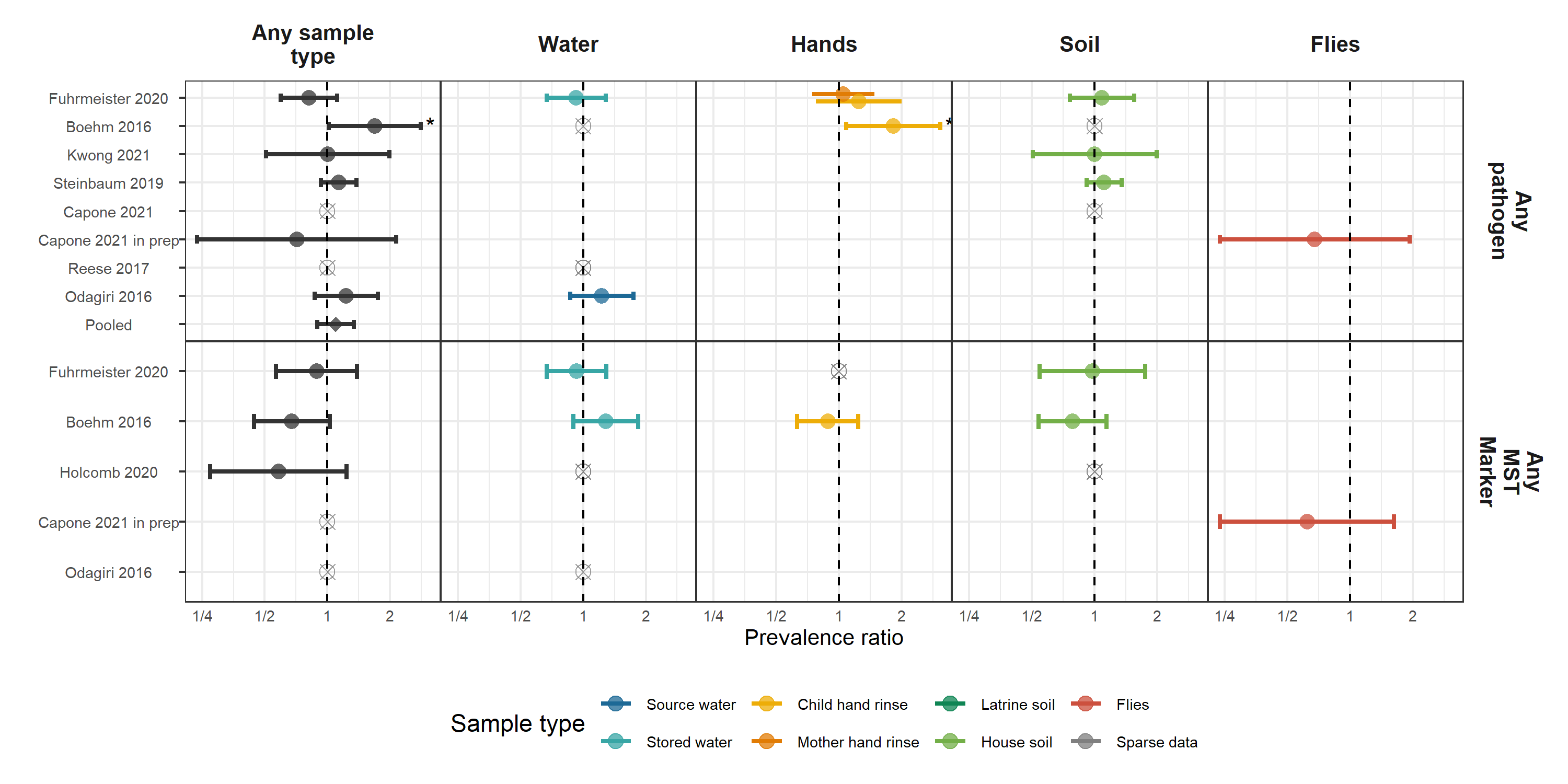
Figures and Tables

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

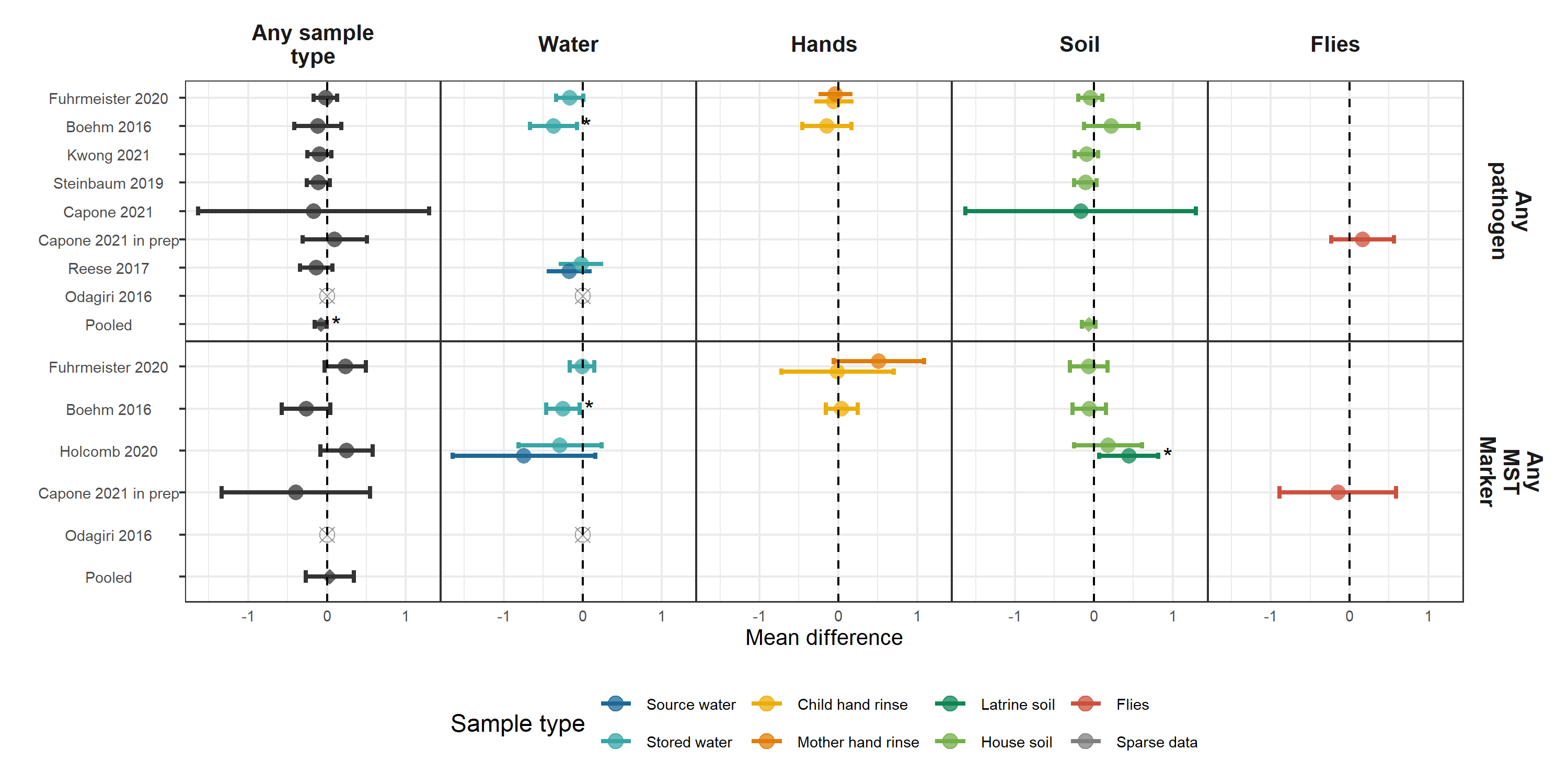
Primary figures



**Figure 1.** 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.

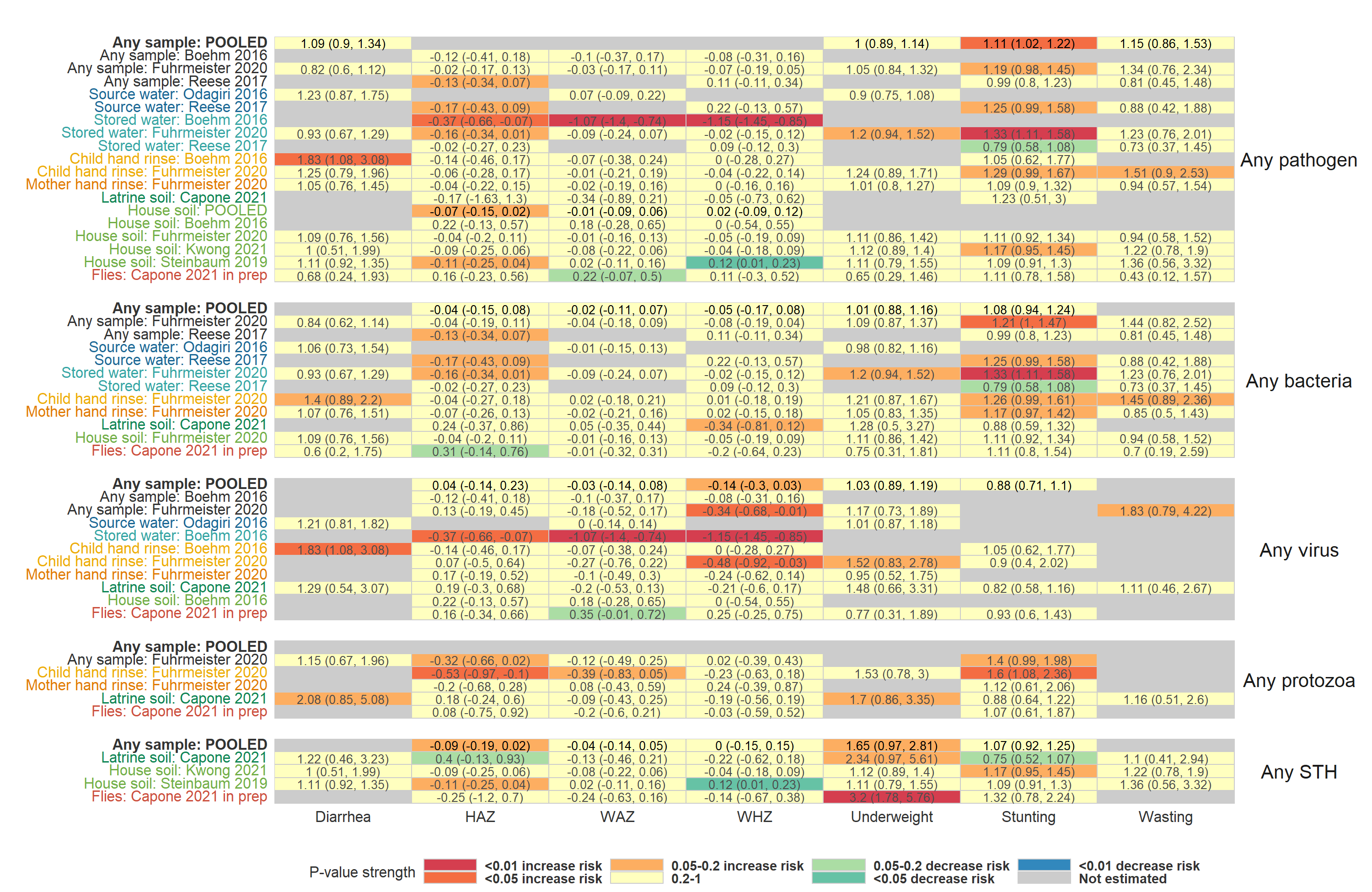


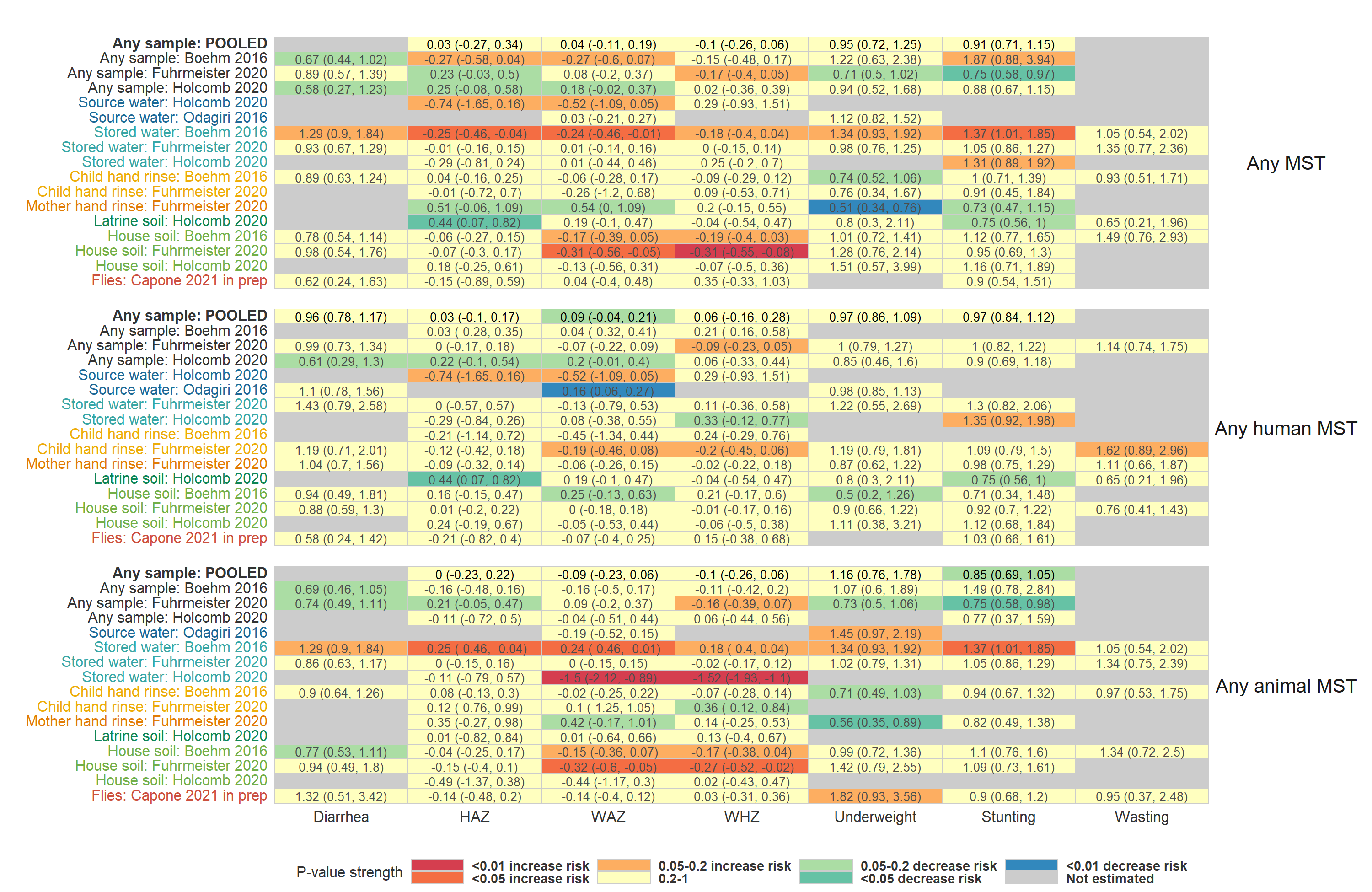
**Figure 2.** Forest plots of associations between child diarrheal disease and the prevalence of any enteropathogen or any MST markers in different types of environmental samples. Pooled estimates are presented when there are four or more