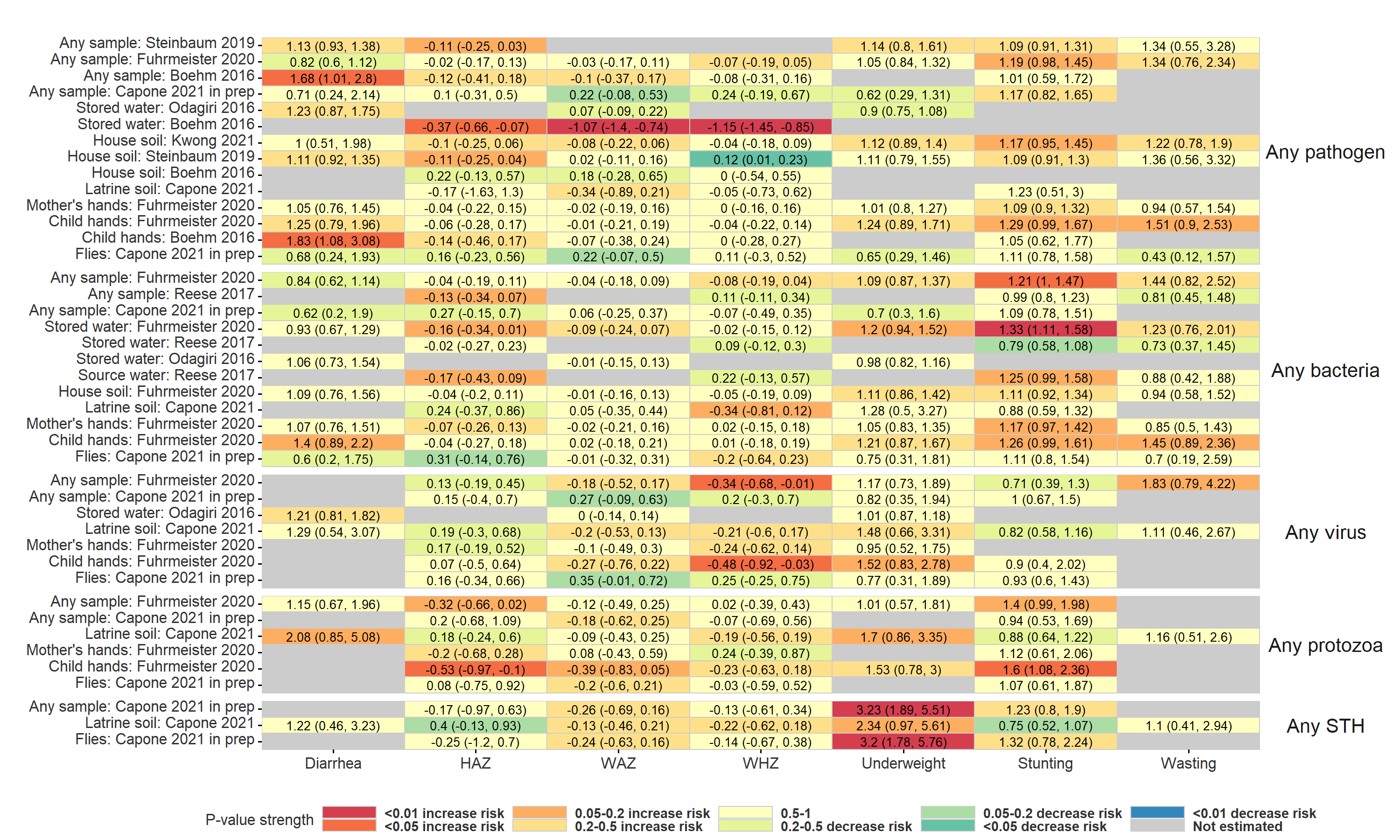


Figure 3: Forest plots of associations between specific enteropathogens in environmental samples and child infections with the same enteropathogens. Samples of the same type from different locations (source vs. stored water, flies in kitchen vs. latrine, soil from courtyard vs. latrine) or different individuals (child vs. mother’s hands) are plotted separately and denoted by different colors. All estimates are adjusted for potential confounders.

