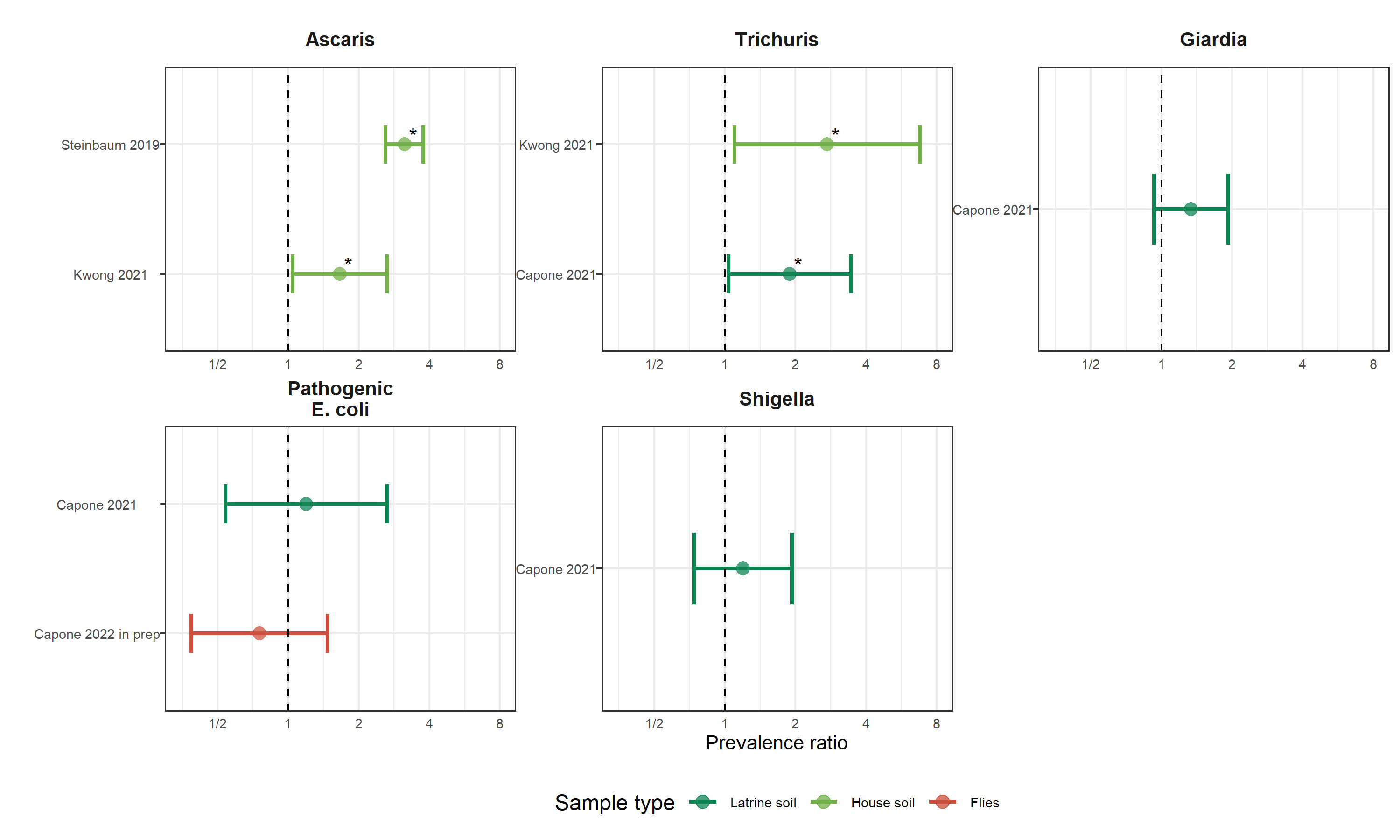
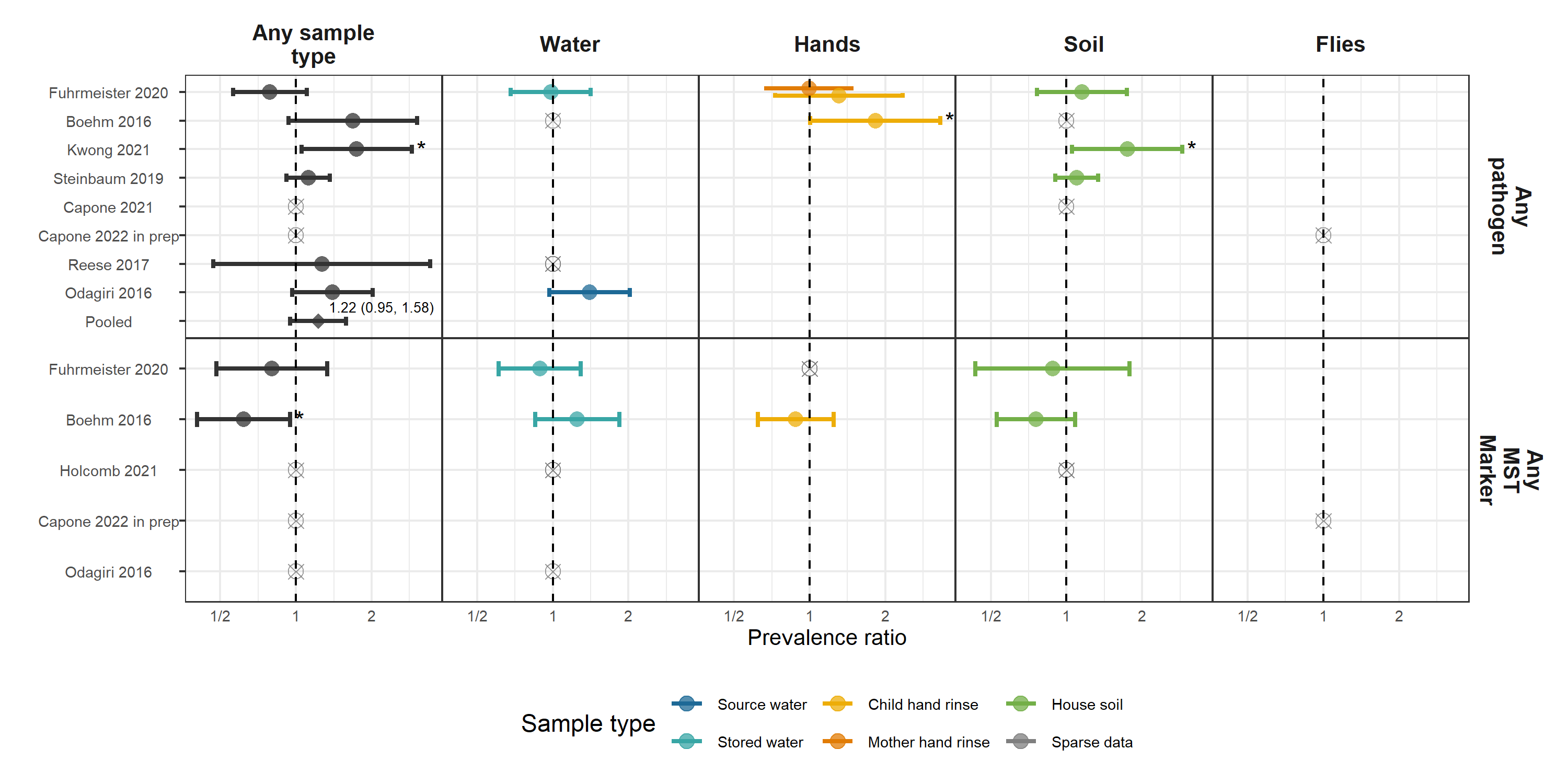
Figures and Tables

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

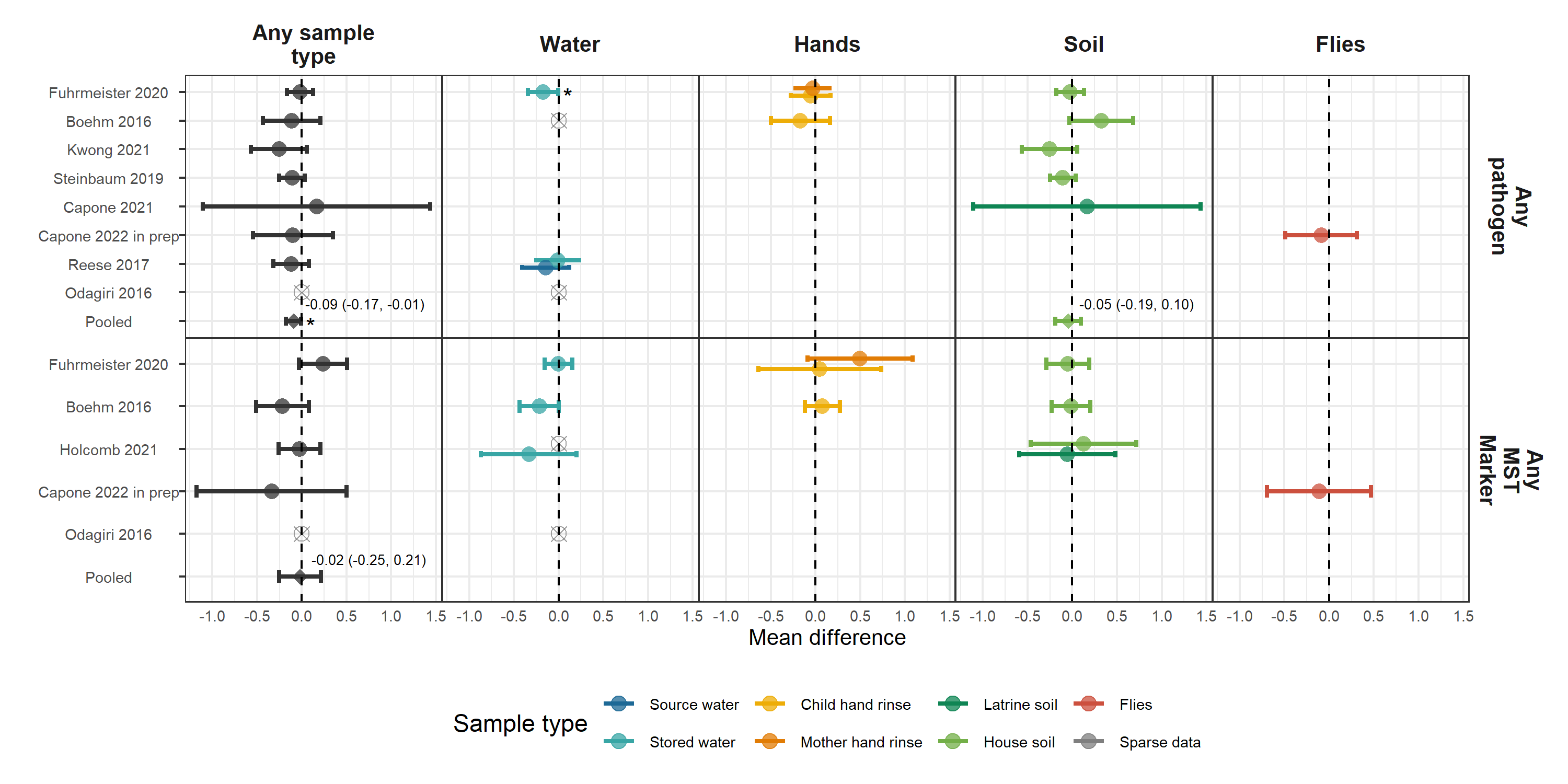
# Primary figures



**Figure 1.** Forest plots of associations between specific enteropathogens in environmental samples and child infections with the same enteropathogens. The presented prevalence ratios compare the detection prevalence of a pathogen in stool between children from compounds where the pathogen was detected vs. not detected in environmental samples. 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. Starred points indicate P-values < 0.05.