study-specific estimates for a specific sample type and target 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.



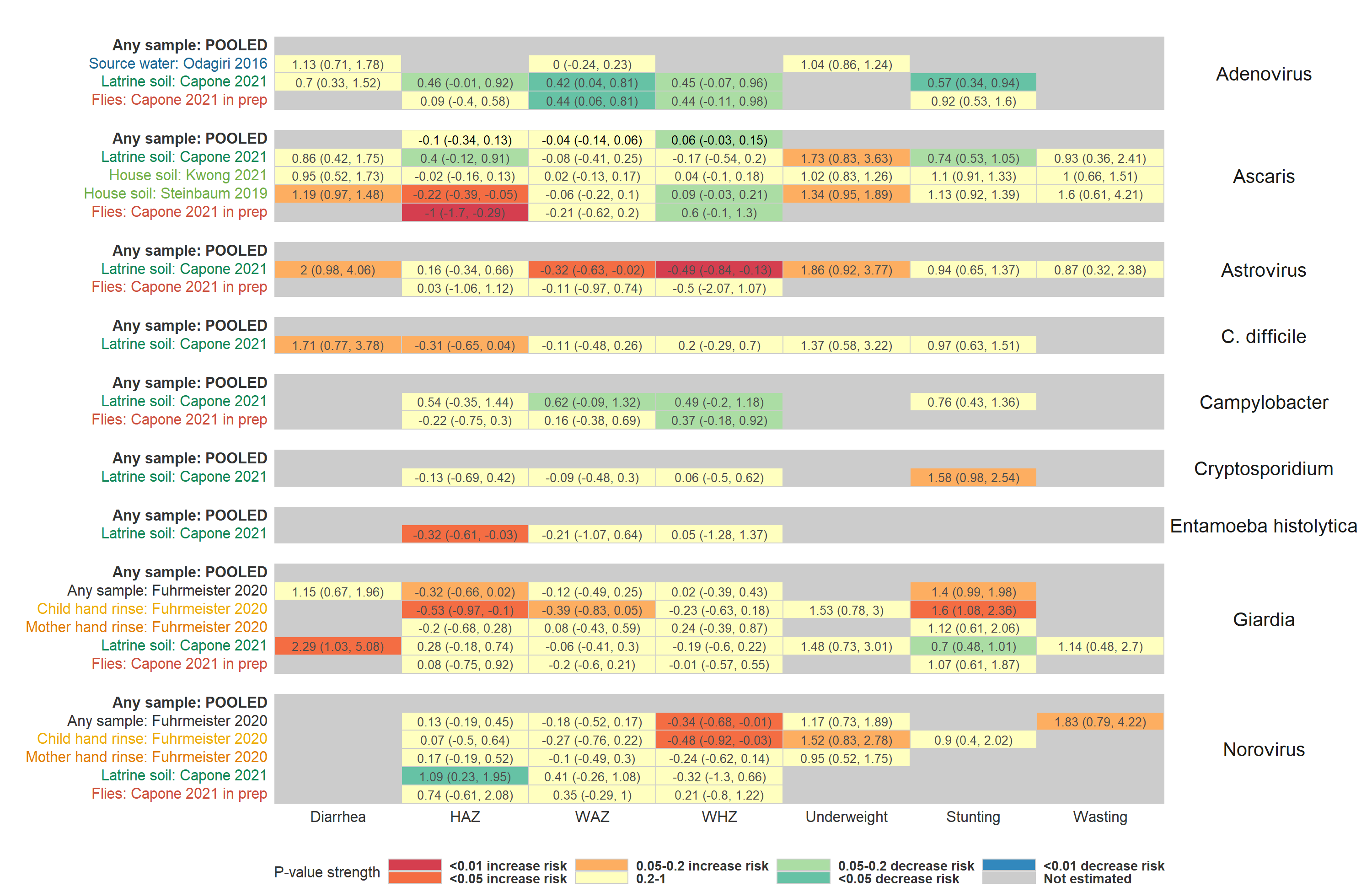
**Figure 3.** Forest plots of associations between child HAZ and the prevalence of any enteropathogen or any MST markers in different types of environmental samples. Pooled estimates are presented when there are four or more study-specific estimates for a specific sample type and target combination and 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. 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.