# Supplimentary figures



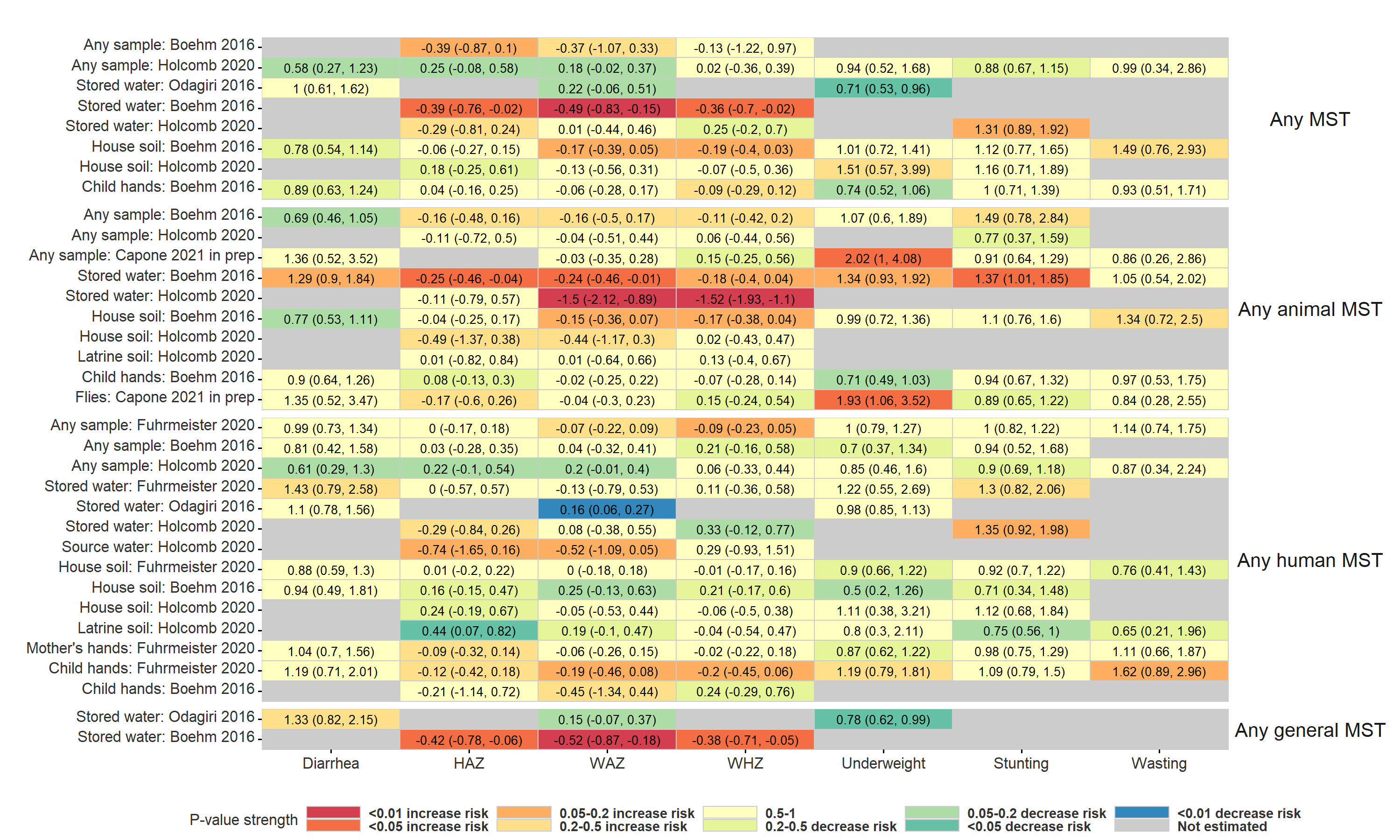
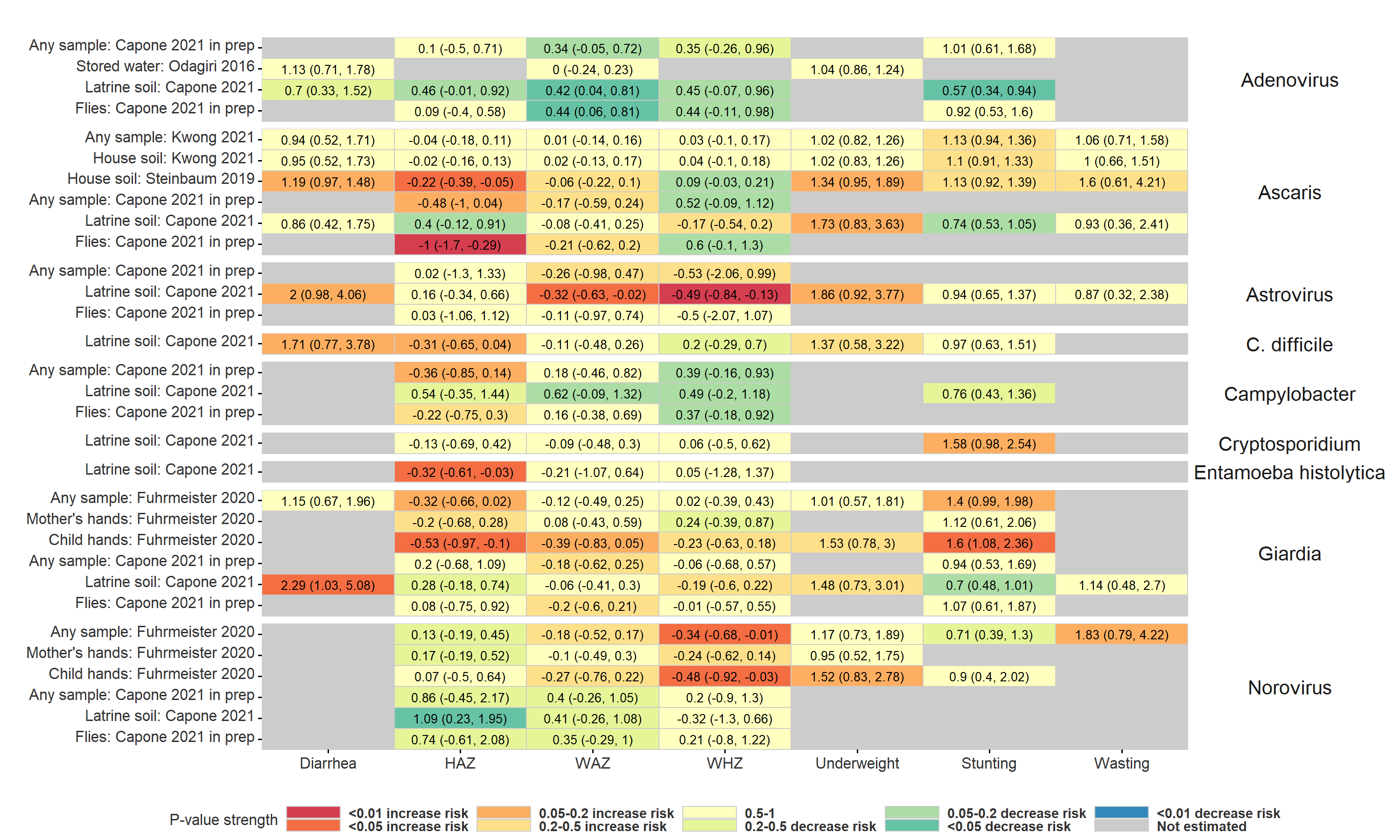


Figure 4: Heatmap of significance and direction of associations between aggregate measures of environmental contamination and child diarrhea and growth outcomes. Cells are colored by the strength of significance and direction of association, and the point estimate and confidence intervals are printed within cells, with relative risks printed for binary outcomes and mean differences for continuous outcomes. Each row is for a different sample type in a specific study, and pooled estimates are not plotted. Estimates aggregated across any sample type are only plotted if there are multiple sample types for a study. All estimates are adjusted for potential confounders.