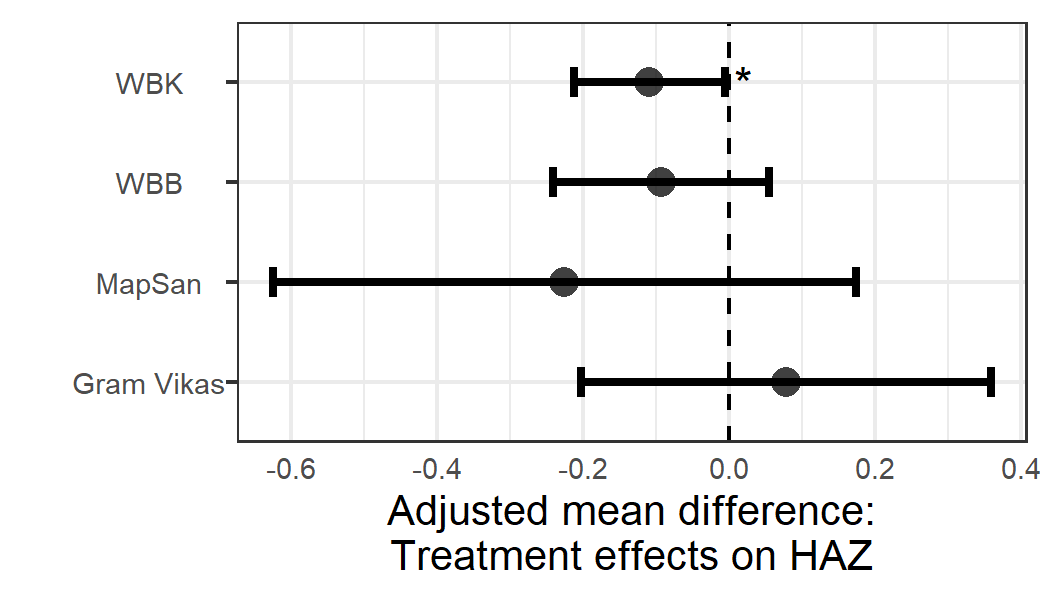


**Figure 2.** Forest plots of associations between the prevalence of any enteropathogen or any MST markers in different types of environmental samples and child diarrhoeal disease. The presented prevalence ratios compare diarrhoea prevalence between children from compounds where any pathogen/MST marker was detected vs. not detected in 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.

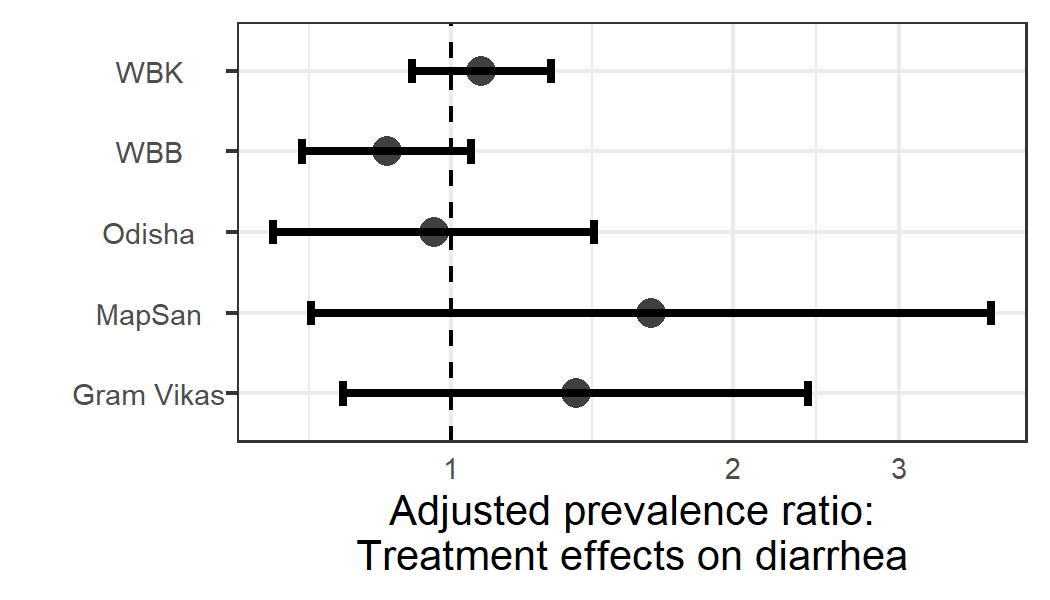


**Figure 3.** Forest plots of associations between the prevalence of any enteropathogen or any MST markers in different types of environmental samples and heigh-for-age Z-scores (HAZ). The presented differences compare HAZ between children from compounds where any pathogen/MST marker was detected vs. not detected in 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.