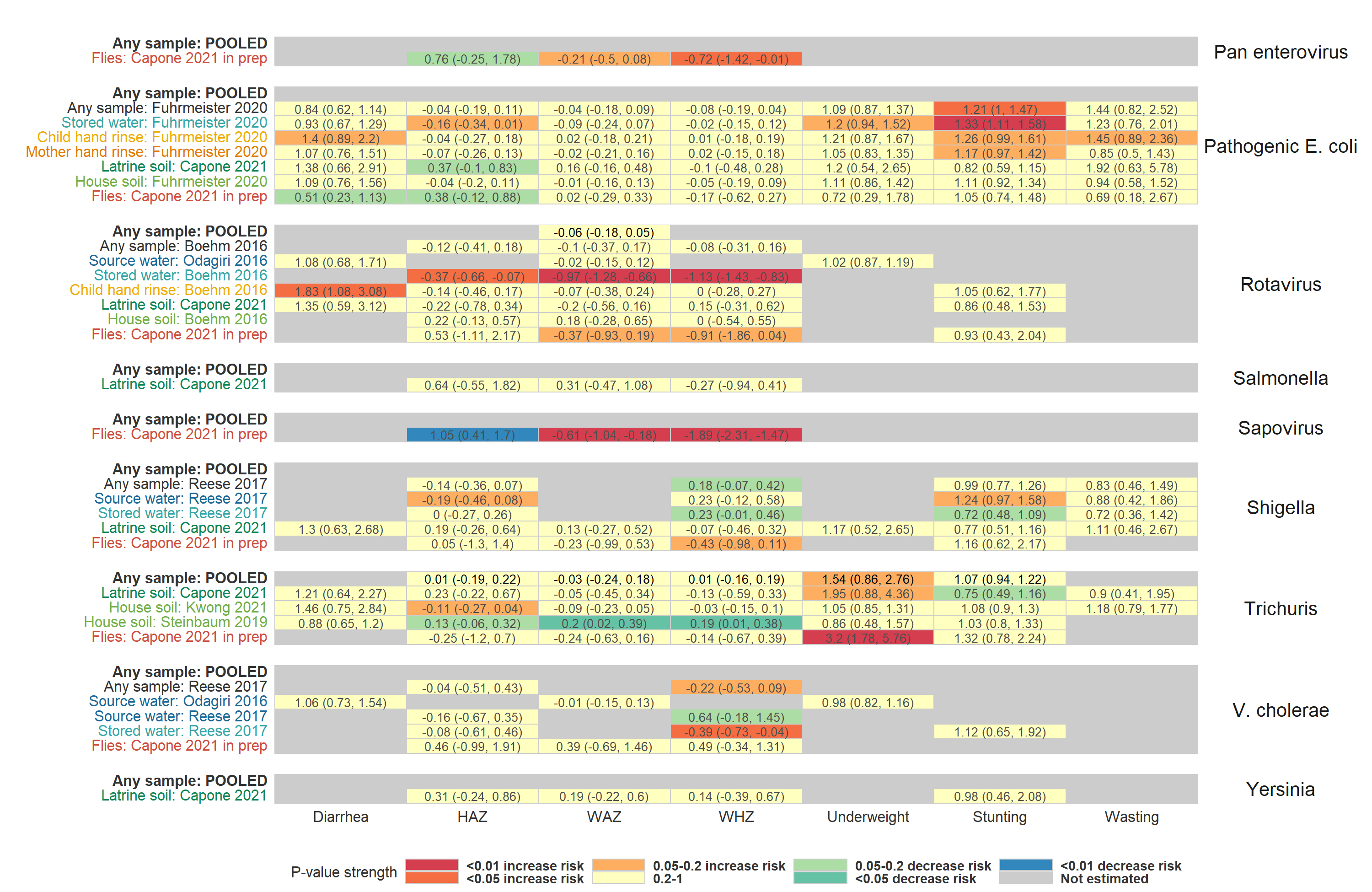
# Supplementary figures



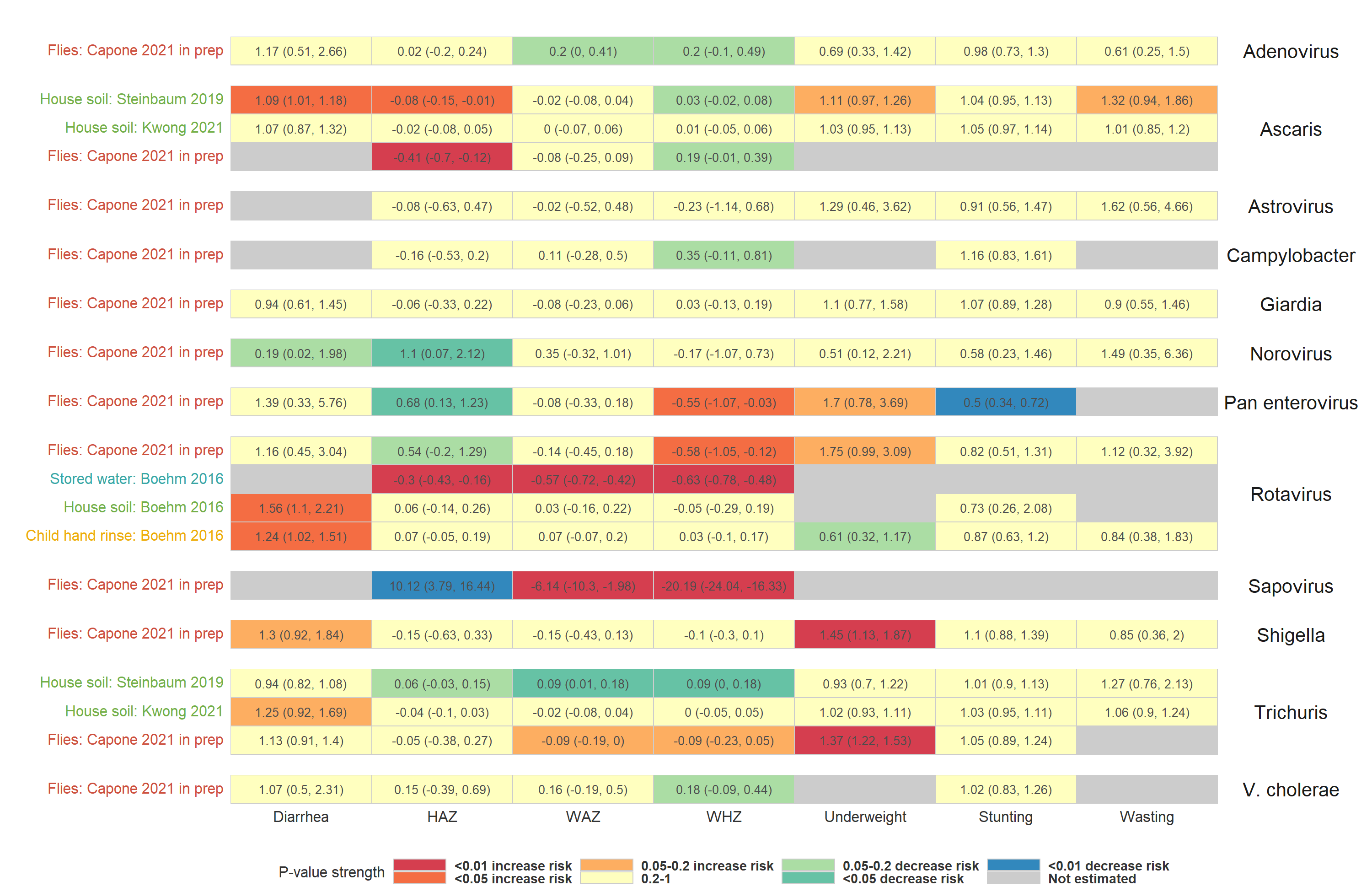


**Figure S1.** Heatmap of significance and direction of associations between aggregate measures of environmental contamination and child diarrhea and growth outcomes. Cells are colored by the strength of significance and direction of association, and the point estimate and confidence intervals are printed 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 or in a pooled estimate across studies. 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.

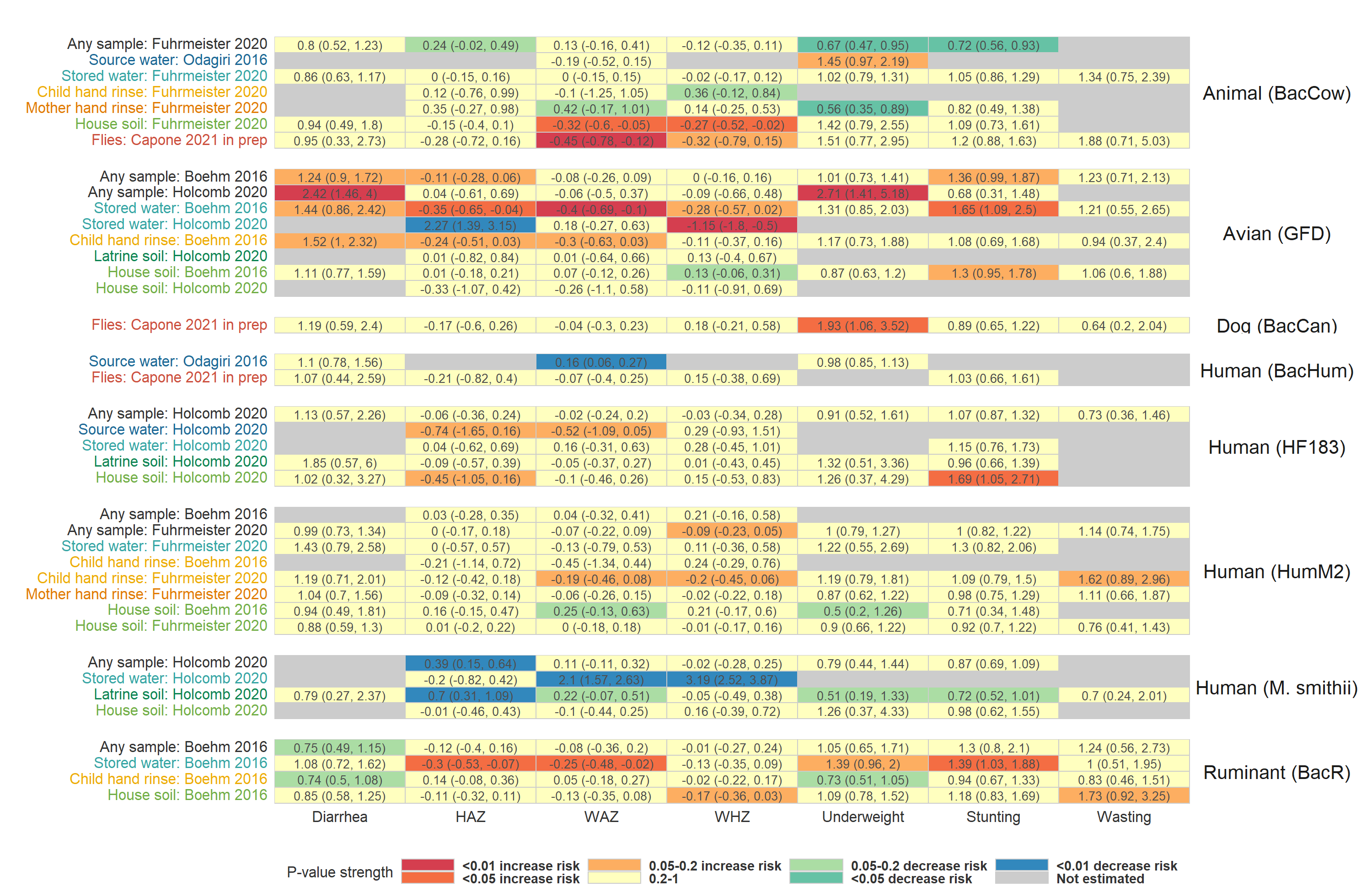




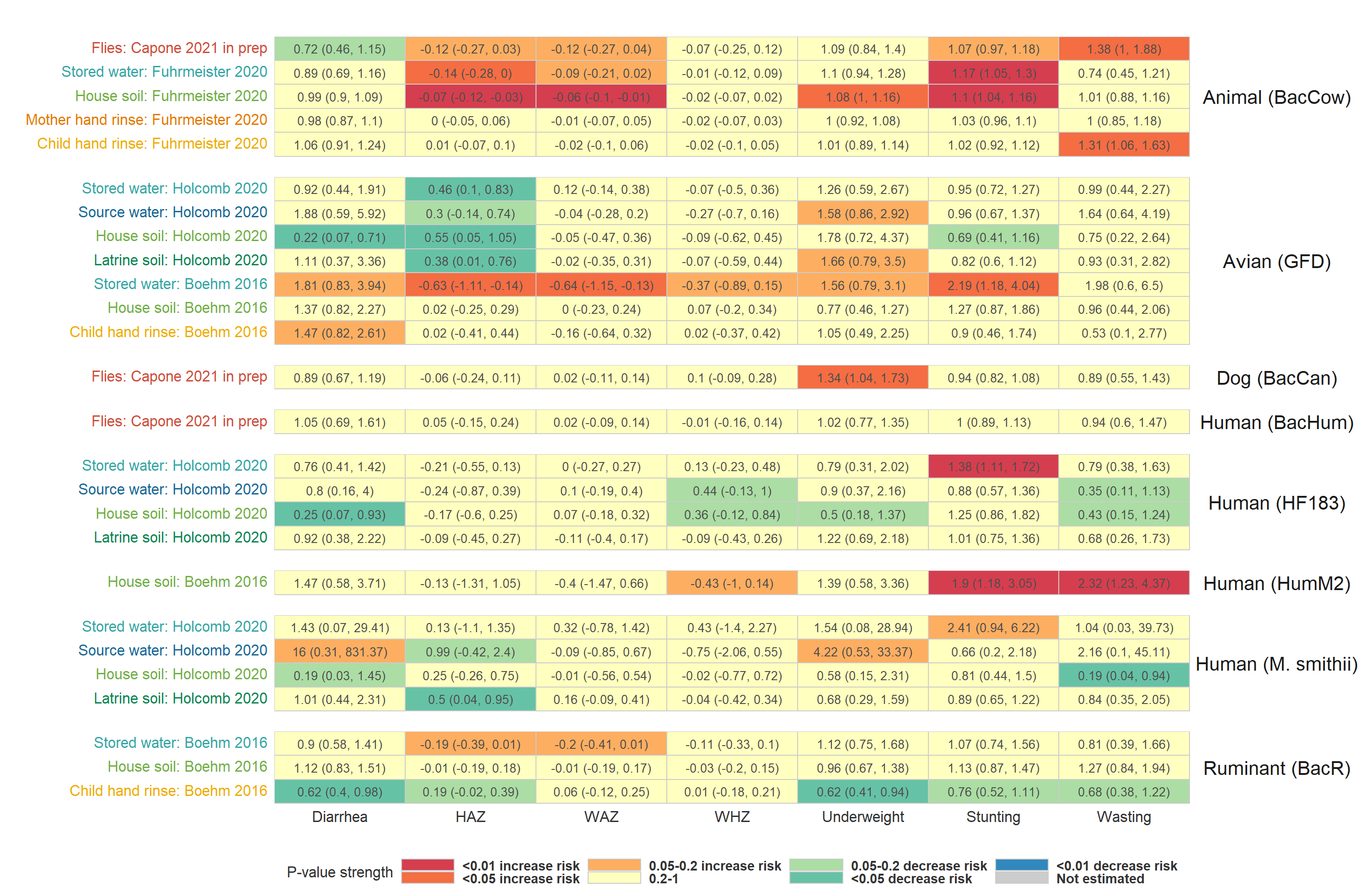
**Figure S2.** Heatmap of significance and direction of associations between specific pathogens in environmental samples and child diarrhea and growth outcomes. Cells are colored by the strength of significance and direction of association, and the point estimate and confidence intervals are printed 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 or in a pooled estimate across studies. Estimates aggregated across any sample type are only plotted if there are multiple sample types for a study. Grey cells mark missing outcomes or exposure-outcome combinations too sparse to estimate. All estimates are adjusted for potential confounders.