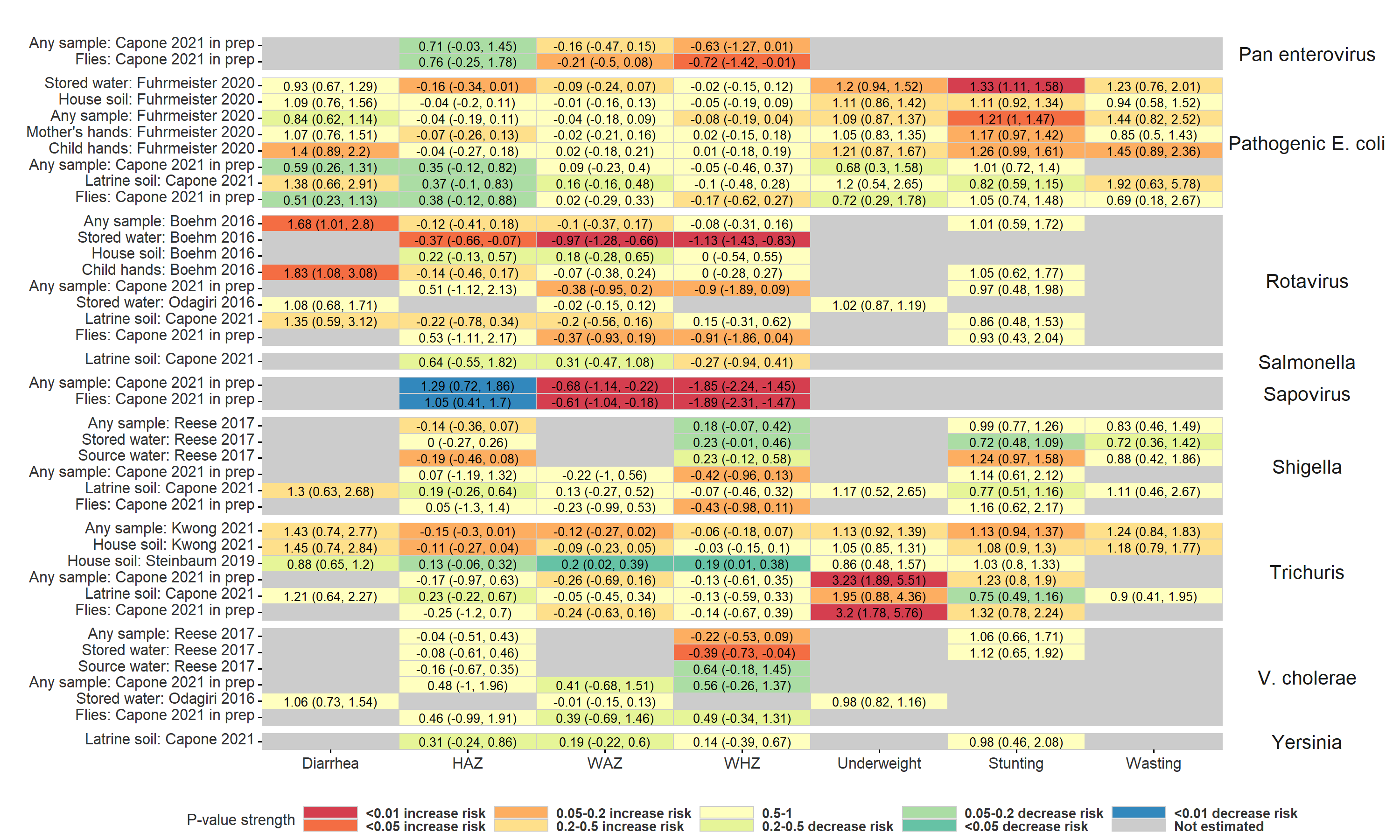


Figure 5: Heatmap of significance and direction of associations between specific pathogens in environmental samples and child diarrhea and growth outcomes. Cells are colored by the strength of significance and direction of association, and the point estimate and confidence intervals are printed within cells, with relative risks printed for binary outcomes and mean differences for continuous outcomes. Each row is for a different sample type in a specific study, and pooled estimates are not plotted. Estimates aggregated across any sample type are only plotted if there are multiple sample types for a study. All estimates are adjusted for potential confounders.

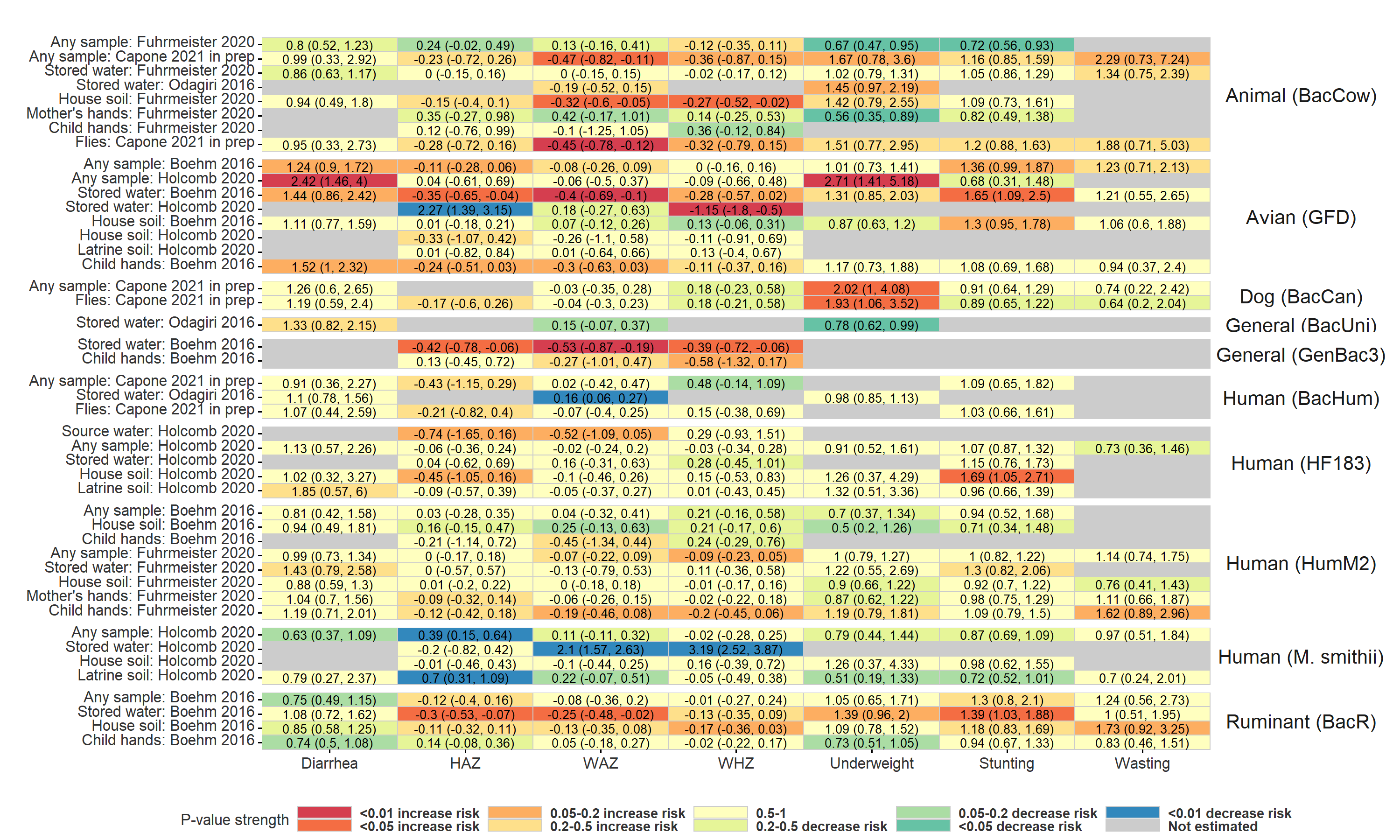


Figure 6: Heatmap of significance and direction of associations between specific microbial source tracking markers in environmental samples and child diarrhea and growth outcomes. Cells are colored by the strength of significance and direction of association, and the point estimate and confidence intervals are printed within cells, with relative risks printed for binary outcomes and mean differences for continuous outcomes. Each row is for a different sample type in a specific study, and pooled estimates are not plotted. Estimates aggregated across any sample type are only plotted if there are multiple sample types for a study. All estimates are adjusted for potential confounders.