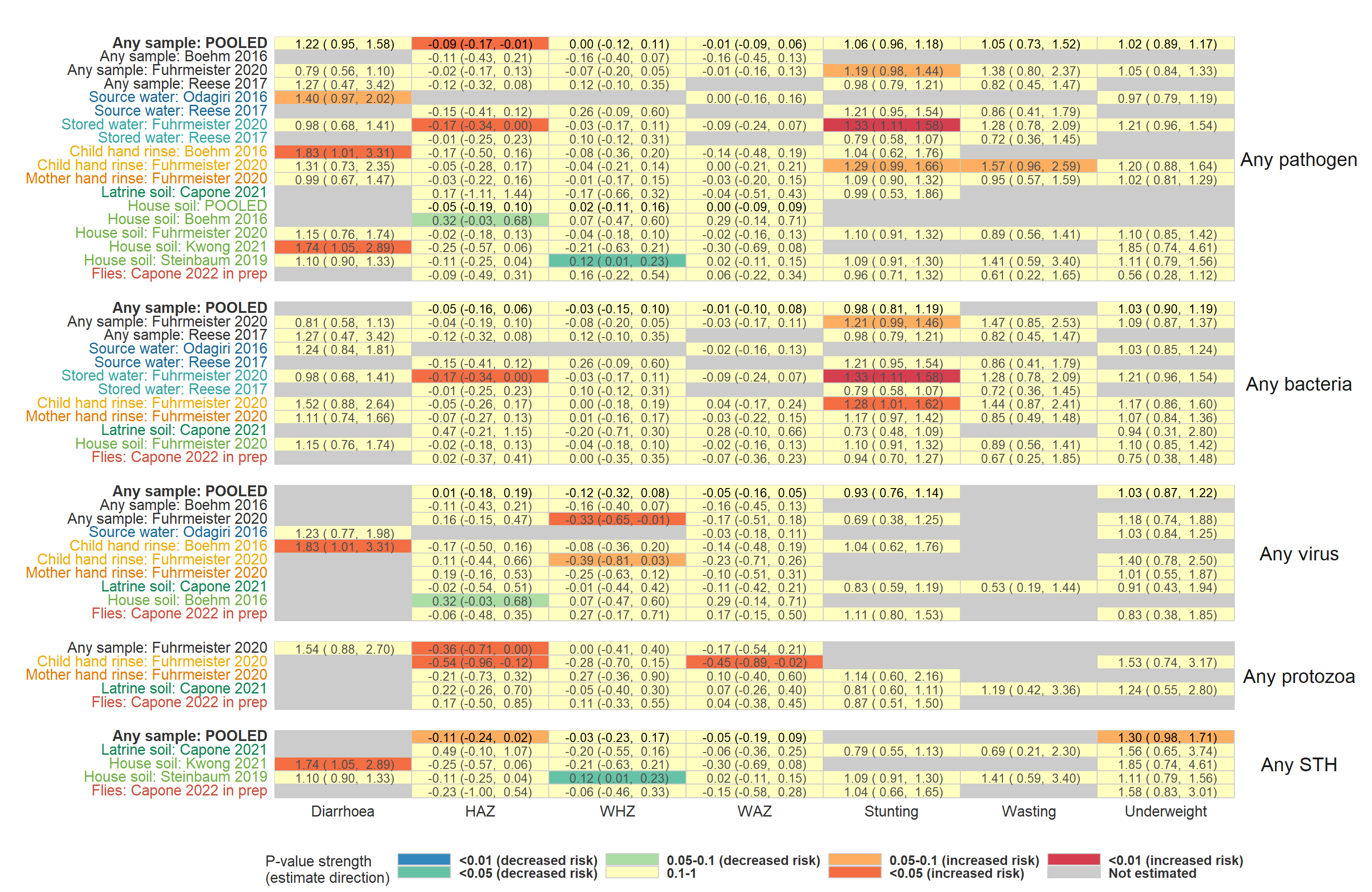
# Supplementary figures

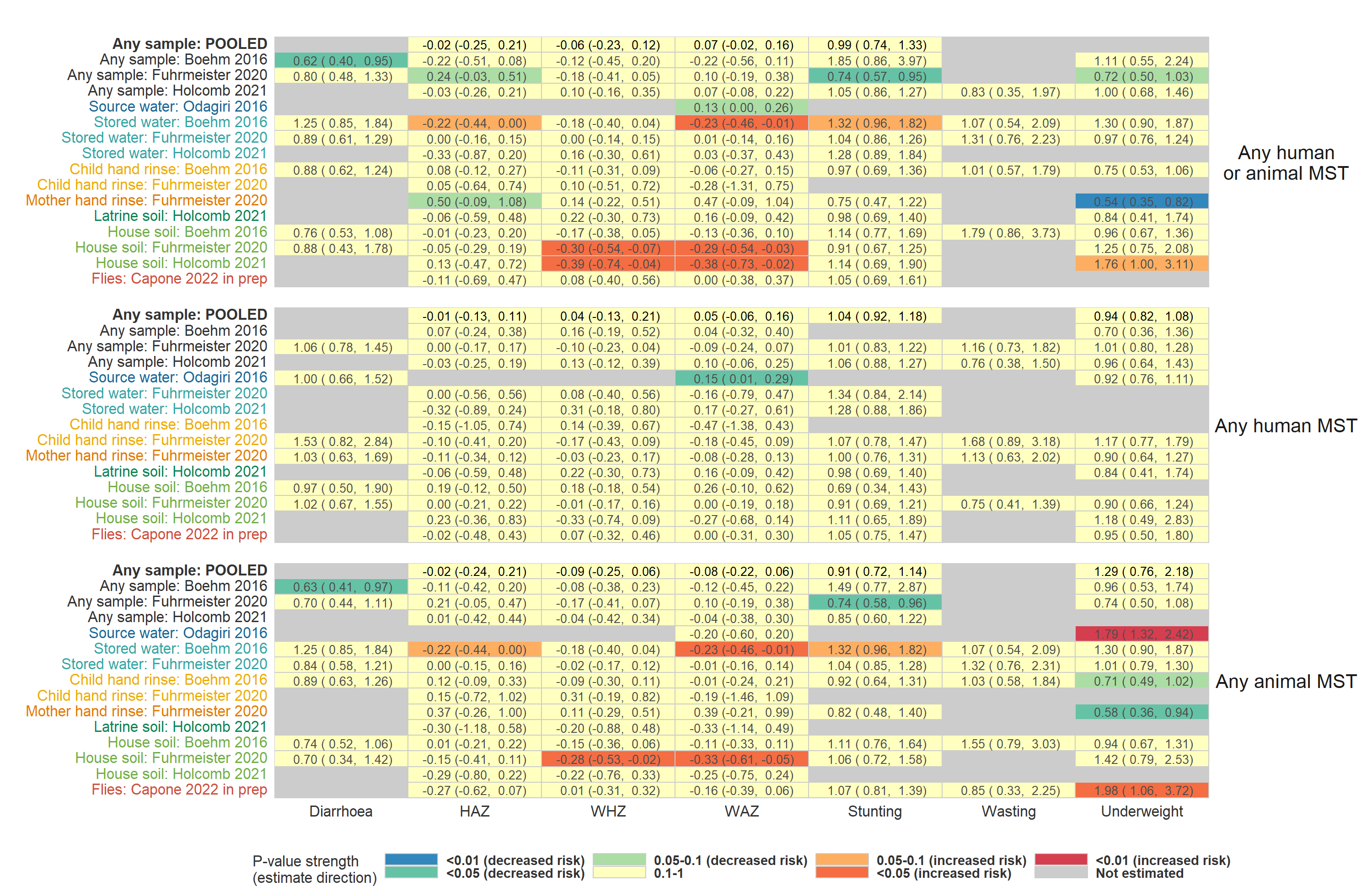


**Figure S2.** WASH intervention effects on child height-for-age Z-scores within the subset of children used in the primary analysis who had time-matched growth measurements and environmental samples.

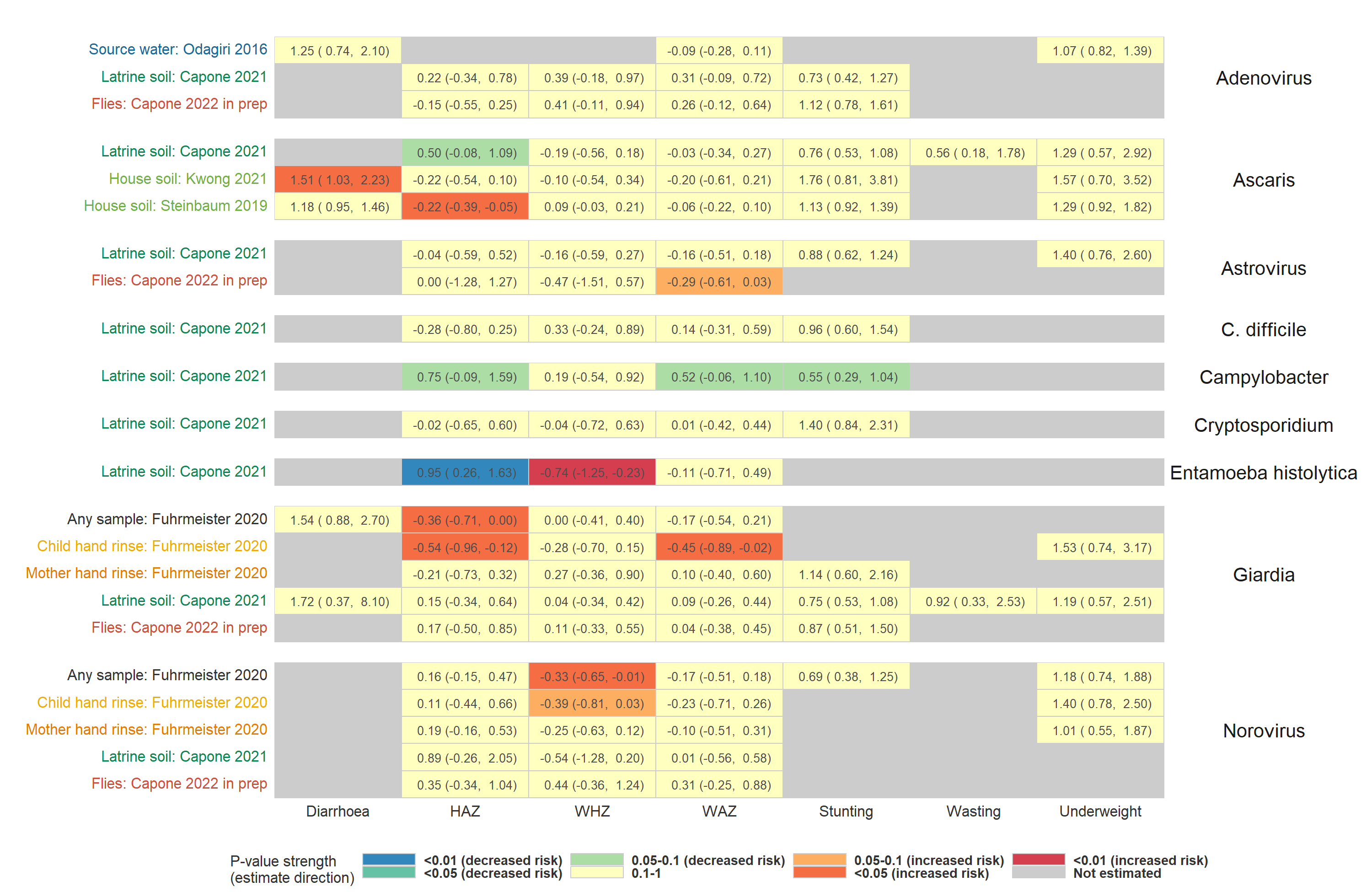


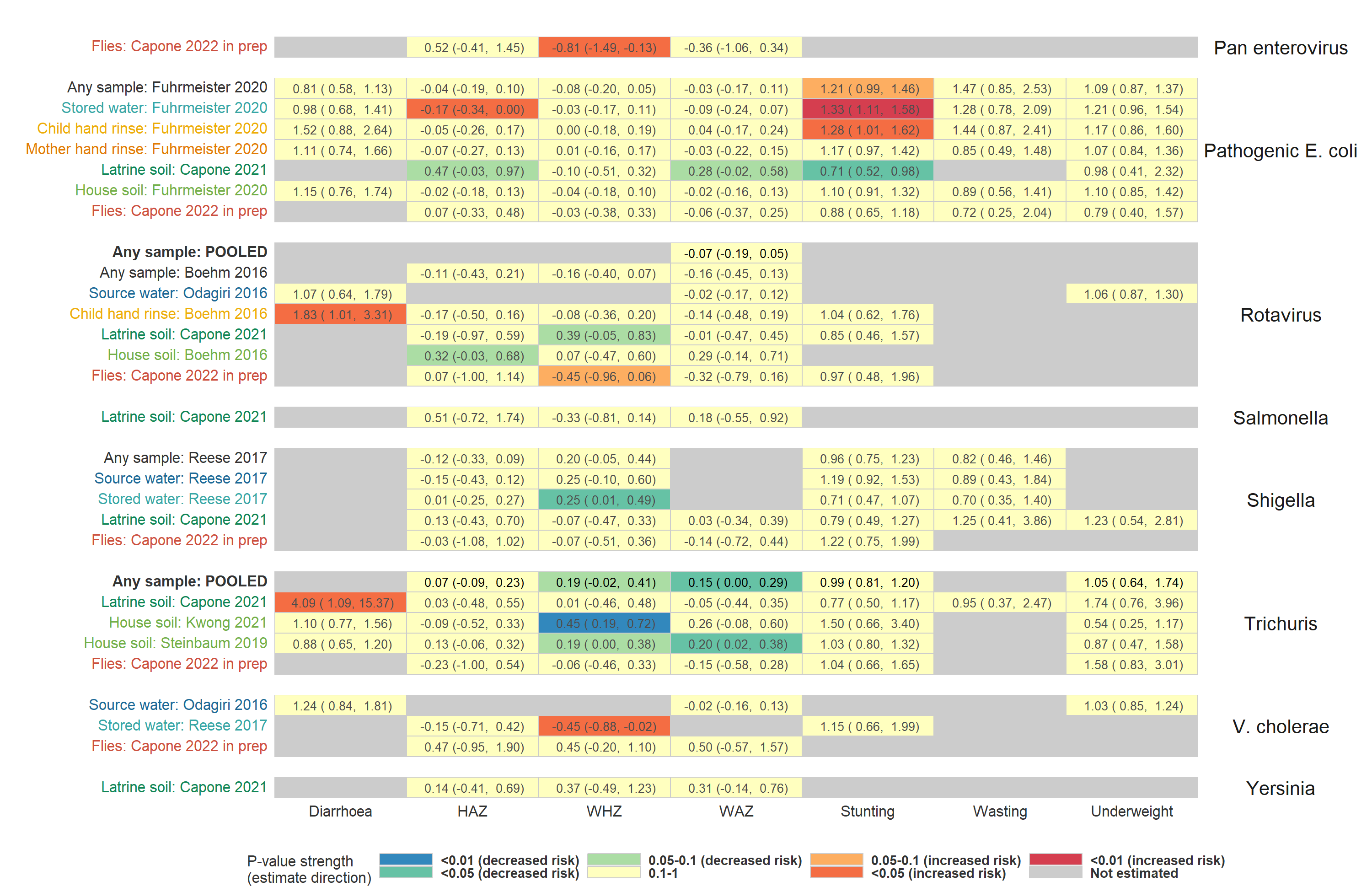
**Figure S3.** WASH intervention effects on child diarrhoeal disease within the subset of children used in the primary analysis who had time-matched diarrhoea observations and environmental samples.