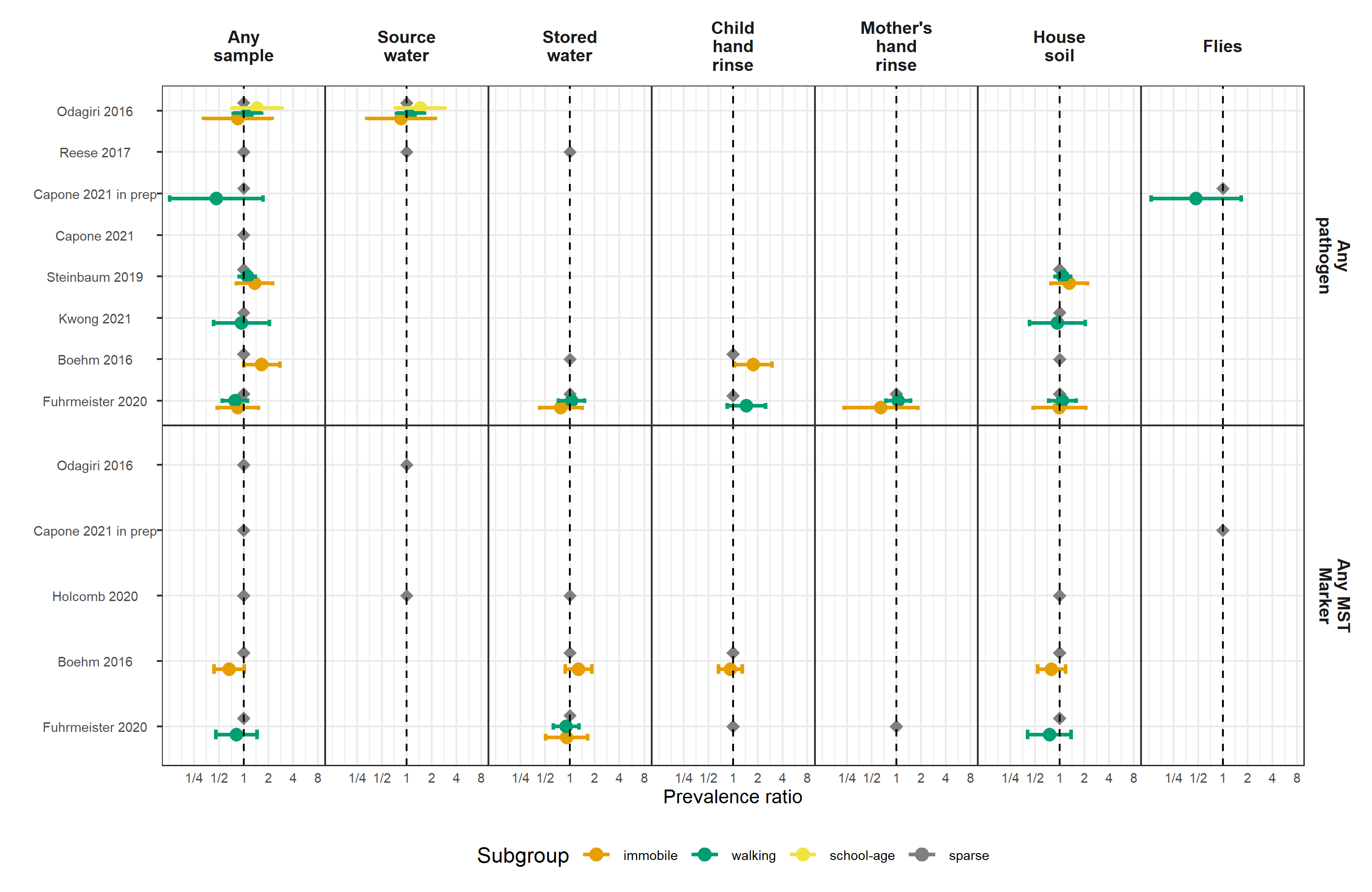
**Figure S3.** Heatmap of significance and direction of associations between the abundance of specific pathogens in environmental samples and child diarrhea and growth outcomes. Cells are colored by the strength of significance and direction of association, and the point estimate and confidence intervals 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 or in a pooled estimate across studies. Estimates aggregated across any sample type are only plotted if there are multiple sample types for a study. Grey cells mark missing outcomes or exposure-outcome combinations too sparse to estimate. All estimates are adjusted for potential confounders.



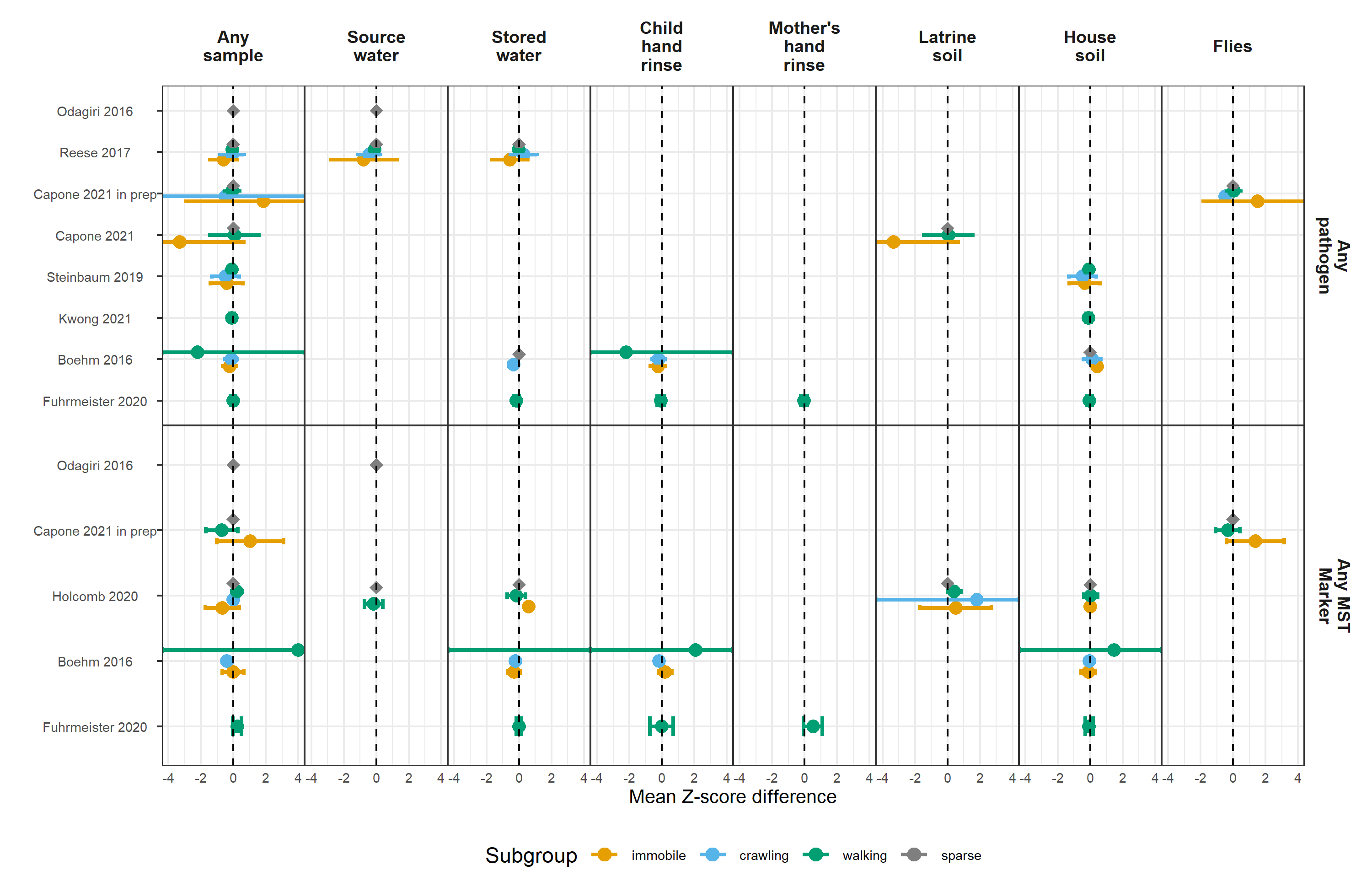
**Figure S4.** Heatmap of significance and direction of associations between specific microbial source tracking markers in environmental samples and child diarrhea and growth outcomes. Cells are colored by the strength of significance and direction of association, and the point estimate and confidence 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 or in a pooled estimate across studies. Estimates aggregated across any sample type are only plotted if there are multiple sample types for a study. Grey cells mark missing outcomes or exposure-outcome combinations too sparse to estimate. All estimates are adjusted for potential confounders.