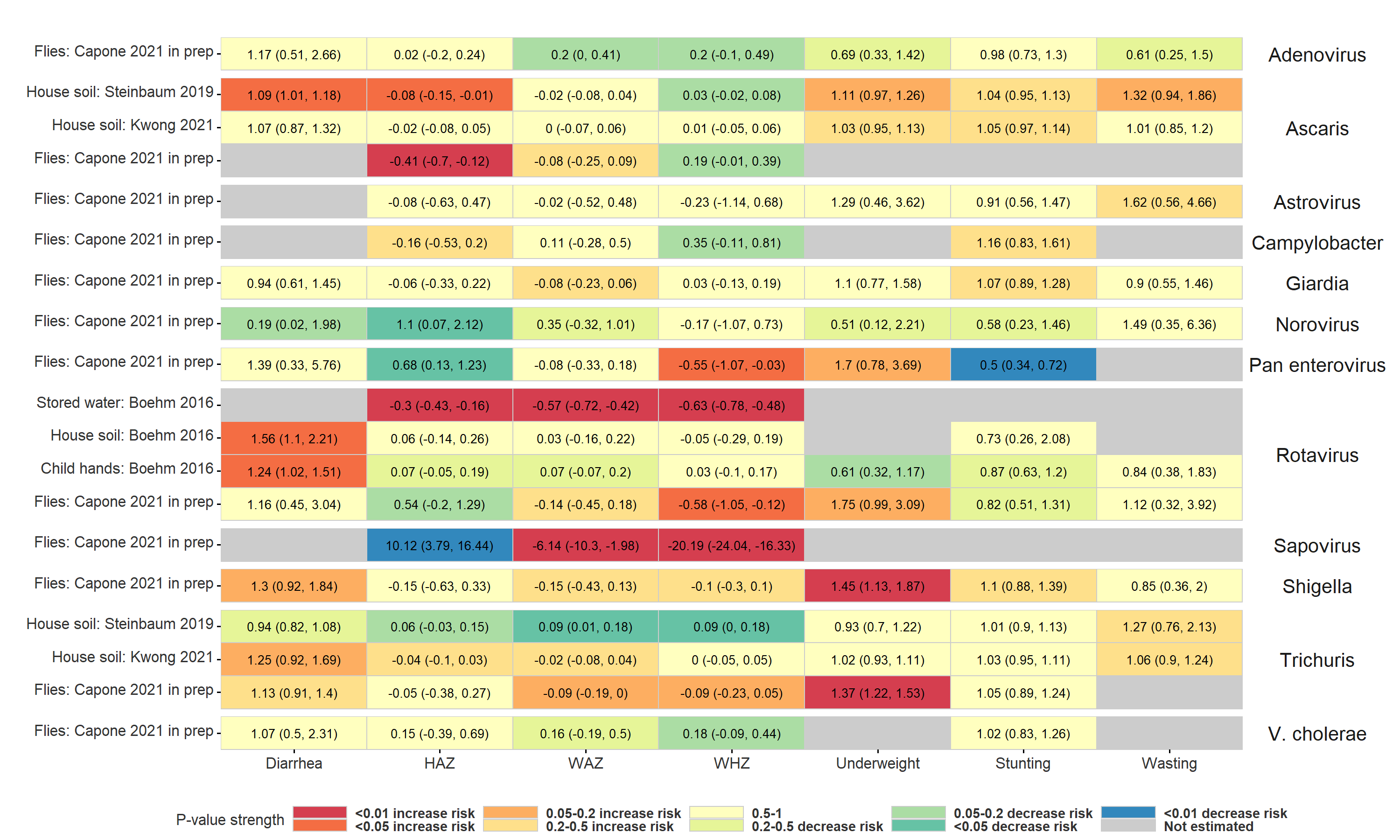


Figure 7: Heatmap of significance and direction of associations between the abundance of specific pathogens in environmental samples and child diarrhea and growth outcomes. Cells are colored by the strength of significance and direction of association, and the point estimate and confidence intervals are printed within cells, with relative risks printed for binary outcomes and mean differences for continuous outcomes. Each row is for a different sample type in a specific study, and pooled estimates are not plotted. Estimates aggregated across any sample type are only plotted if there are multiple sample types for a study. All estimates are adjusted for potential confounders.