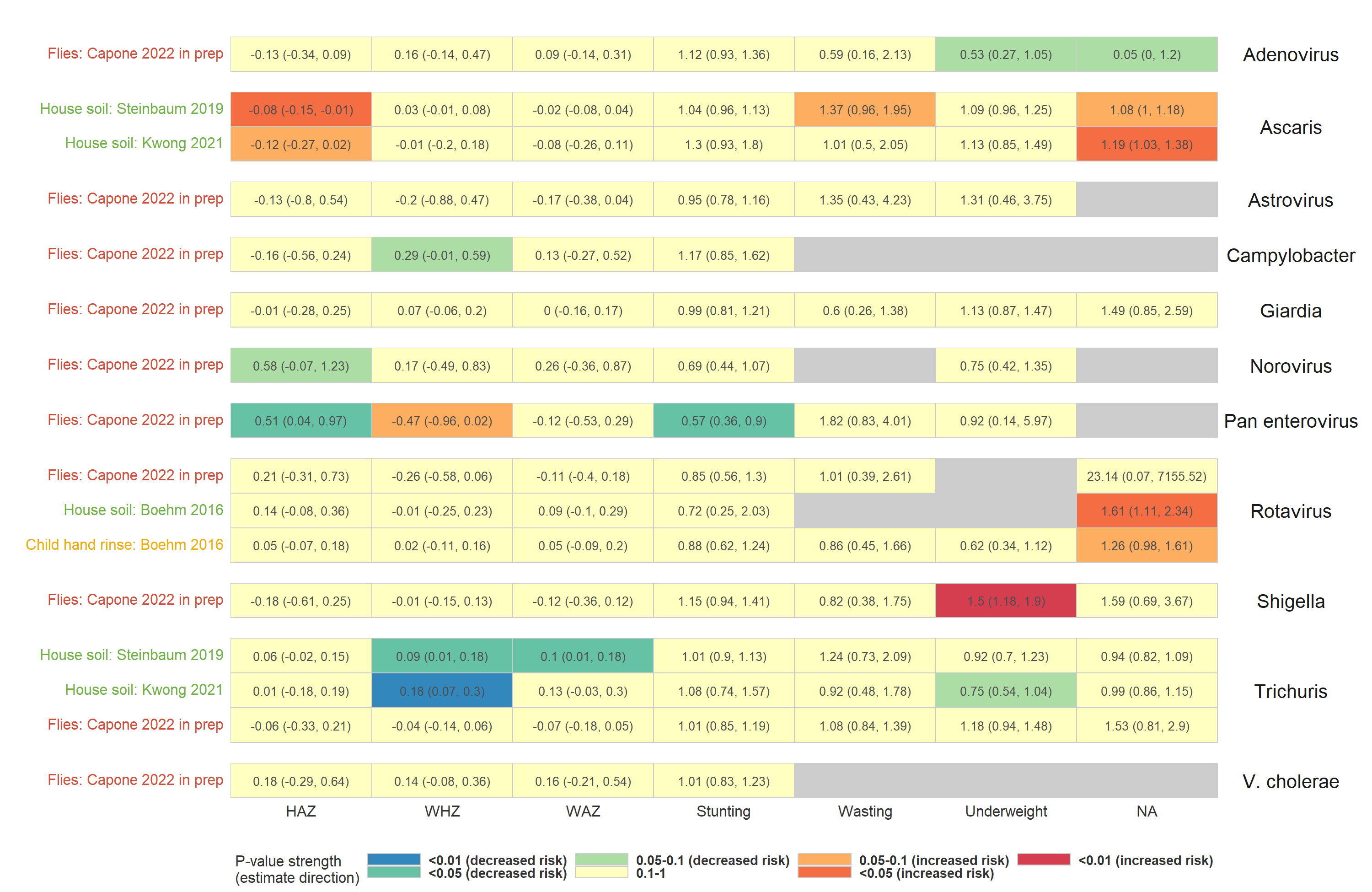


**Figure S4.** Heatmap of significance and direction of associations between aggregate measures of environmental contamination (rows) and child diarrhoea and growth outcomes (columns). Cells are colored by the strength of significance and direction of association, and the point estimate and confidence 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, and axis labels are colored by sample type, matching the primary figure legends. 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.

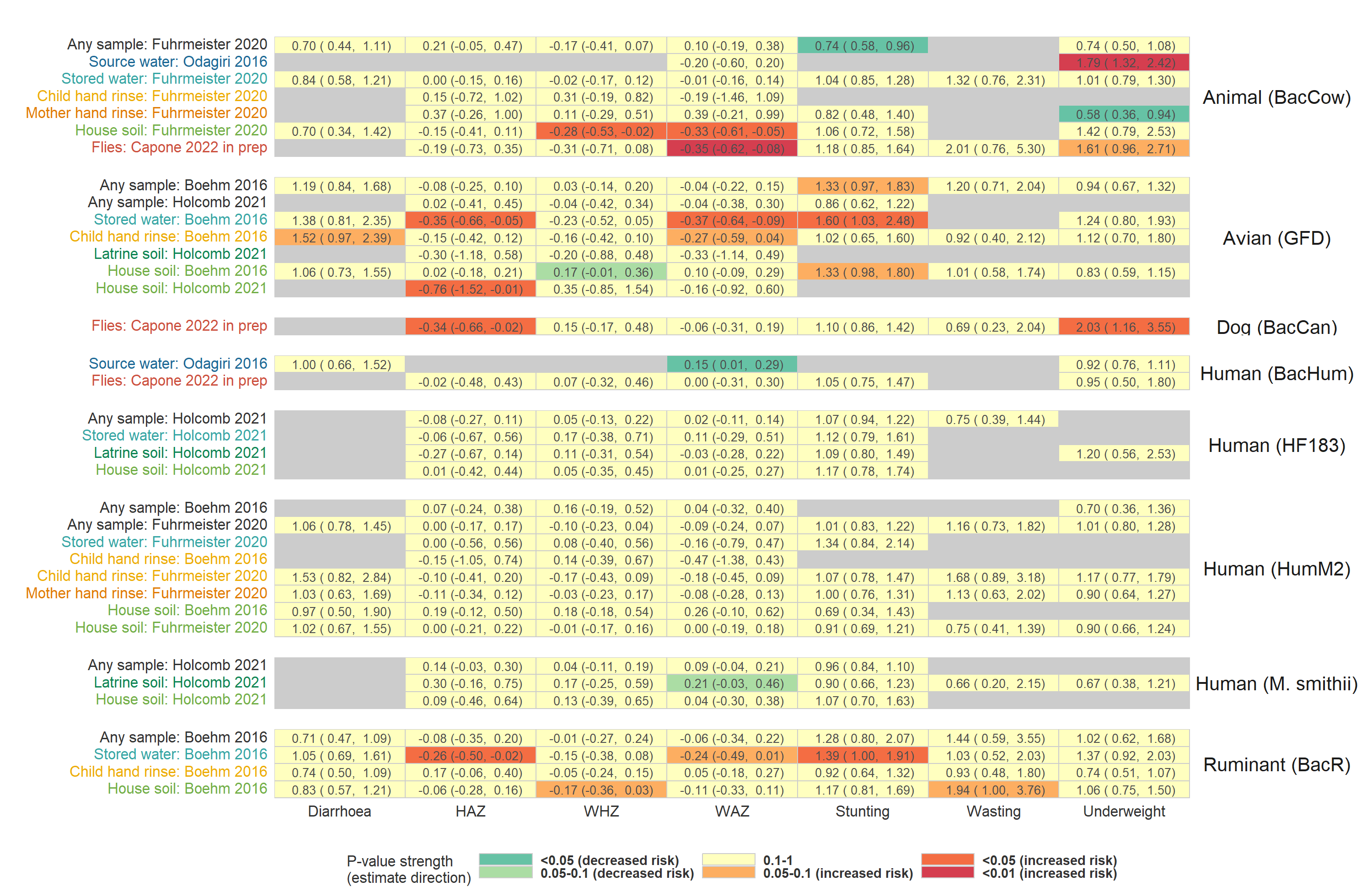




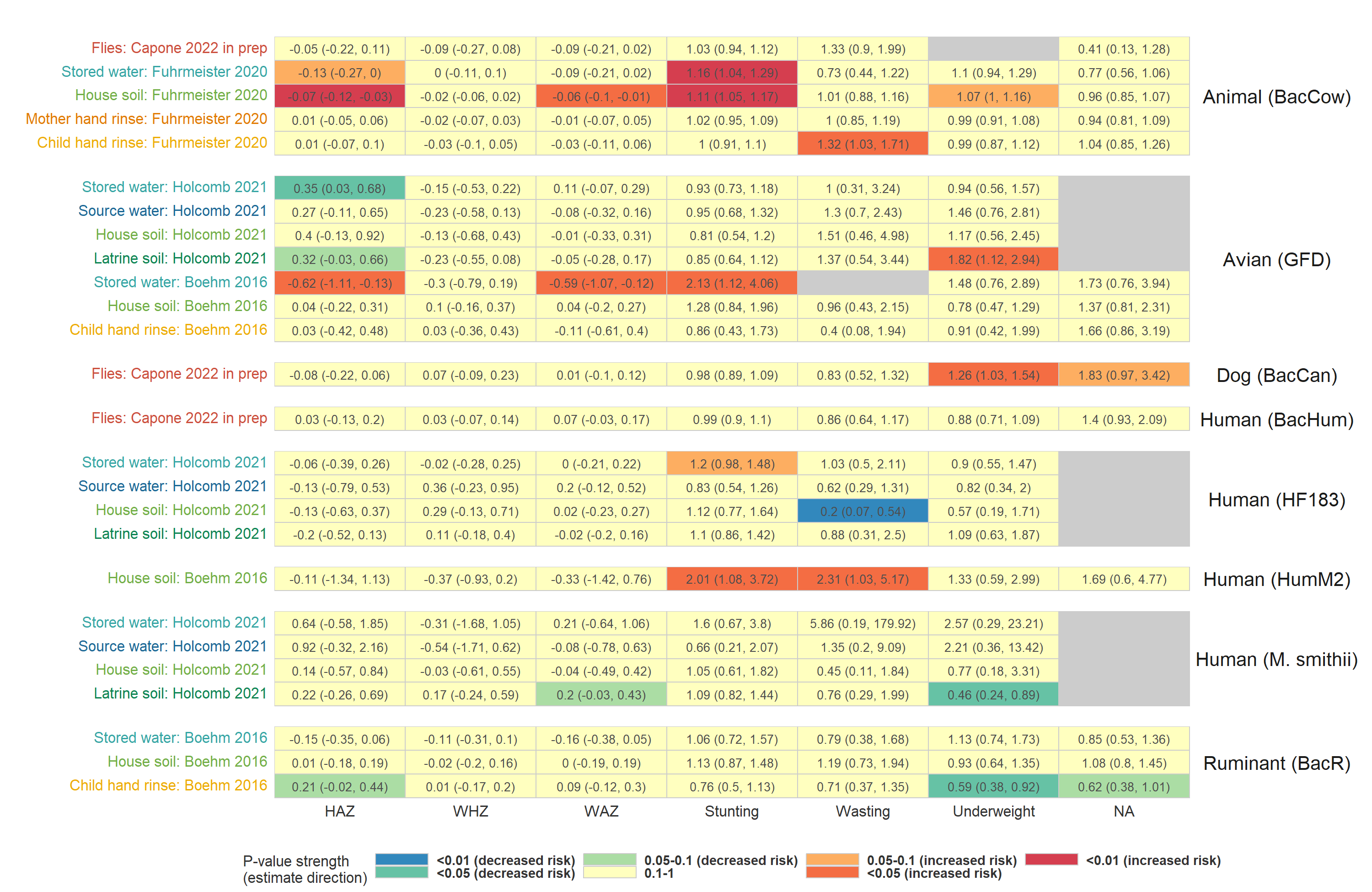
**Figure S5.** Heatmap of significance and direction of associations between specific pathogens in environmental samples (rows) and child diarrhoea and growth outcomes (columns). Cells are colored by the strength of significance and direction of association, and the point estimate and confidence intervals 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, and axis labels are colored by sample type, matching the primary figure legends. 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.