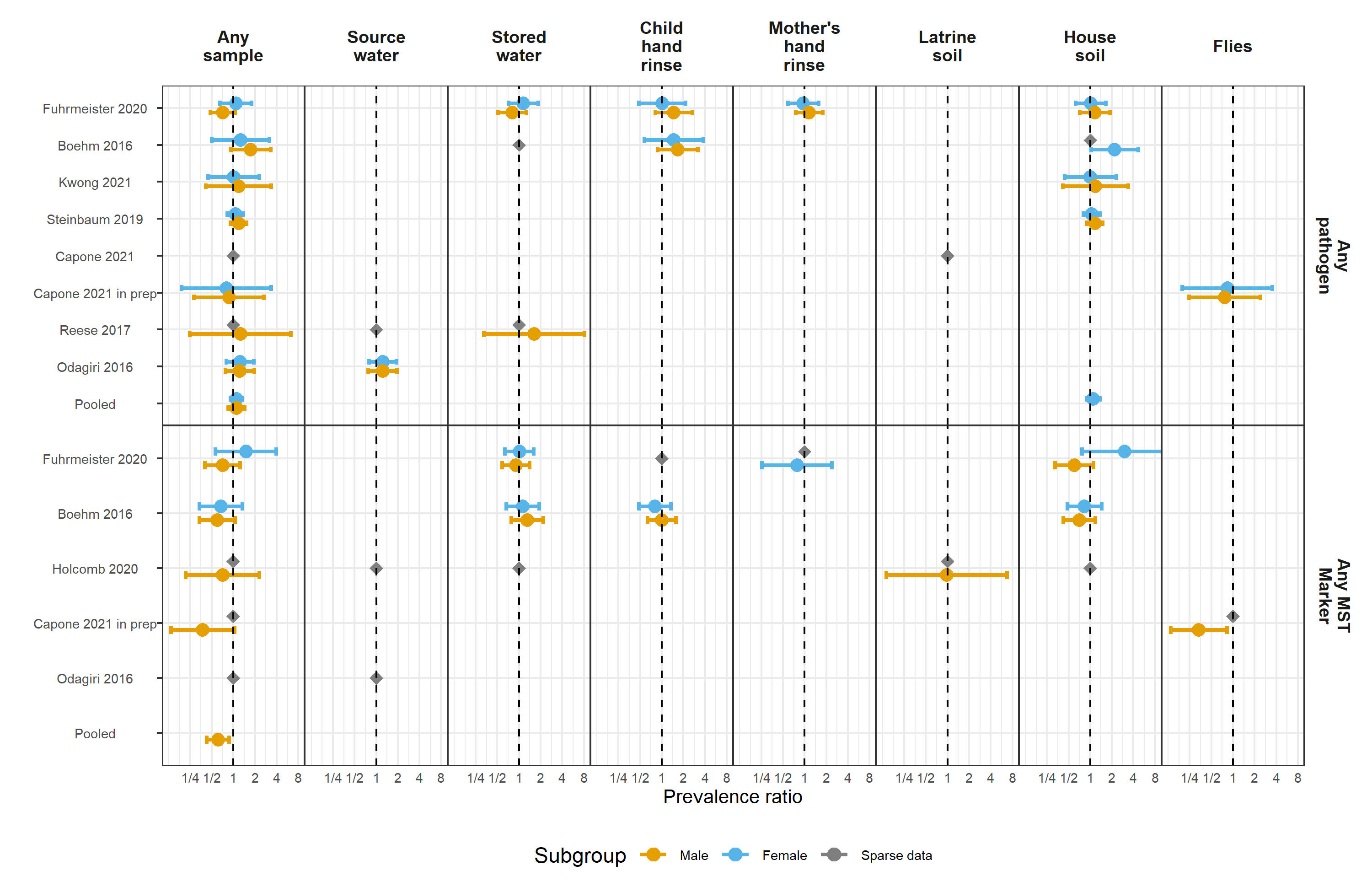
**Figure S5.** Heatmap of significance and direction of associations between the abundance of specific microbial source tracking markers in environmental samples and child diarrhea and growth outcomes. Cells are colored by the strength of significance and direction of association, and the point estimate 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 or in a pooled estimate across studies. Estimates aggregated across any sample type are only plotted if there are multiple sample types for a study. Grey cells mark missing outcomes or exposure-outcome combinations too sparse to estimate. All estimates are adjusted for potential confounders.



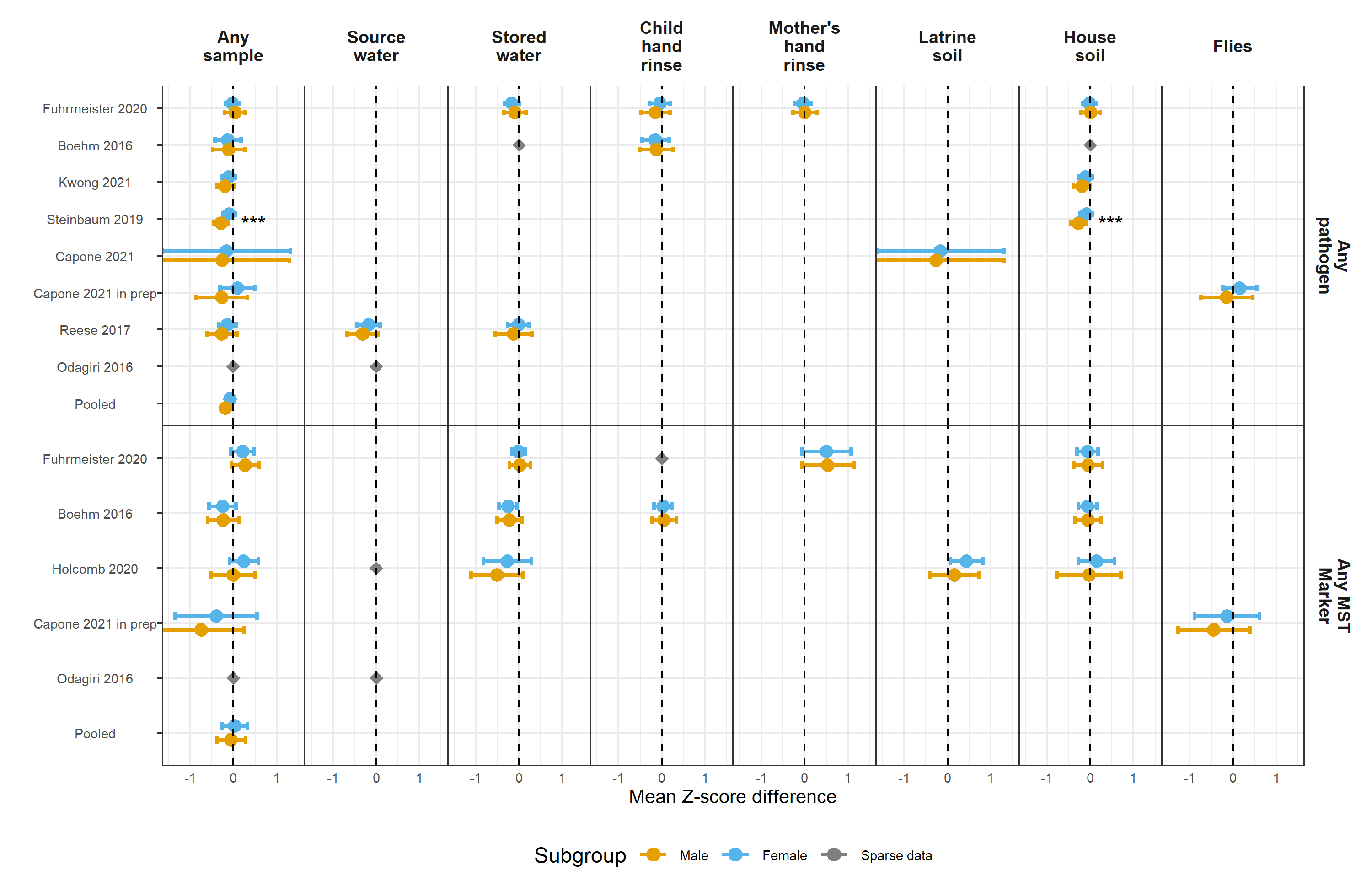
**Figure S6.** Forest plots of associations between child diarrheal disease and any enteropathogen, and any MST markers, in different types of environmental samples, stratified by child age. Grey points mark sparse age strata without estimated relative risks. Significant effect modification, as determined by the p-values on the regression model interaction term, is marked with asterisks (P < 0.05 = \*, P < 0.01 = \*\*, P < 0.001 = \*\*\*).



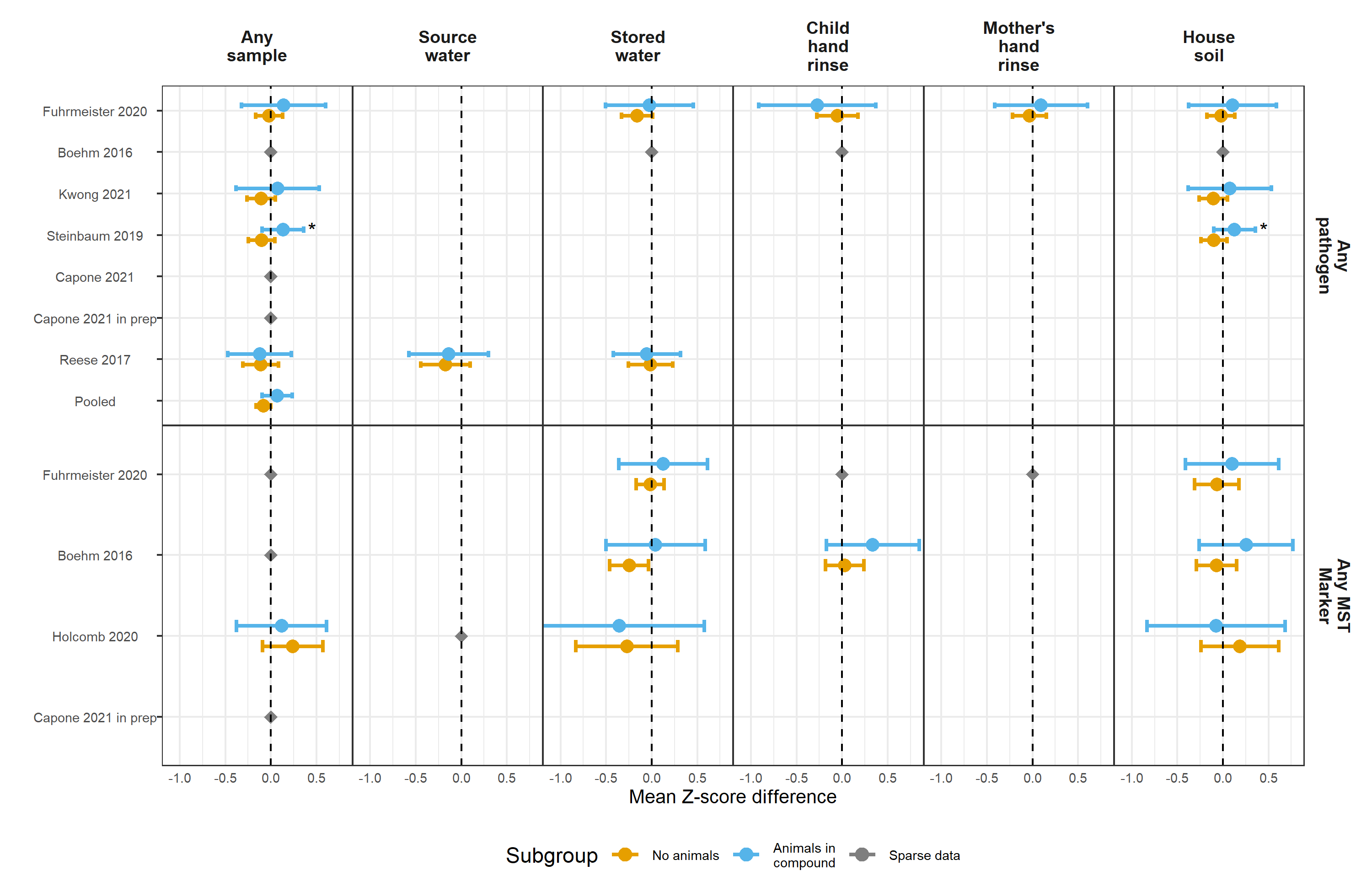
**Figure S7.** Forest plots of associations between child HAZ and any enteropathogen, and any MST markers, in different types of environmental samples, stratified by child age. Grey points mark sparse age strata without estimated mean differences. Significant effect modification, as determined by the p-values on the regression model interaction term, is marked with asterisks (P < 0.05 = \*, P < 0.01 = \*\*, P < 0.001 = \*\*\*).