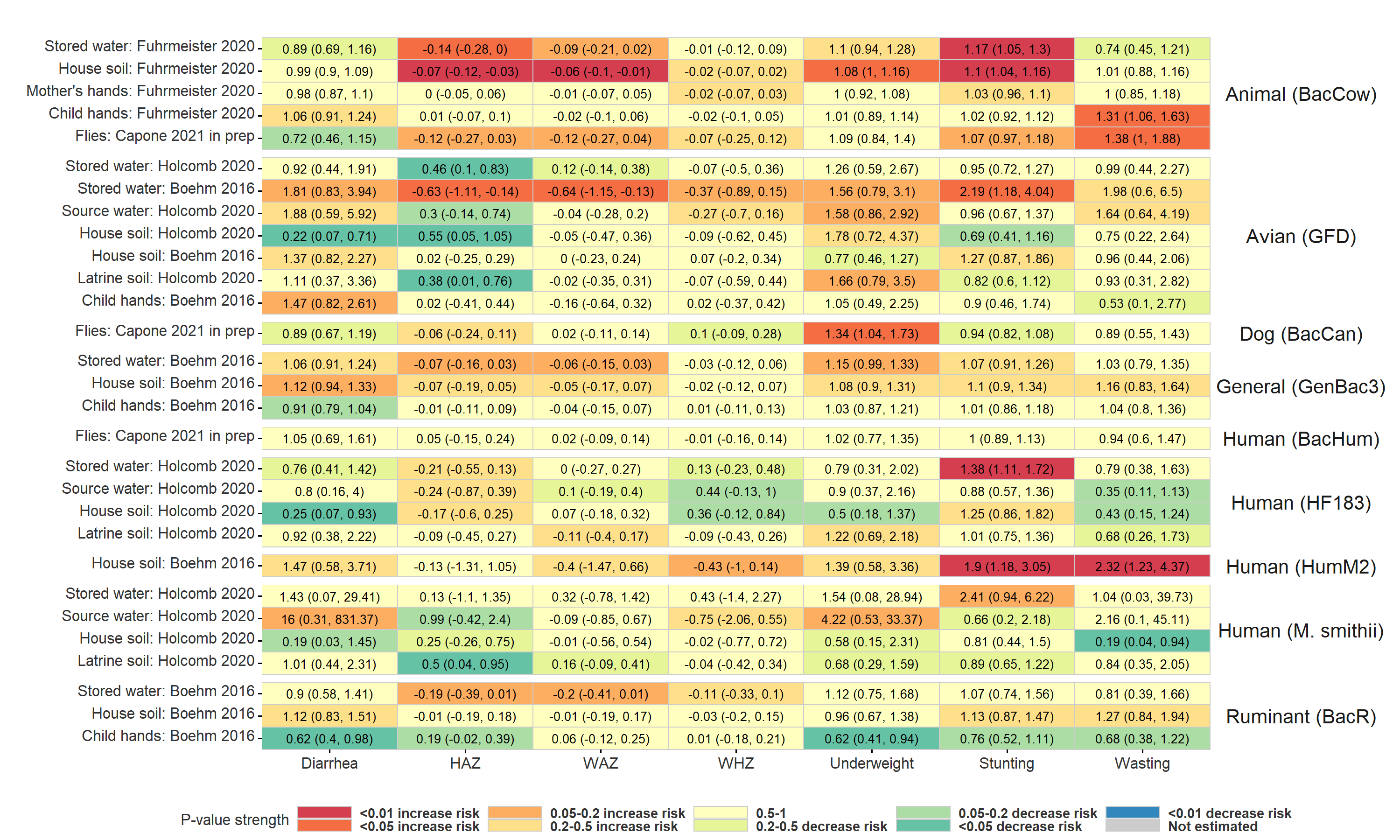


Figure 8: Heatmap of significance and direction of associations between the abundance of specific microbial source tracking markers in environmental samples and child diarrhea and growth outcomes. Cells are colored by the strength of significance and direction of association, and the point estimate and confidence intervals are printed within cells, with relative risks printed for binary outcomes and mean differences for continuous outcomes. Each row is for a different sample type in a specific study, and pooled estimates are not plotted. Estimates aggregated across any sample type are only plotted if there are multiple sample types for a study. All estimates are adjusted for potential confounders.

## Subgroup analysis

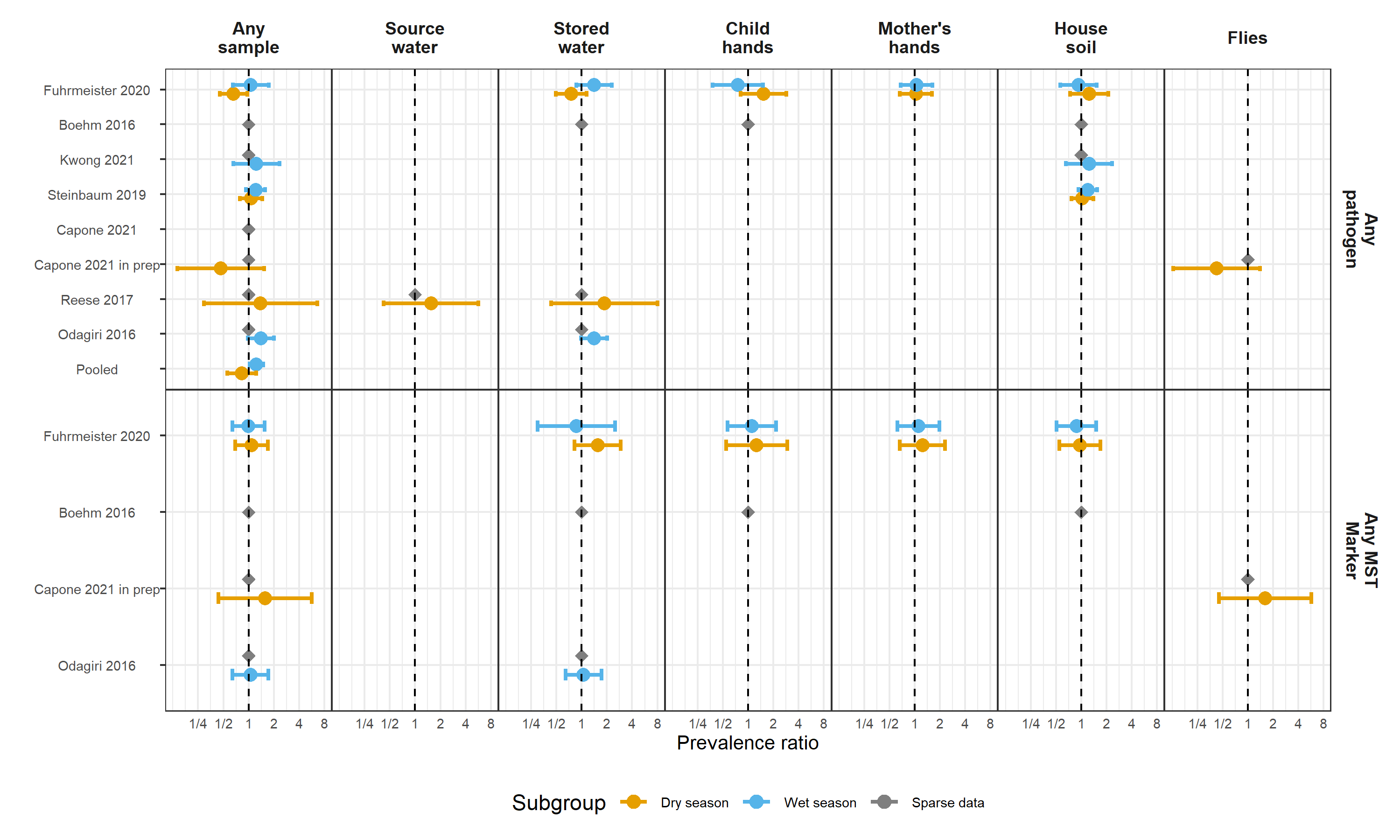


Figure 9: Forest plots of relative risks between diarrhea disease and any enteropathogen, and any MST markers, in different types of environmental samples, stratified by whether the diarrheal disease occurred during the wet versus dry season (defined by the 6 months of highest average rainfall). Significant effect modification, as determined by the p-values on the regression model interaction term, is marked with asterisks (P < 0.05 = \*, P < 0.01 = \*\*, P < 0.001 = \*\*\*).