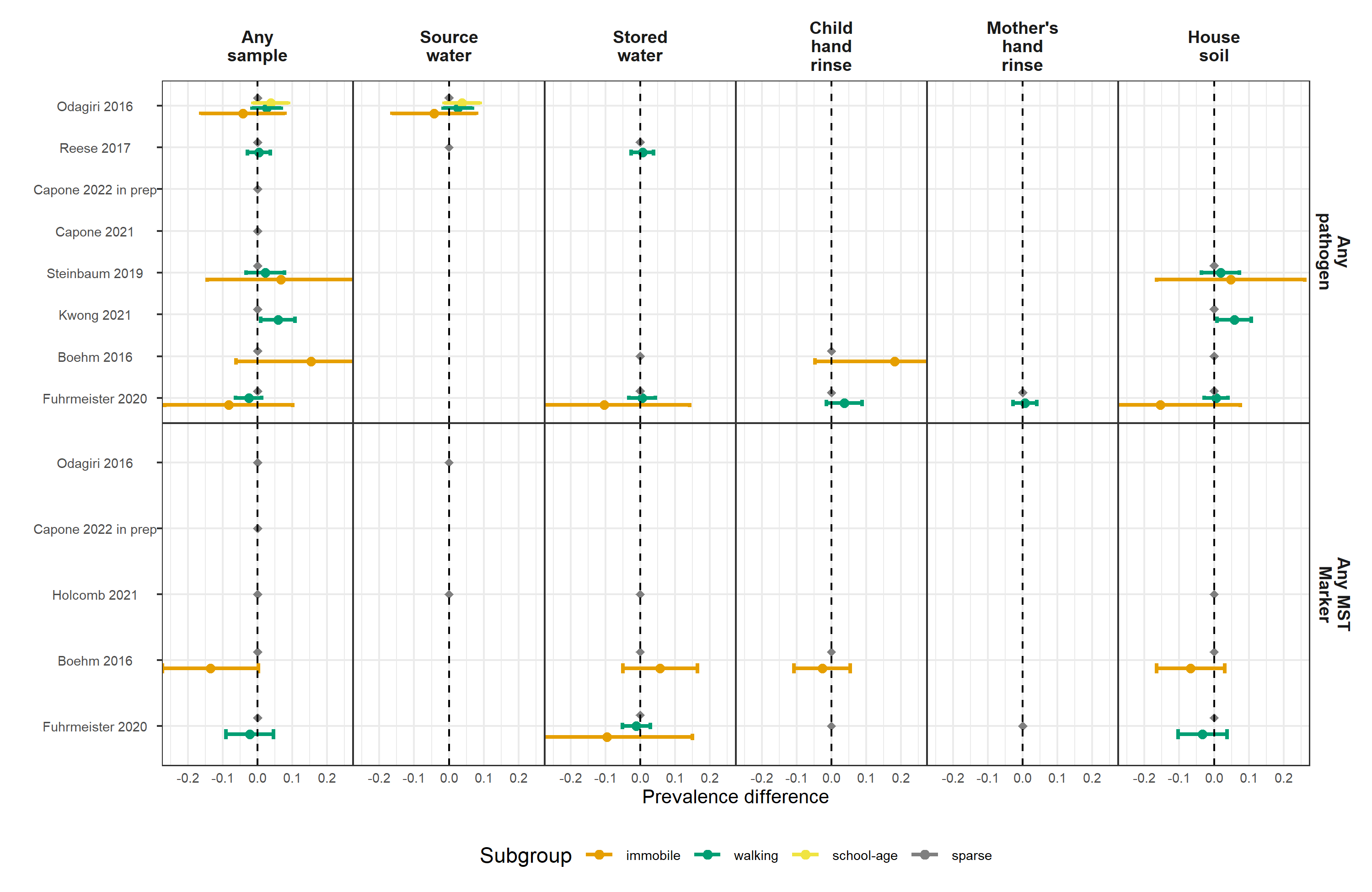
**Figure S6.** Heatmap of significance and direction of associations between the abundance of specific pathogens in environmental samples (rows) and child diarrhoea and growth outcomes (columns). Cells are colored by the strength of significance and direction of association, and the point estimate and 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, and axis labels are colored by sample type, matching the primary figure legends. 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.



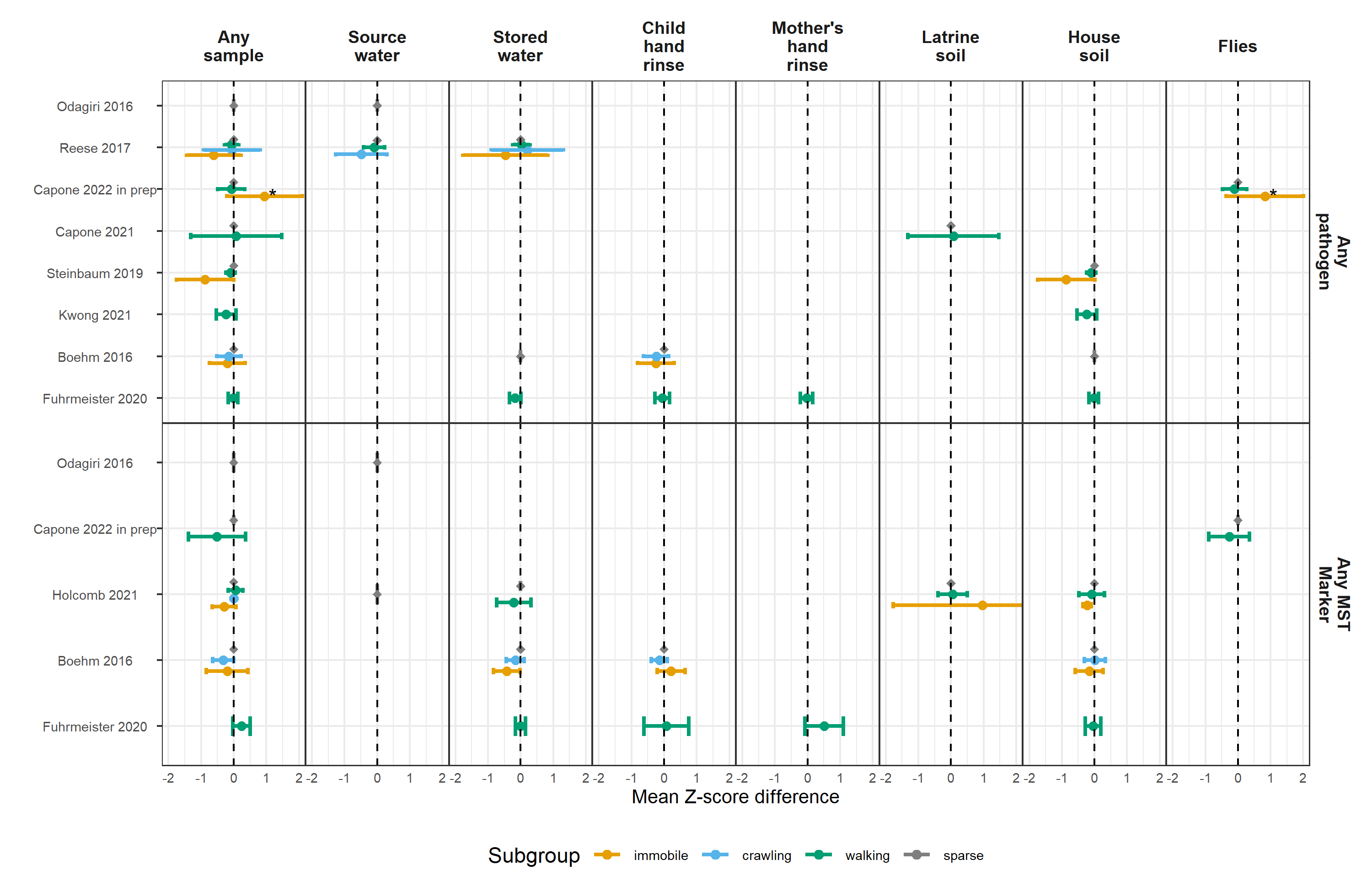
**Figure S7.** Heatmap of significance and direction of associations between specific microbial source tracking markers in environmental samples (rows) and child diarrhoea and growth outcomes (columns). Cells are colored by the strength of significance and direction of association, and the point estimate 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, and axis labels are colored by sample type, matching the primary figure legends. 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.



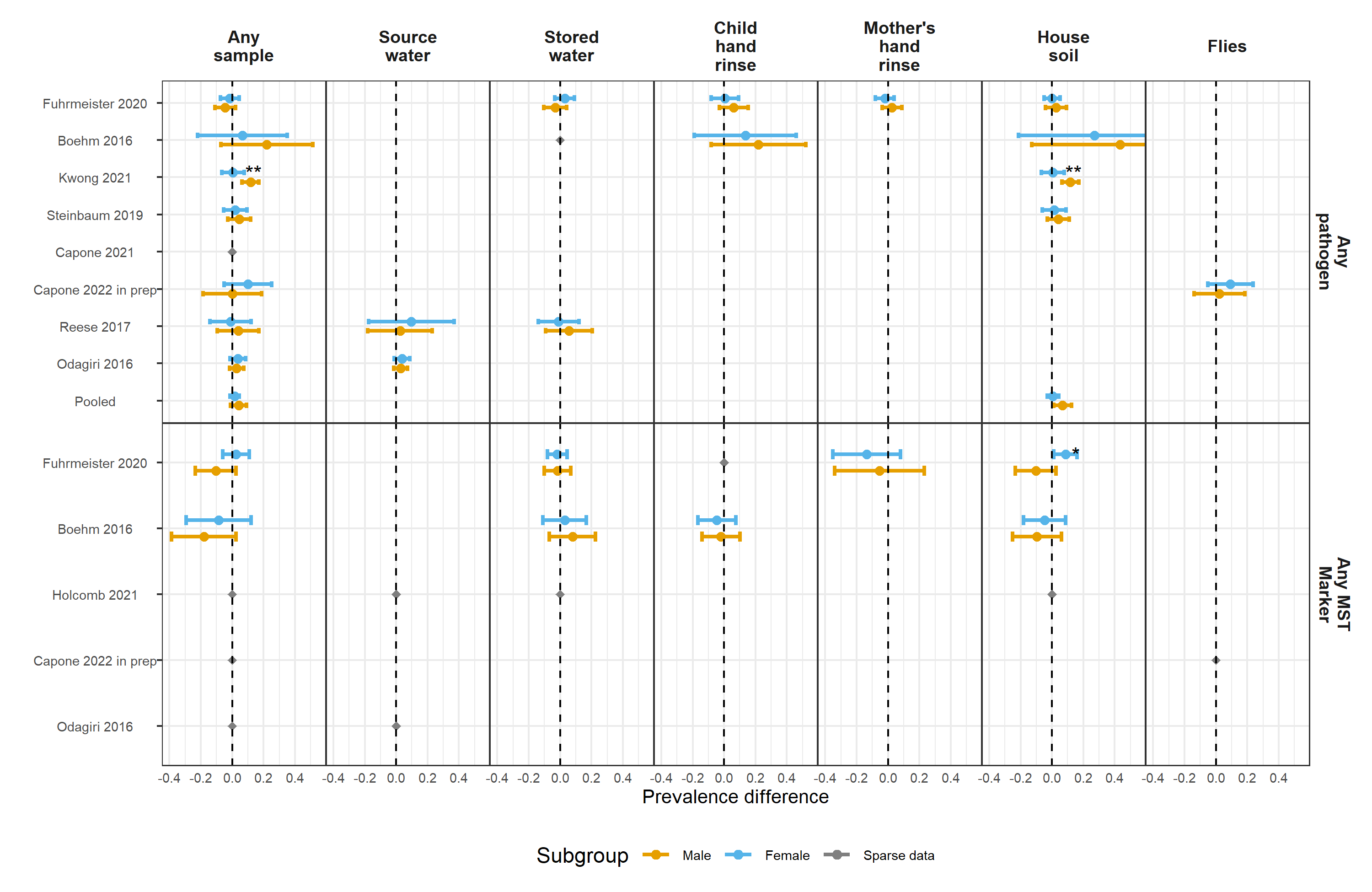
**Figure S8.** Heatmap of significance and direction of associations between the abundance of specific microbial source tracking markers in environmental samples (rows) and child diarrhoea and growth outcomes (columns). Cells are colored by the strength of significance and direction of association, and 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, and axis labels are colored by sample type, matching the primary figure legends. 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.