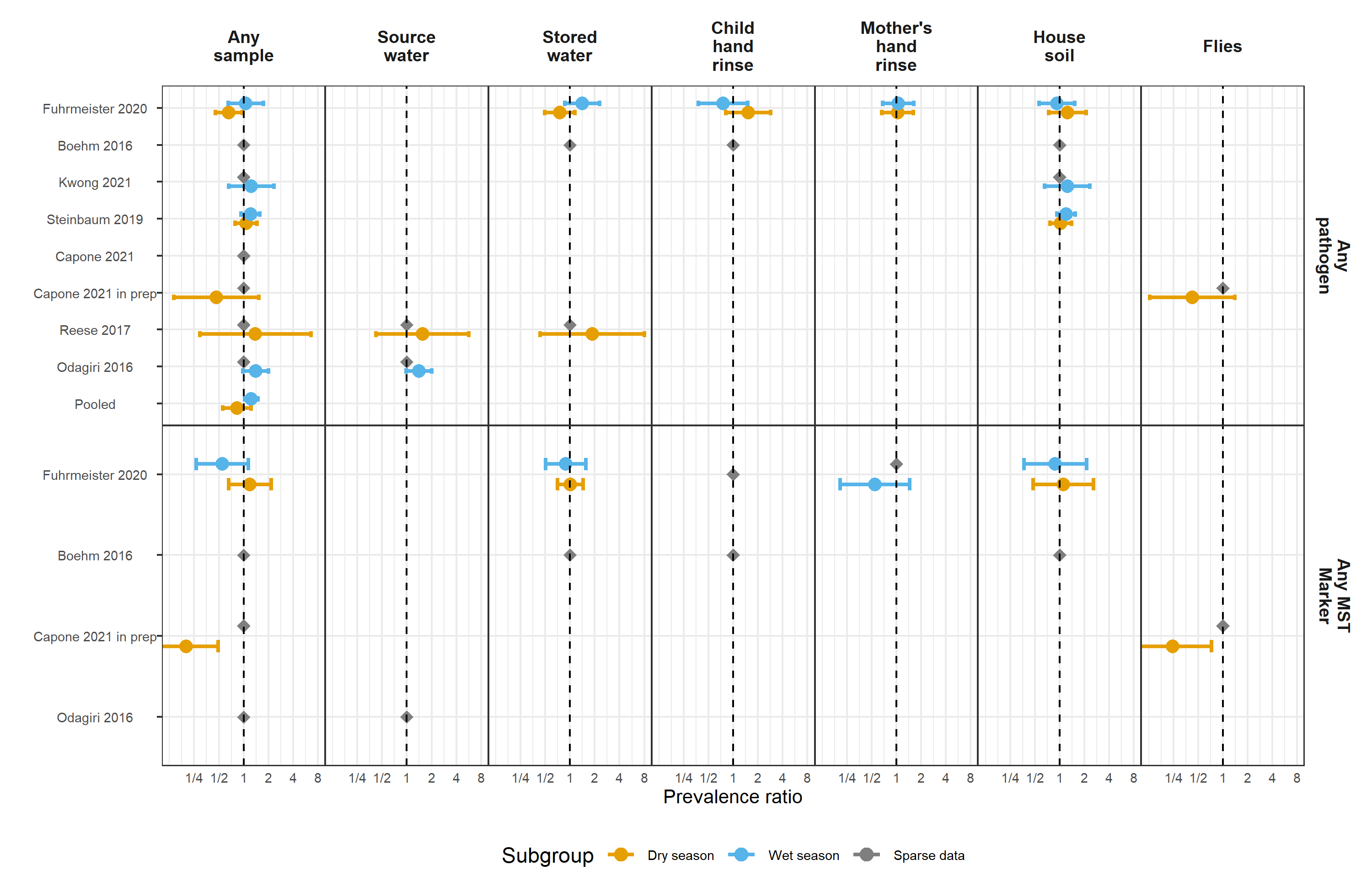
**Figure S8.** Forest plots of associations between child diarrheal disease and any enteropathogen, and any MST markers, in different types of environmental samples, stratified by child sex. Significant effect modification, as determined by the p-values on the regression model interaction term, is marked with asterisks (P < 0.05 = \*, P < 0.01 = \*\*, P < 0.001 = \*\*\*).



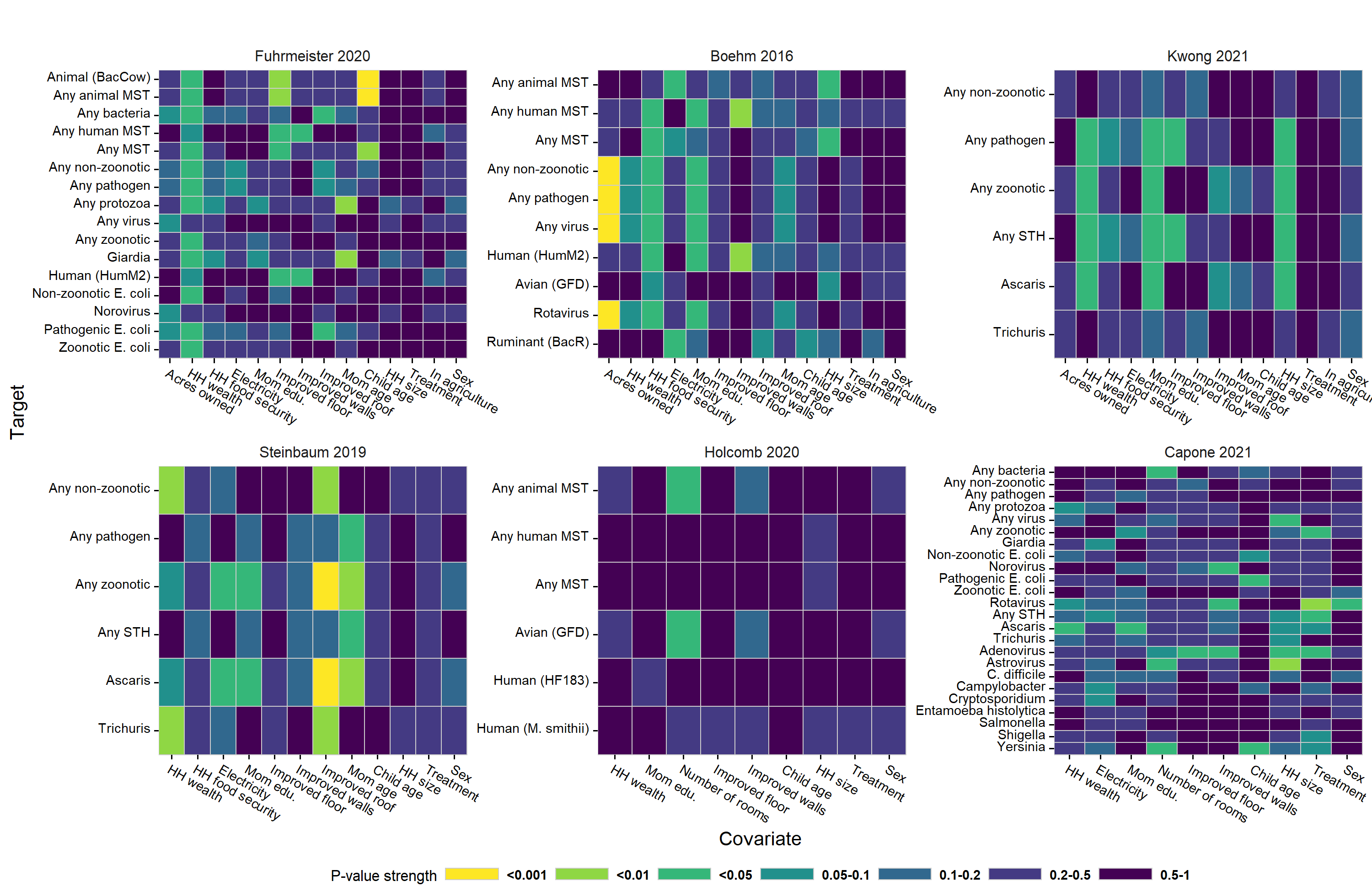
**Figure S9.** Forest plots of associations between child HAZ and any enteropathogen, and any MST markers, in different types of environmental samples, stratified by child sex. Significant effect modification, as determined by the p-values on the regression model interaction term, is marked with asterisks (P < 0.05 = \*, P < 0.01 = \*\*, P < 0.001 = \*\*\*).