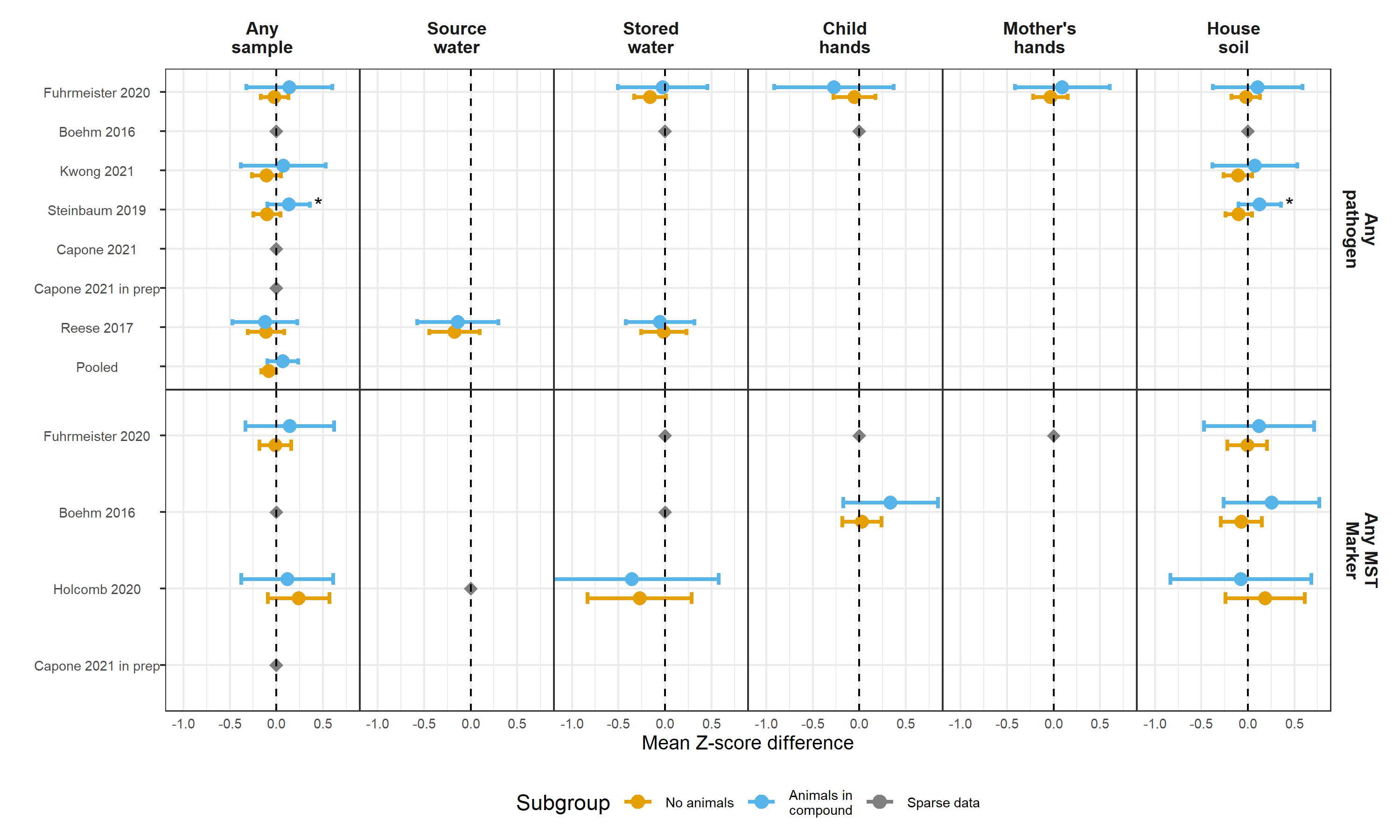


Figure 10: Forest plots of associations between child HAZ and any enteropathogen, and any MST markers, in different types of environmental samples, stratified by whether any animals were present in the compound. Significant effect modification, as determined by the p-values on the regression model interaction term, is marked with asterisks (P < 0.05 = \*, P < 0.01 = \*\*, P < 0.001 = \*\*\*).

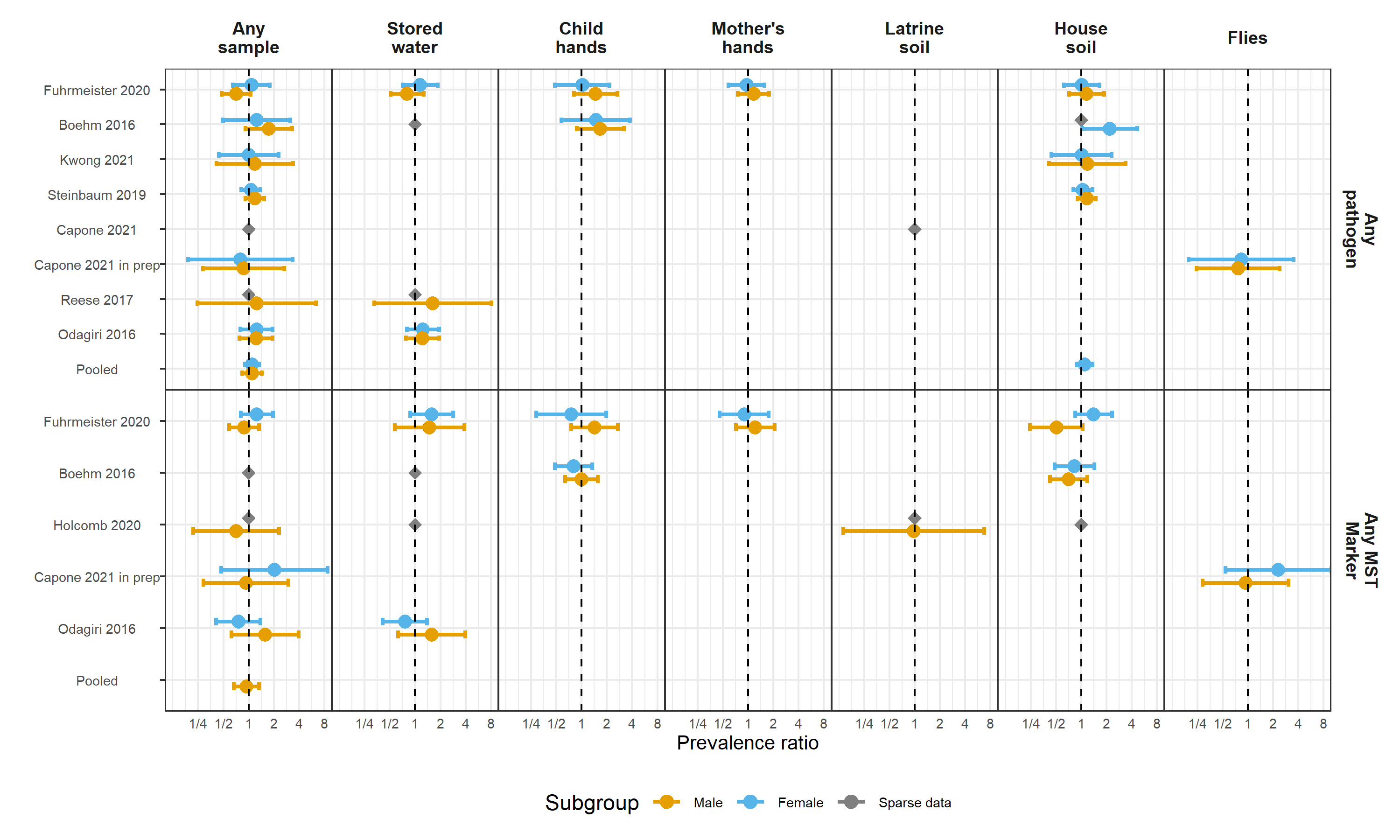


Figure 11: Forest plots of associations between child diarrheal disease and any enteropathogen, and any MST markers, in different types of environmental samples, stratified by child sex. Significant effect modification, as determined by the p-values on the regression model interaction term, is marked with asterisks (P < 0.05 = \*, P < 0.01 = \*\*, P < 0.001 = \*\*\*).