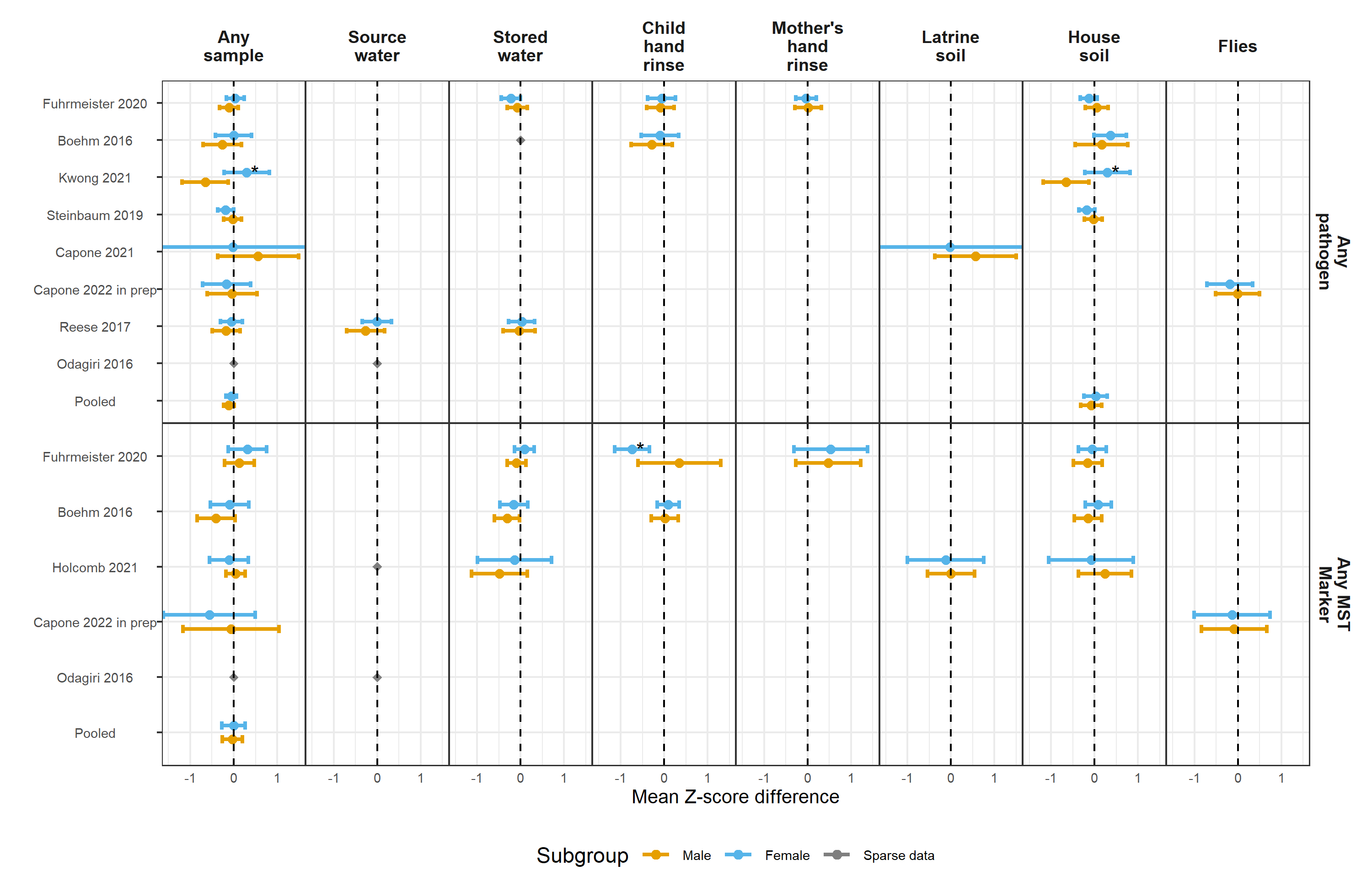
**Figure S9.** Forest plots of child diarrhoeal disease prevalence differences between environmental samples with and without any enteropathogen or any MST marker detected, stratified by child age. Grey points mark sparse age strata without estimated relative risks. Significant effect modification, as determined by the p-values on the regression model interaction term, is marked above points with asterisks (P < 0.05 = \*, P < 0.01 = \*\*, P < 0.001 = \*\*\*).



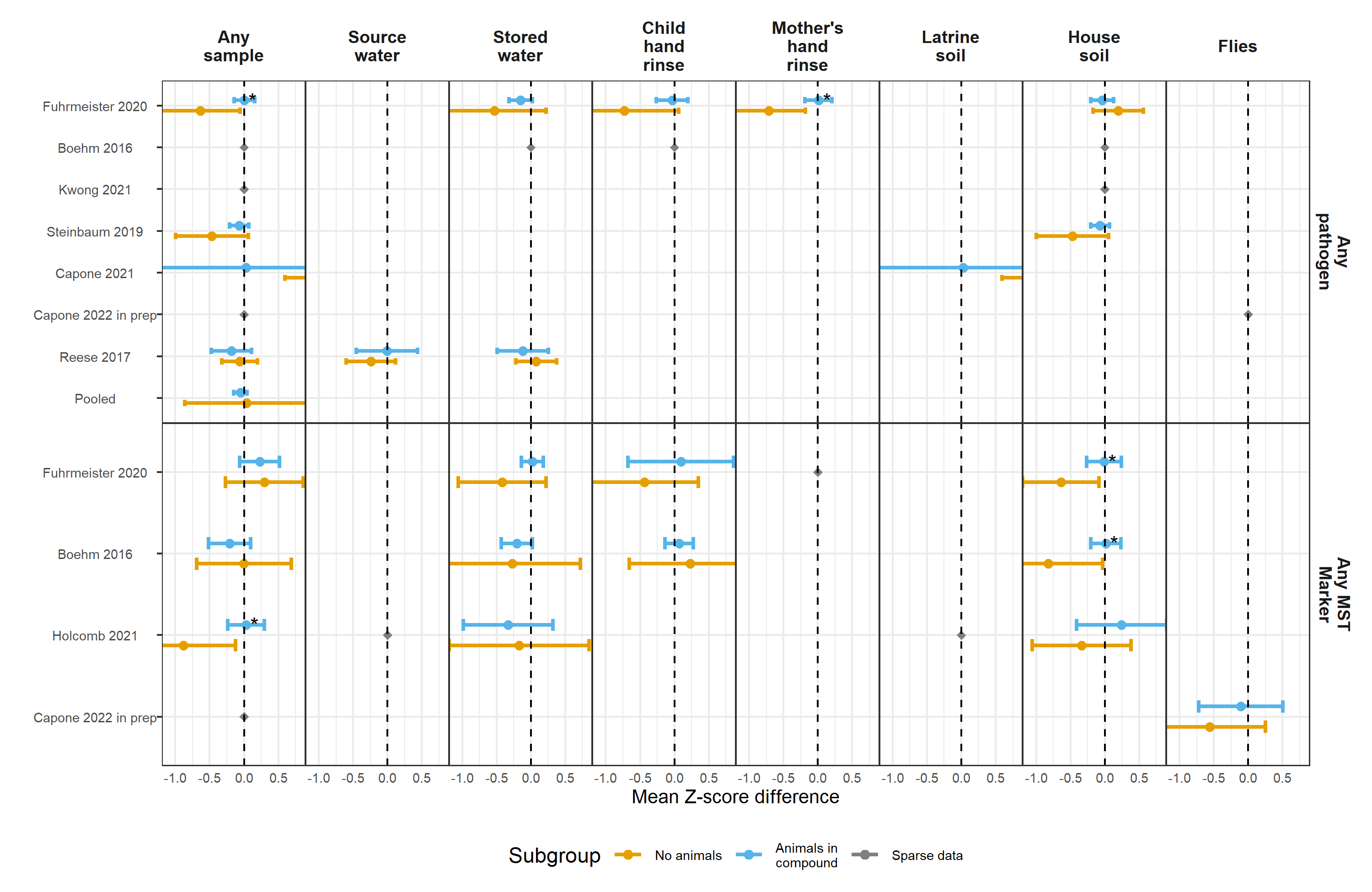
**Figure S10.** Forest plots of associations between any enteropathogen/any MST markers in different types of environmental samples and child height-for-age Z-score (HAZ), stratified by child age. Grey points mark sparse age strata without estimated mean differences. Significant effect modification, as determined by the p-values on the regression model interaction term, is marked above points with asterisks (P < 0.05 = \*, P < 0.01 = \*\*, P < 0.001 = \*\*\*).