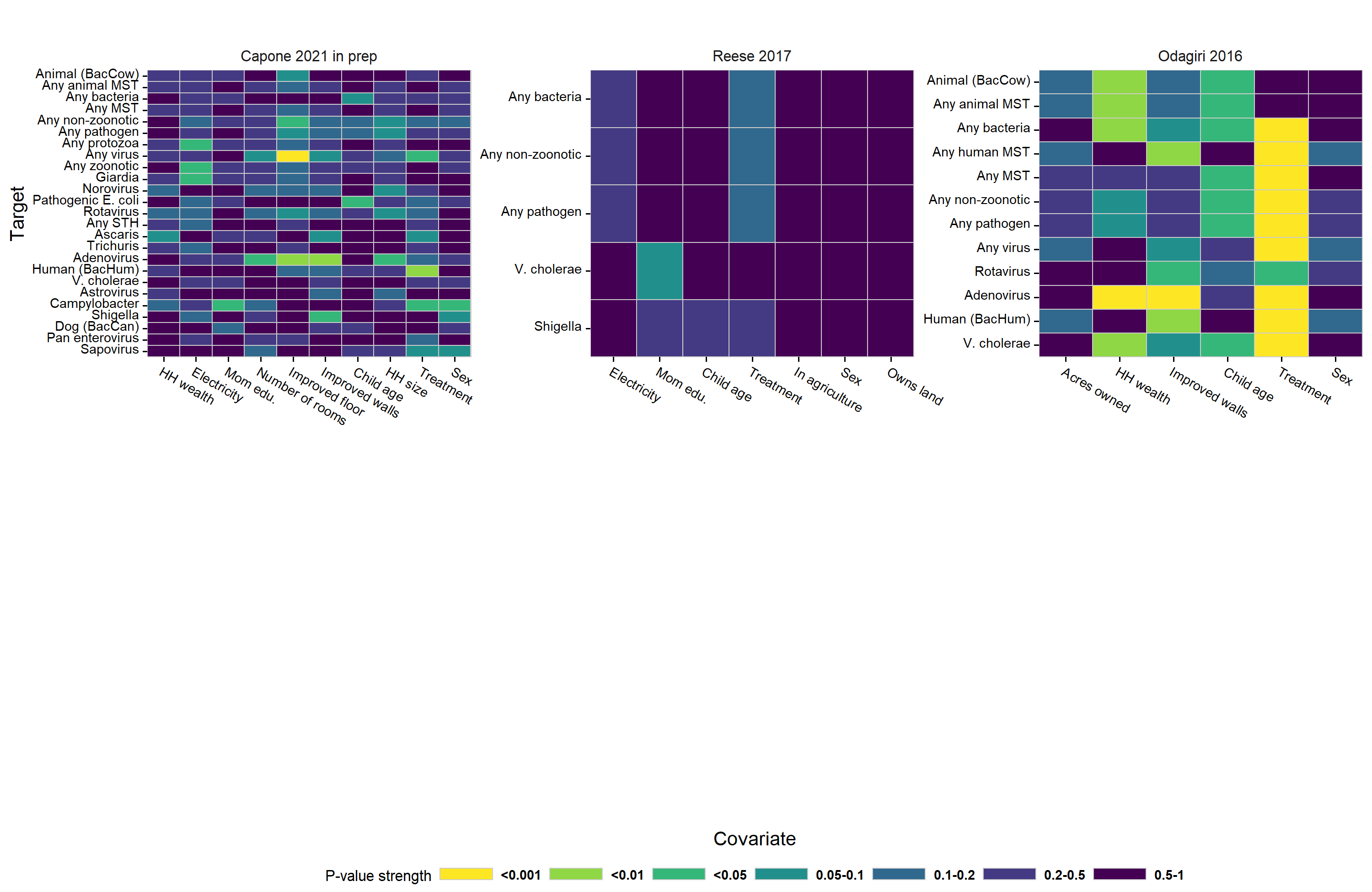


**Figure S10.** Forest plots of associations between child HAZ and any enteropathogen, and any MST markers, in different types of environmental samples, stratified by whether any animals were present in the compound. Significant effect modification, as determined by the p-values on the regression model interaction term, is marked with asterisks (P < 0.05 = \*, P < 0.01 = \*\*, P < 0.001 = \*\*\*).

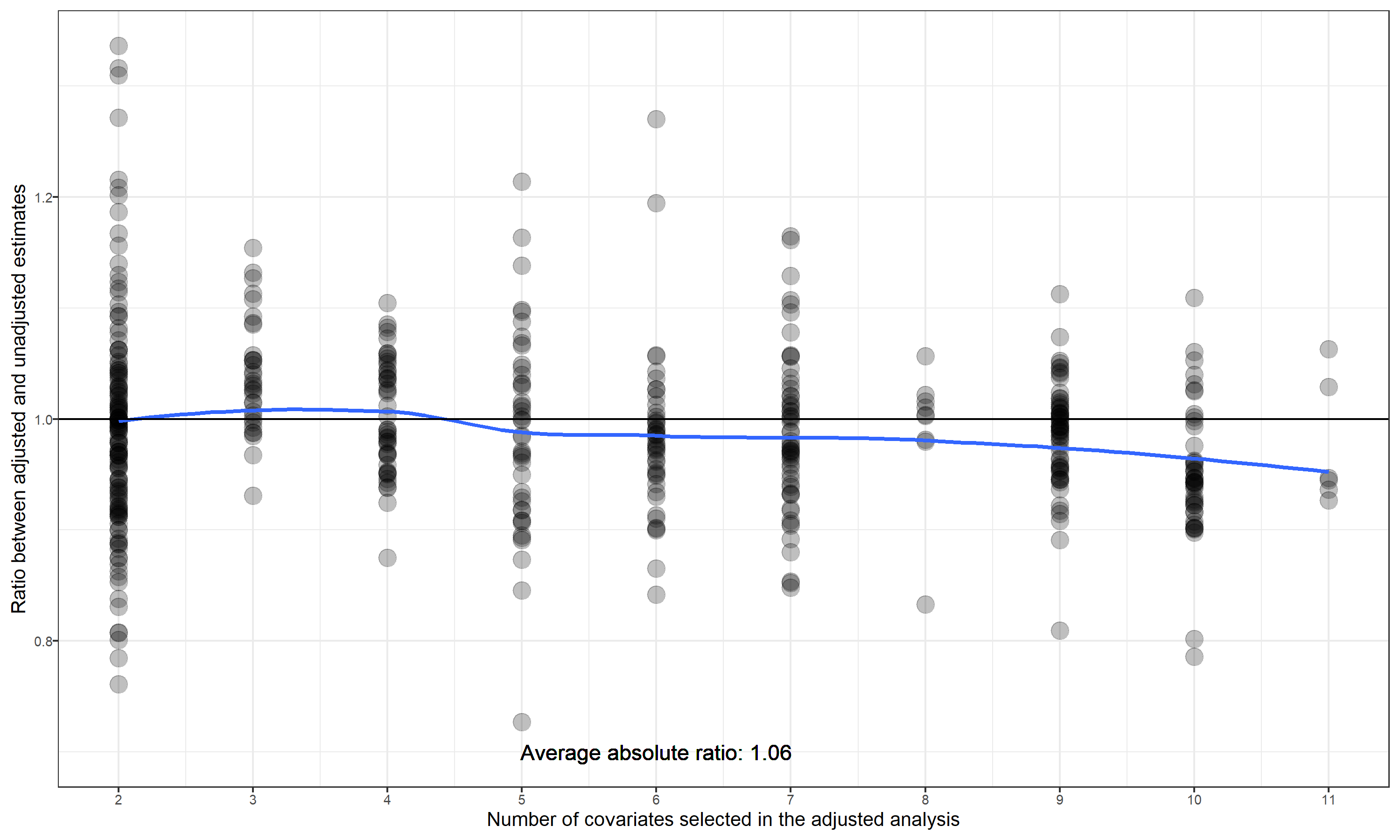


**Figure S11.** Forest plots of relative risks between diarrhea disease and any enteropathogen, and any MST markers, in different types of environmental samples, stratified by whether the diarrheal disease occurred during the wet versus dry season (defined by the 6 months of highest average rainfall). 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 = \*\*\*).