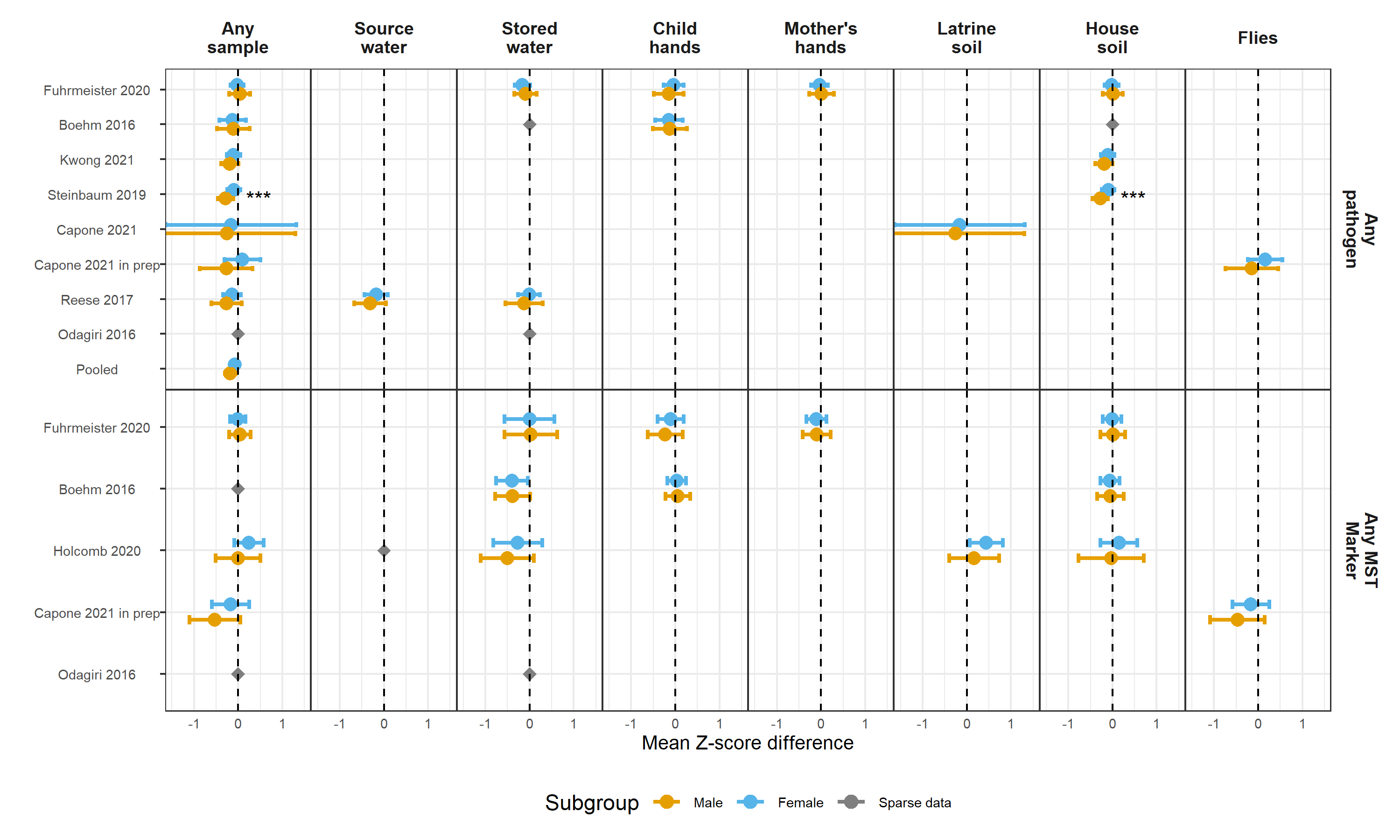


Figure 12: Forest plots of associations between child HAZ and any enteropathogen, and any MST markers, in different types of environmental samples, stratified by child sex. Significant effect modification, as determined by the p-values on the regression model interaction term, is marked with asterisks (P < 0.05 = \*, P < 0.01 = \*\*, P < 0.001 = \*\*\*).

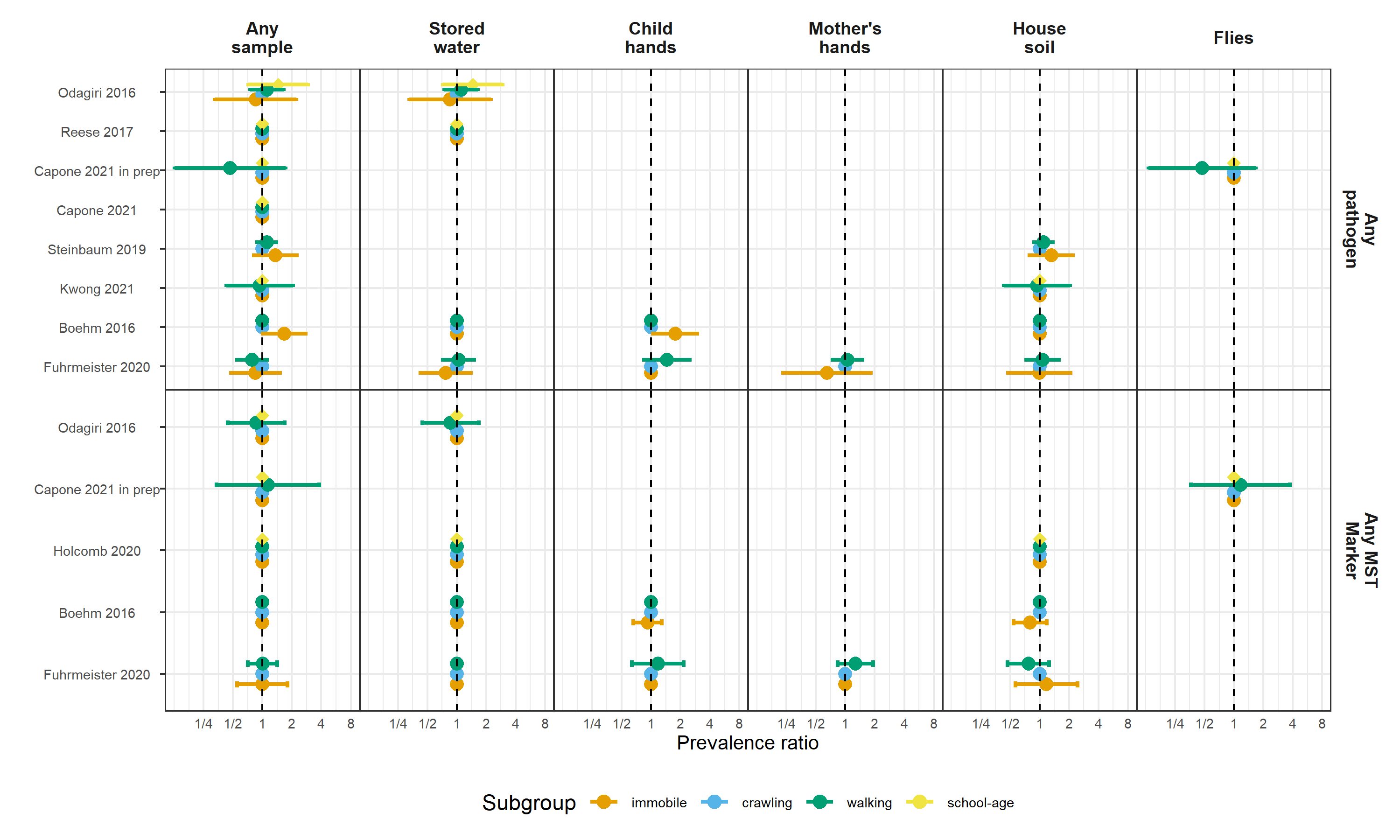


Figure 13: Forest plots of associations between child diarrheal disease and any enteropathogen, and any MST markers, in different types of environmental samples, stratified by child age,