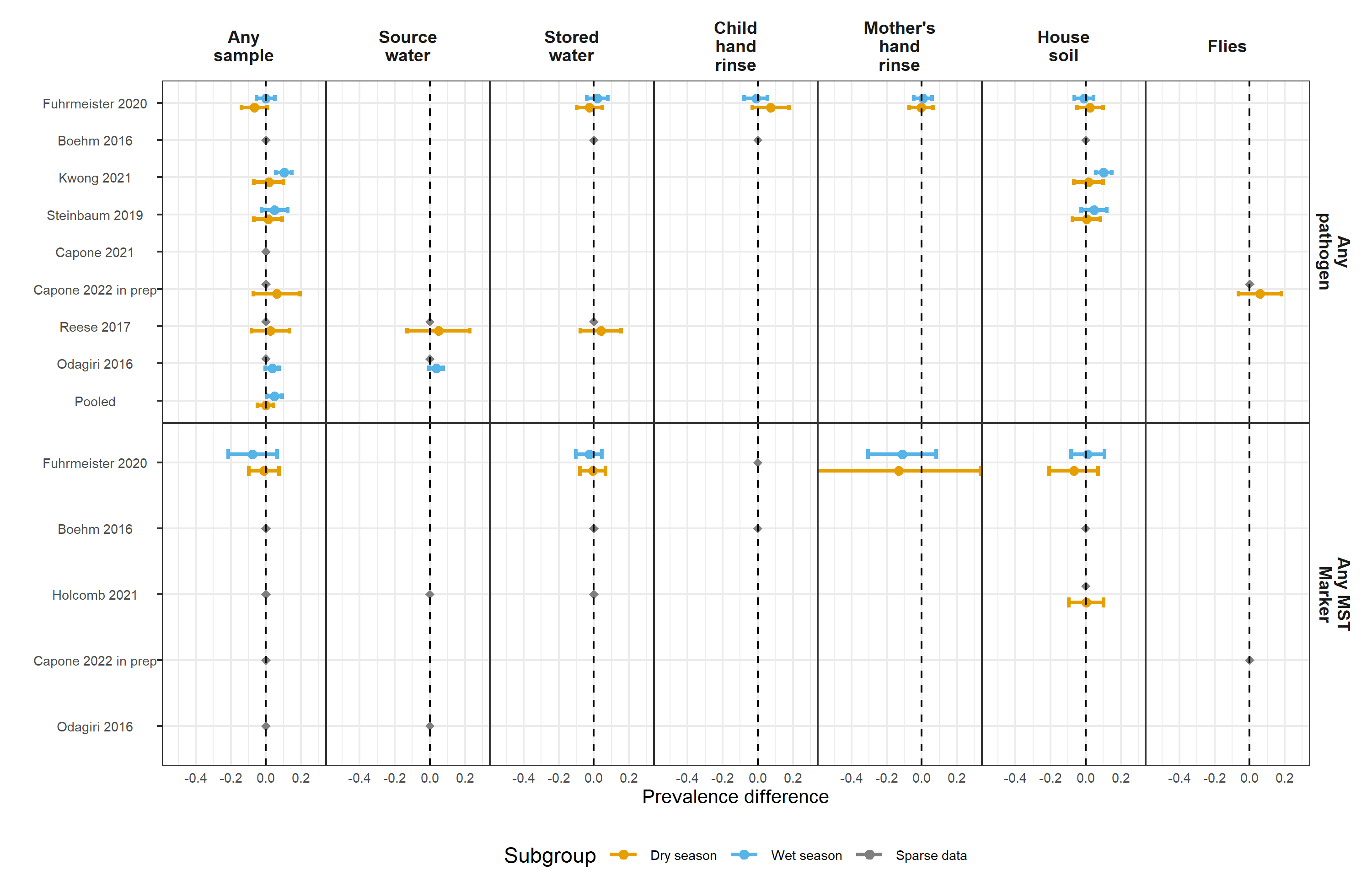
**Figure S11.** Forest plots of child diarrhoeal disease prevalence differences between environmental samples with and without any enteropathogen or any MST marker detected, stratified by child sex. Significant effect modification, as determined by the p-values on the regression model interaction term, is marked above points with asterisks (P < 0.05 = \*, P < 0.01 = \*\*, P < 0.001 = \*\*\*).



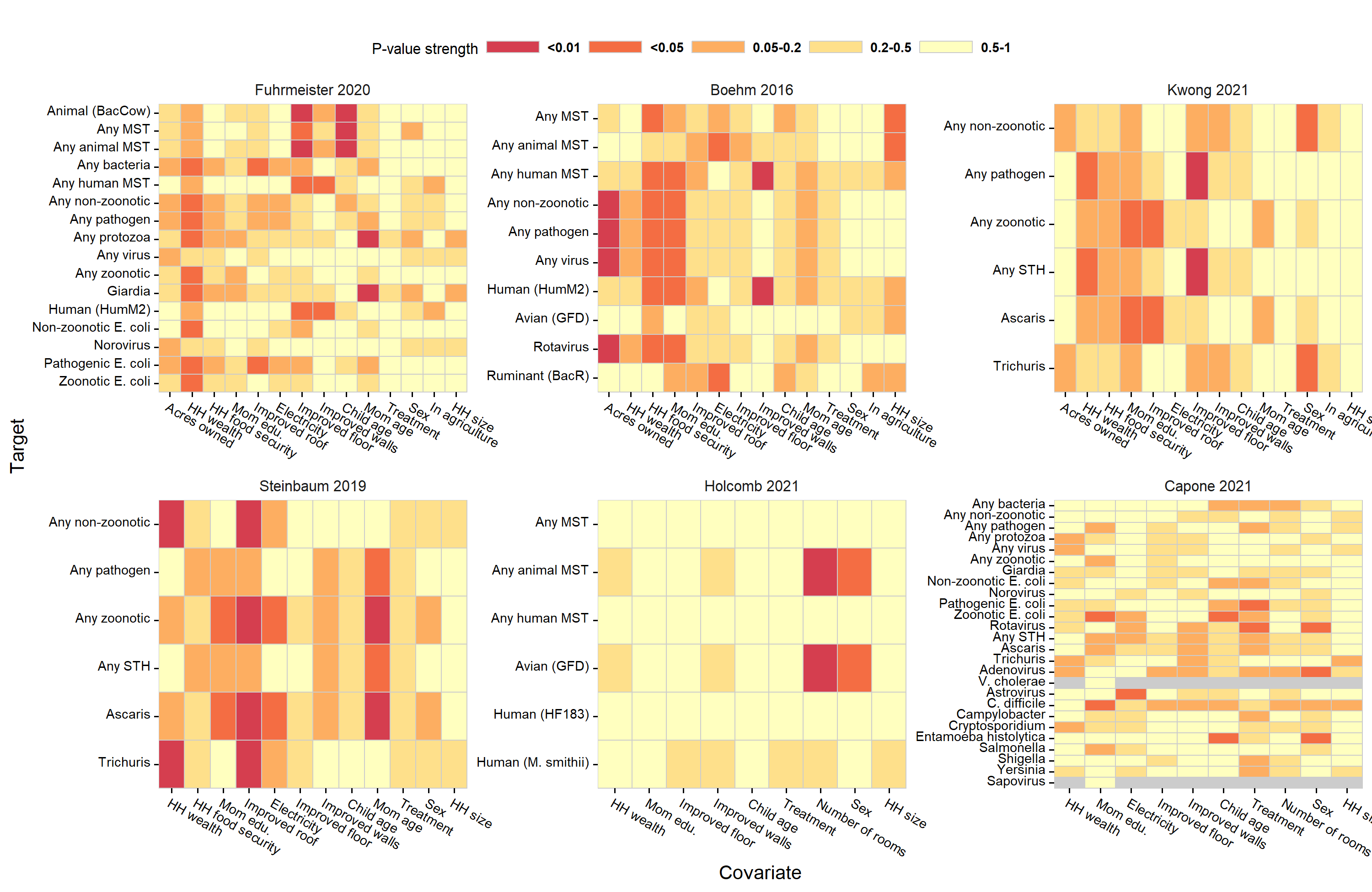
**Figure S12.** Forest plots of associations between any enteropathogen/any MST markers in different types of environmental samples and child height-for-age Z-scores (HAZ), stratified by child sex. Significant effect modification, as determined by the p-values on the regression model interaction term, is marked above points with asterisks (P < 0.05 = \*, P < 0.01 = \*\*, P < 0.001 = \*\*\*).

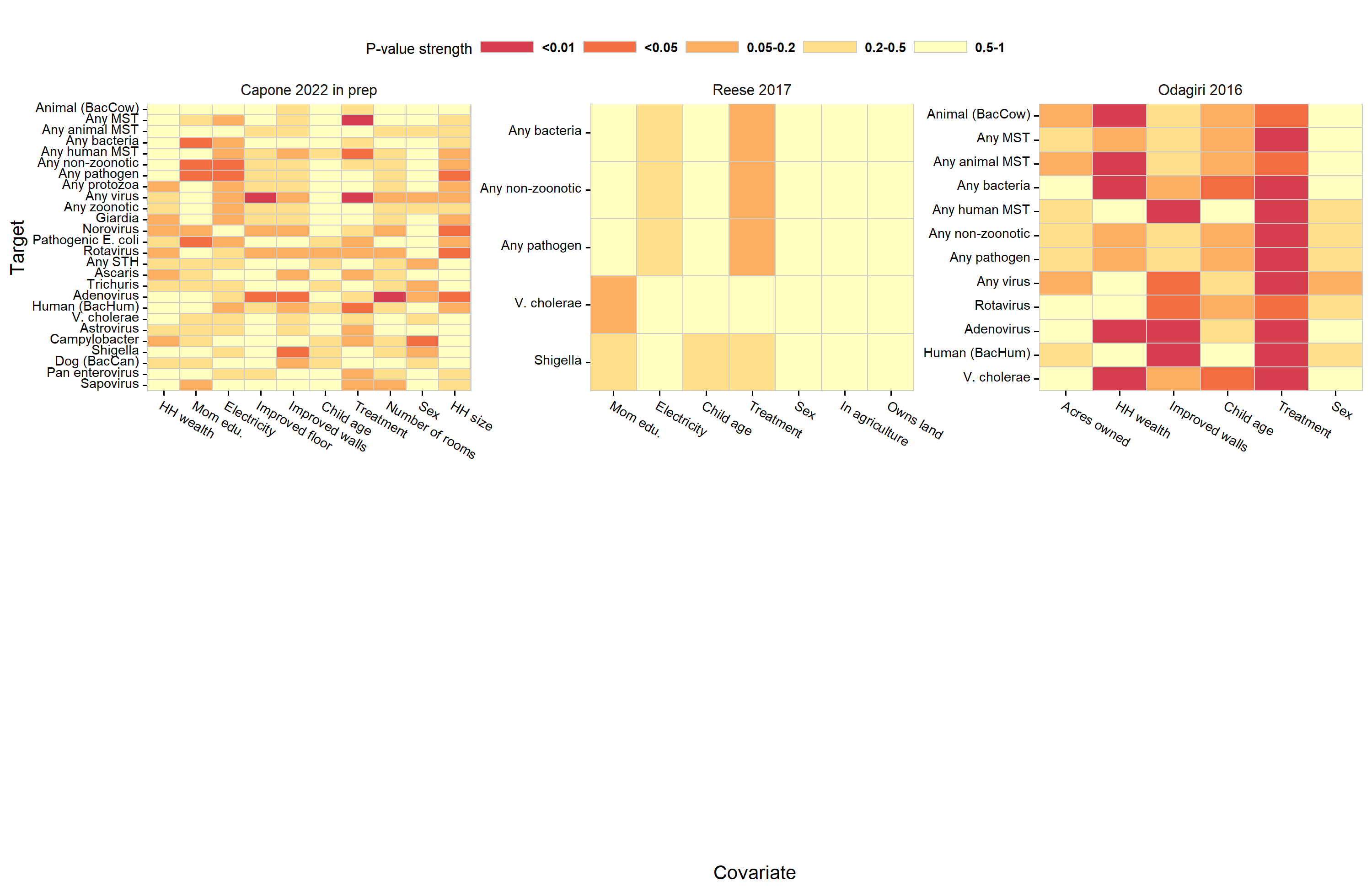


**Figure S13.** Forest plots of associations between any enteropathogen/any MST markers in different types of environmental samples and child height-for-age Z-scores (HAZ), stratified by whether any animals were present in the compound. Significant effect modification, as determined by the p-values on the regression model interaction term, is marked above points with asterisks (P < 0.05 = \*, P < 0.01 = \*\*, P < 0.001 = \*\*\*).