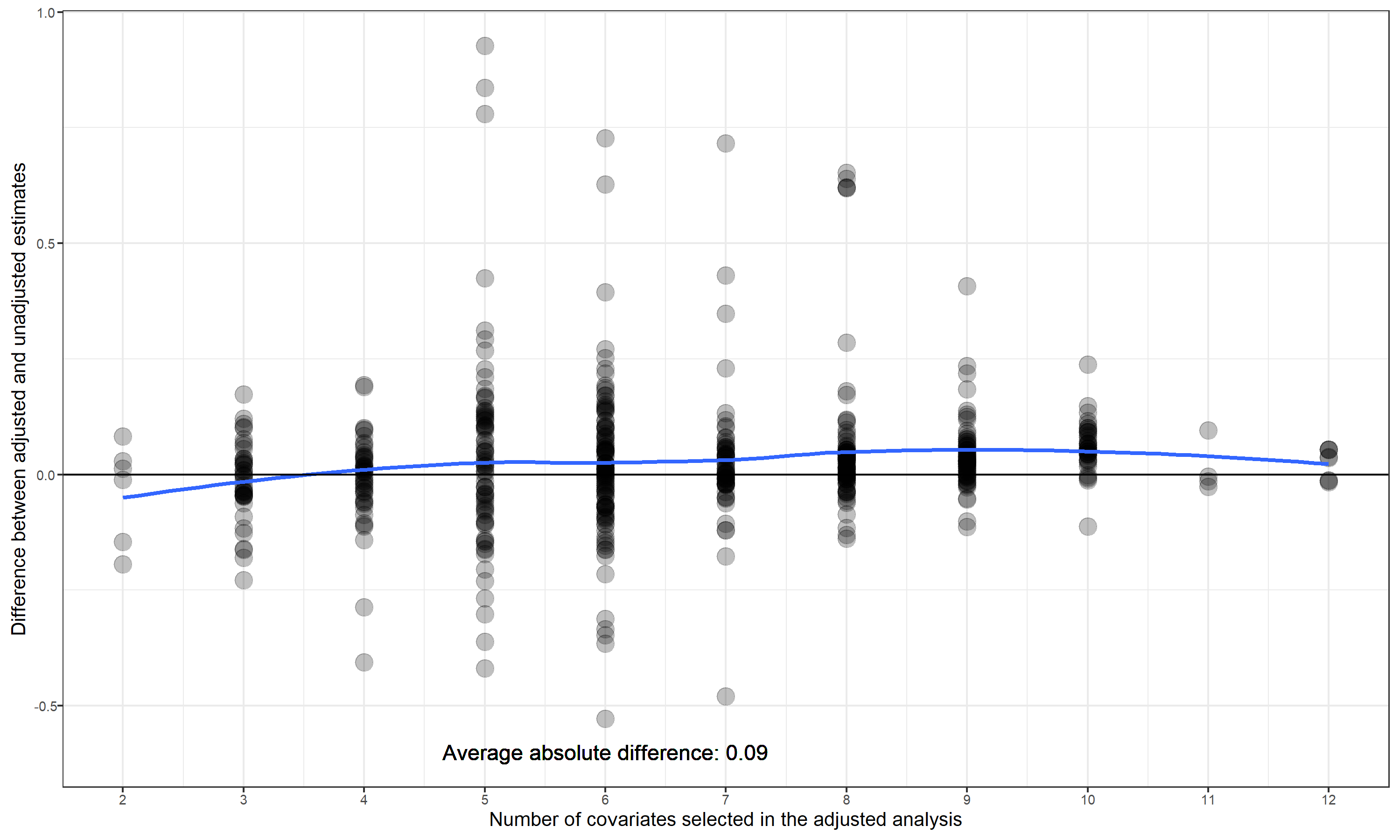




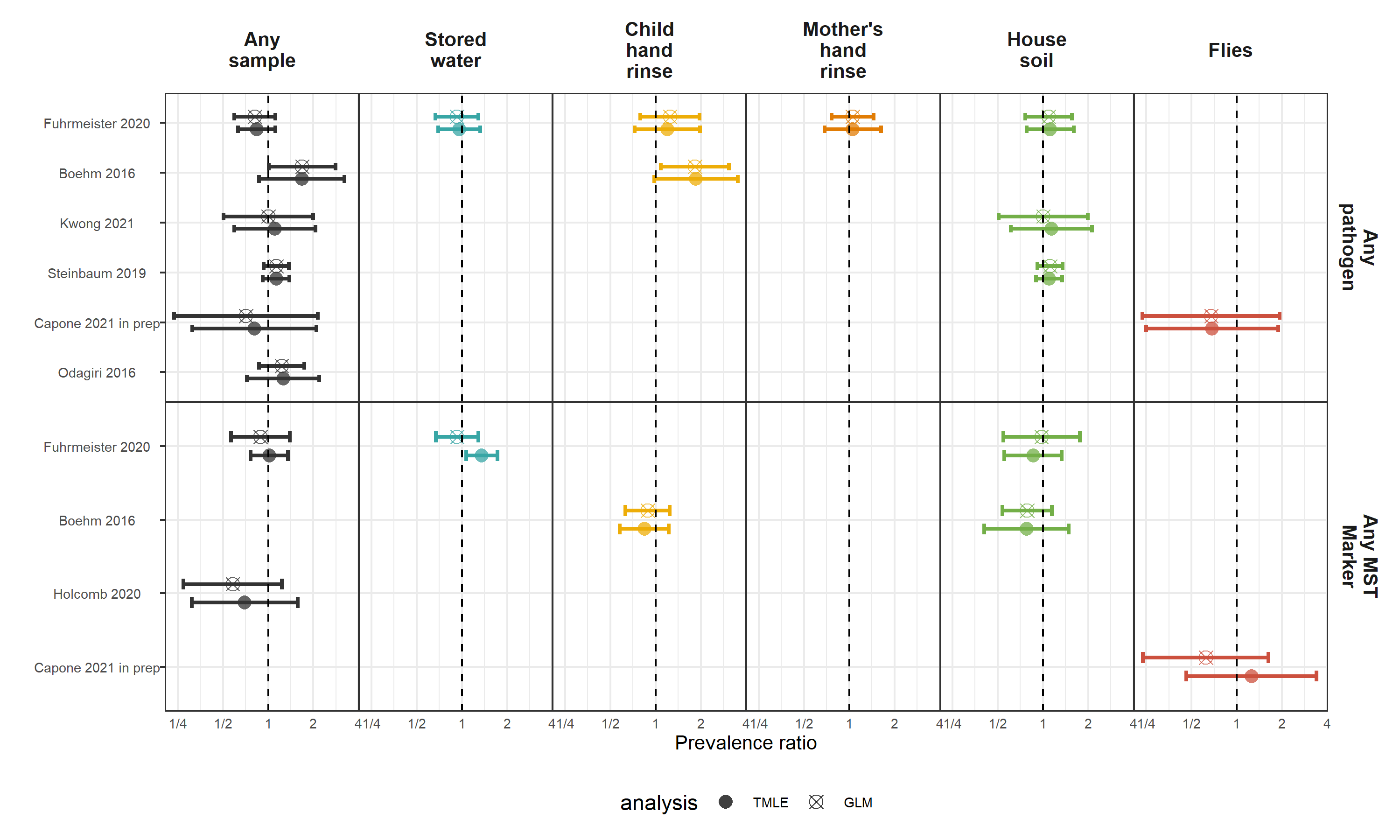
**Figure S12.** Study-specific associations between adjustment covariates and the presence of different enteropathogen and MST markers in aggregated environmental samples. The columns are different pre-screened confounders, and the rows are specific enteropathogens and MST markers. Cells of the heatmaps are colored by P-values of bivariate likelihood ratio tests, and heatmaps are stratified by study.



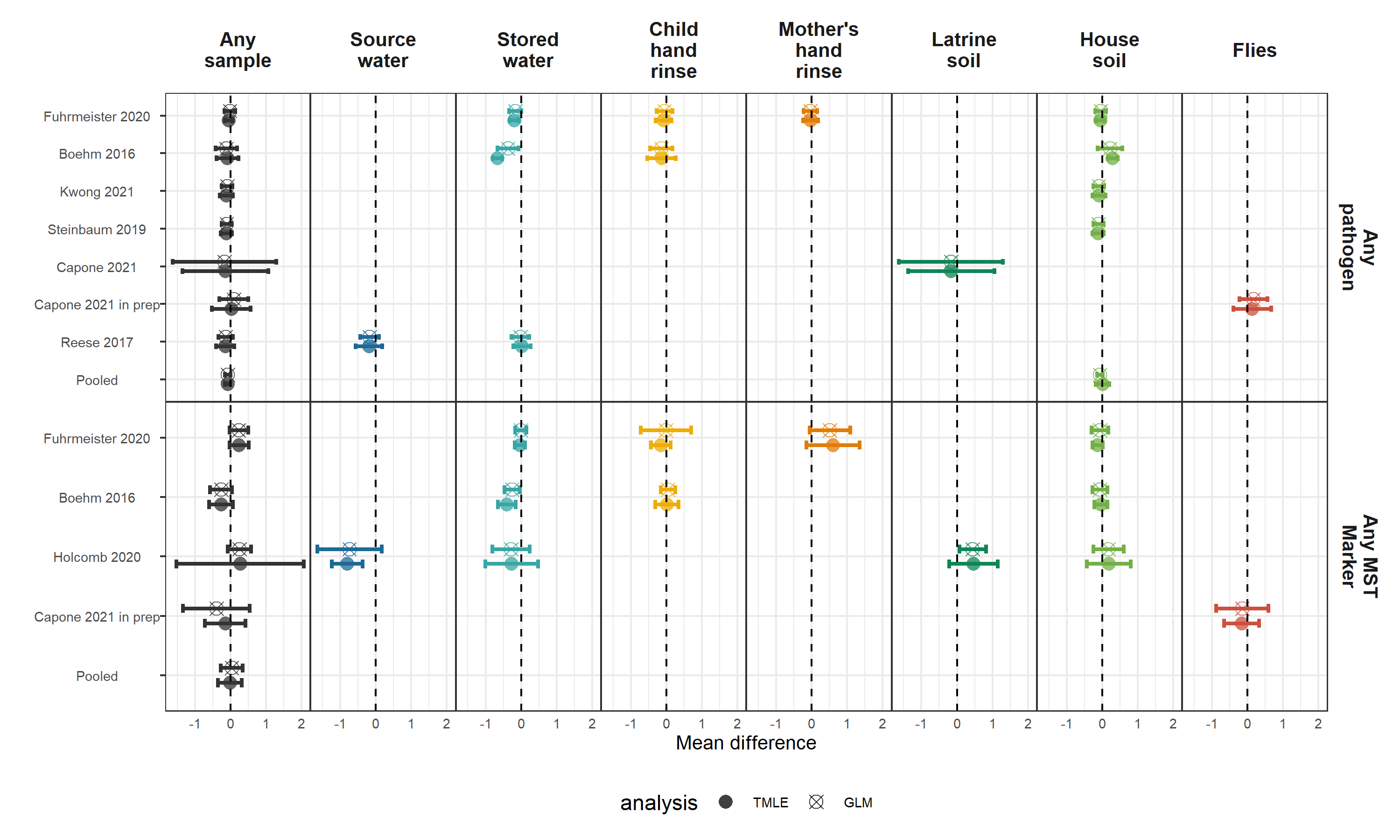
**Figure 13.** Comparison between associations estimated with and without including potential confounders for the binary diarrhea and growth outcomes. Points mark the ratio of relative risks estimated using adjusted and unadjusted generalized linear models. The blue line shows the average ratio between adjusted estimates and unadjusted estimates, fitted using a cubic spline.



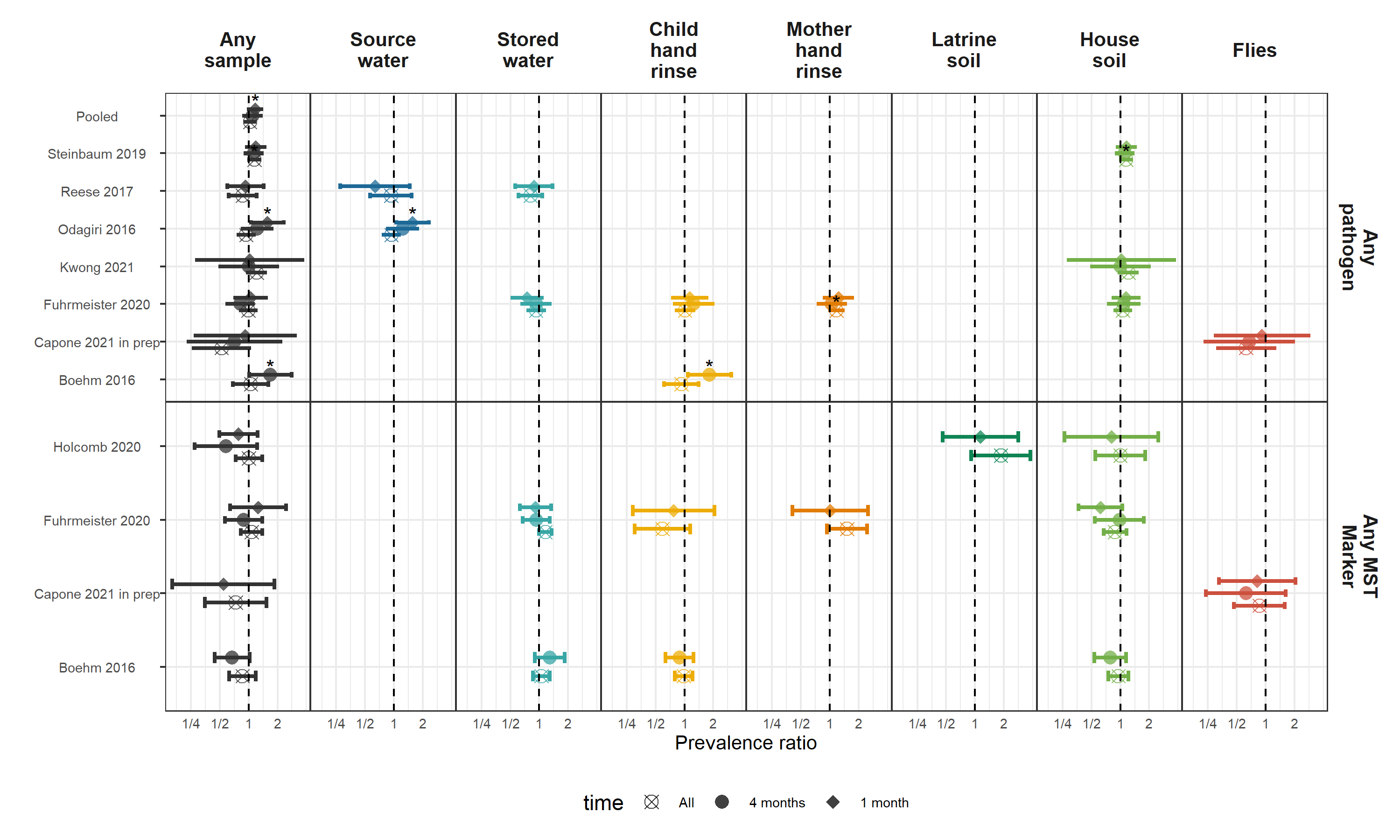
**Figure 14.** Comparison between associations estimated with and without including potential confounders for the continious growth outcomes. Points mark the differences between mean differences estimated using adjusted and unadjusted generalized linear models. The blue line shows the average difference in differences between adjusted estimates and unadjusted estimates, fitted using a cubic spline.



**Figure S15.** Comparison between associations estimated with generalized linear models (GLM) and machine-learning based targeted likelihood estimation models (TMLE) for the diarrhea outcome.



**Figure S16.** Comparison between associations estimated with generalized linear models (GLM) and machine-learning based targeted likelihood estimation models (TMLE) for the height-for-age Z-score outcome.



**Figure S17.** Comparison between associations estimated in the primary diarrhea analysis (diarrheal disease occurring after environmental sampling, but no more than 4 months later with associations estimated only using diarrheal disease cases within 1 month, or occuring at any time). For the analysis of all diarrhea, it included diarrheal cases, even cases occuring prior to sampling, under the hypothesis that enteropathogen presence at one time is a surrogate variable for general environmental contamination.

# Tables

**Table 1.** Descriptive statistics of child health outcomes by study.

| **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 | 412 | 99 | 24.0 | 411 | -1.35 | 26.3 | 412 | -1.35 | 24.3 | 412 | -0.74 | 9.5 |
| 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 |