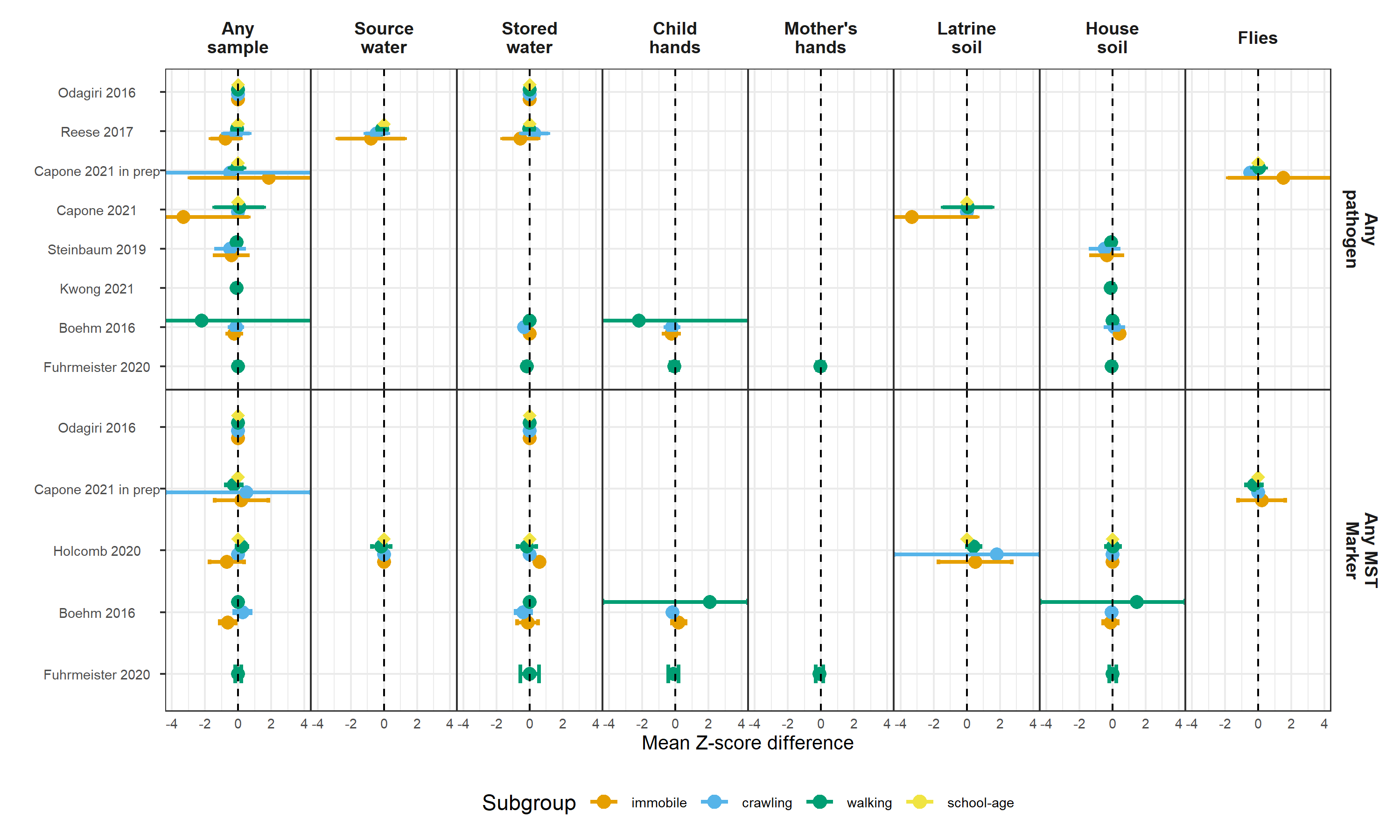


Figure 14: Forest plots of associations between child HAZ and any enteropathogen, and any MST markers, in different types of environmental samples, stratified by child age.

# Tables

| **Trial** | **Study** | **# diarrhea obs.** | **# diarrhea cases** | **Diarrhea prevalence** | **# HAZ obs.** | **Mean HAZ** | **Stunting prevalence** | **# WAZ obs.** | **Mean WAZ** | **Underweight prevalence** | **# WHZ obs.** | **Mean WHZ** | **Wasting prevalence** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Gram Vikas | Reese 2017 | 210 | 17 | 8.1 | 578 | -1.78 | 42.2 |  |  |  | 576 | -0.87 | 13.4 |
| MapSan | Holcomb 2020 | 227 | 20 | 8.8 | 202 | -1.81 | 48.5 | 199 | -0.68 | 11.6 | 203 | 0.22 | 7.9 |
| MapSan | Capone 2021 | 293 | 33 | 11.3 | 266 | -1.63 | 40.6 | 267 | -0.73 | 12.4 | 262 | 0.07 | 8.8 |
| MapSan | Capone 2021 in prep | 247 | 27 | 10.9 | 213 | -1.75 | 41.8 | 213 | -0.66 | 11.7 | 211 | 0.21 | 6.2 |
| Odisha | Odagiri 2016 | 2,036 | 188 | 9.2 |  |  |  | 4,152 | -1.38 | 29.1 |  |  |  |
| WBB | Fuhrmeister 2020 | 1,598 | 189 | 11.8 | 858 | -1.81 | 40.9 | 872 | -1.54 | 30.5 | 860 | -0.85 | 10.0 |
| WBB | Boehm 2016 | 459 | 105 | 22.9 | 458 | -1.33 | 24.9 | 459 | -1.32 | 23.3 | 459 | -0.72 | 8.9 |
| WBB | Kwong 2021 | 703 | 43 | 6.1 | 758 | -1.90 | 44.1 | 760 | -1.70 | 35.8 | 759 | -1.01 | 13.4 |
| WBK | Steinbaum 2019 | 1,913 | 496 | 25.9 | 1,800 | -1.54 | 31.6 | 1,852 | -0.73 | 9.7 | 1,797 | 0.10 | 1.5 |

## 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. 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).

4. 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).

5. Kraay, A. N. M. *et al.* Understanding the Impact of Rainfall on Diarrhea: Testing the Concentration-Dilution Hypothesis Using a Systematic Review and Meta-Analysis. *Environmental Health Perspectives* **128**, 126001.

6. Mertens, A. *et al.* Associations between High Temperature, Heavy Rainfall, and Diarrhea among Young Children in Rural Tamil Nadu, India: A Prospective Cohort Study. *Environmental Health Perspectives* **127**, 047004.