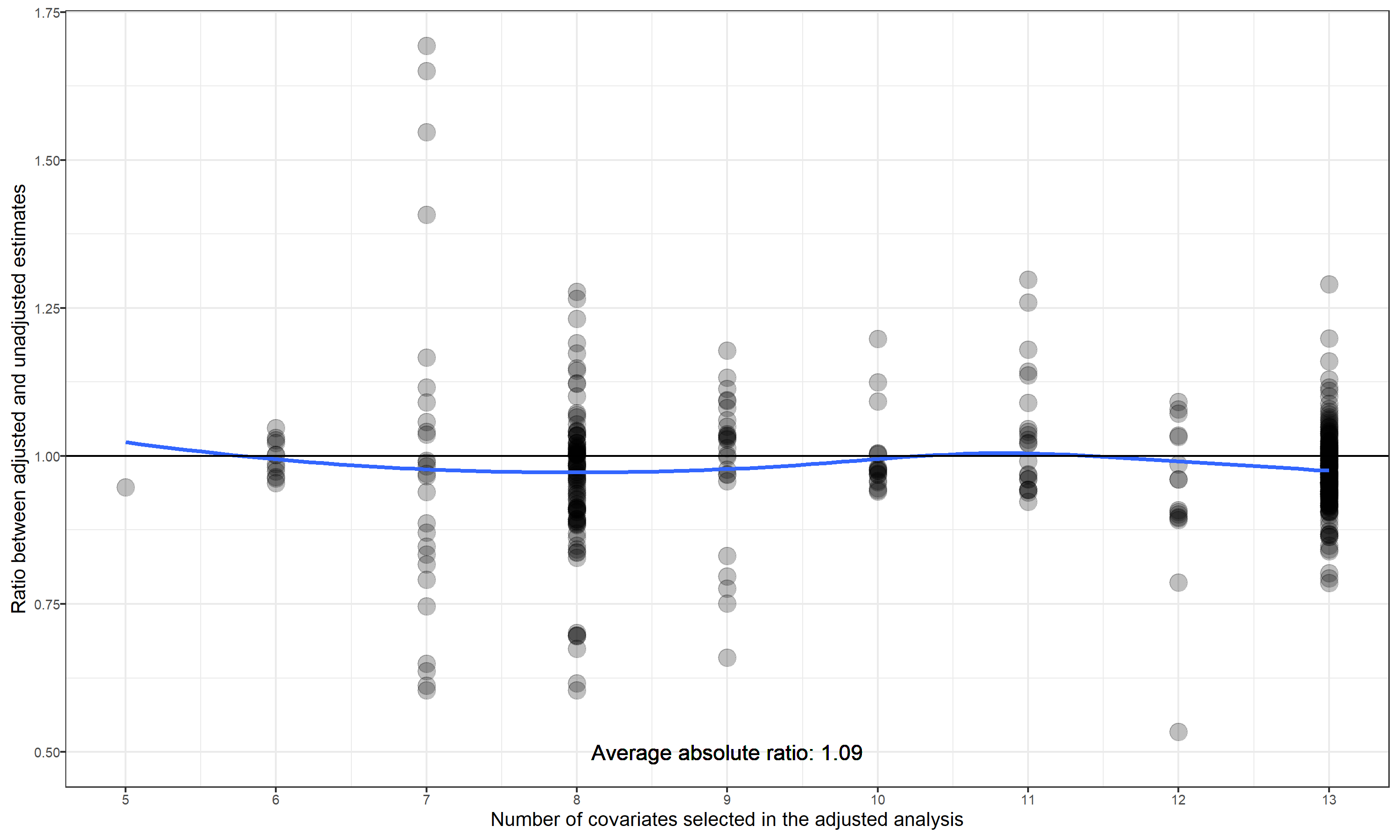


**Figure S14.** Forest plots of child diarrhoeal disease prevalence differences between environmental samples with and without any enteropathogen or any MST marker detected, stratified by whether the diarrhoeal disease occurred during the wet versus dry season (defined by the 6 months of highest average rainfall). Significant effect modification, as determined by the p-values on the regression model interaction term, is marked above points with asterisks (P < 0.05 = \*, P < 0.01 = \*\*, P < 0.001 = \*\*\*).

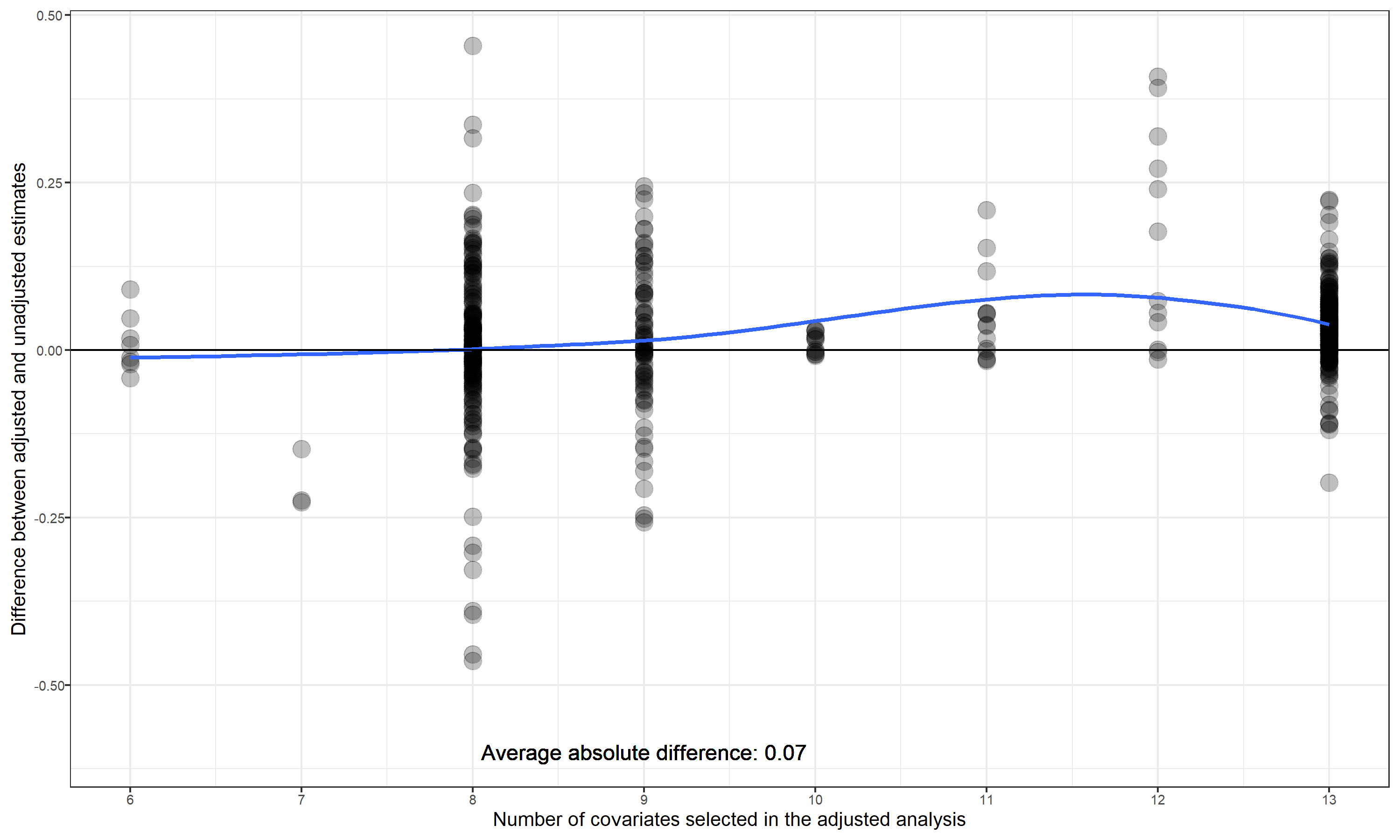




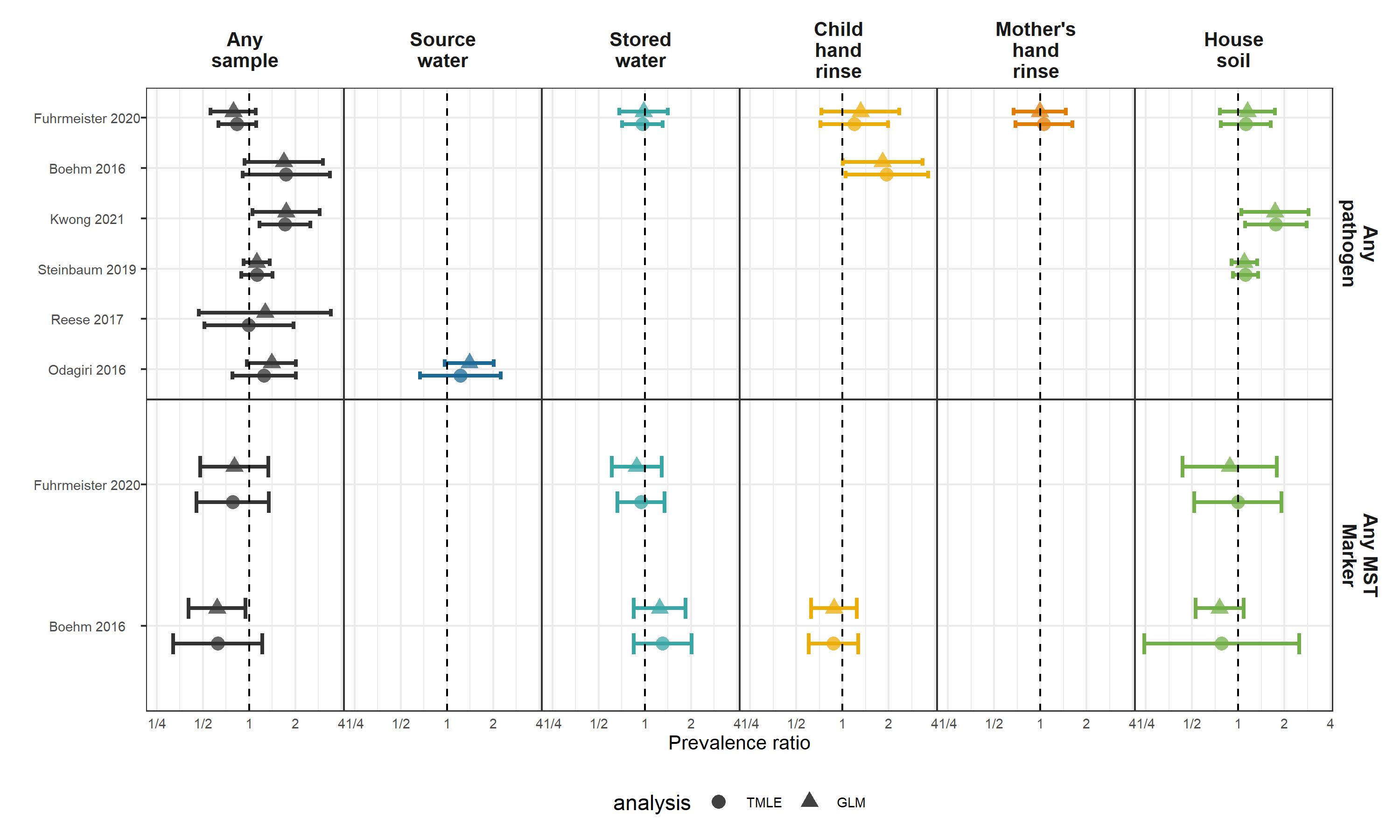
**Figure S15.** Study-specific associations between adjustment covariates and the presence of different enteropathogen and MST markers in aggregated environmental samples. The columns are different pre-screened confounders, and the rows are specific enteropathogens and MST markers. Cells of the heatmaps are colored by P-values of bivariate likelihood ratio tests, and heatmaps are stratified by study.



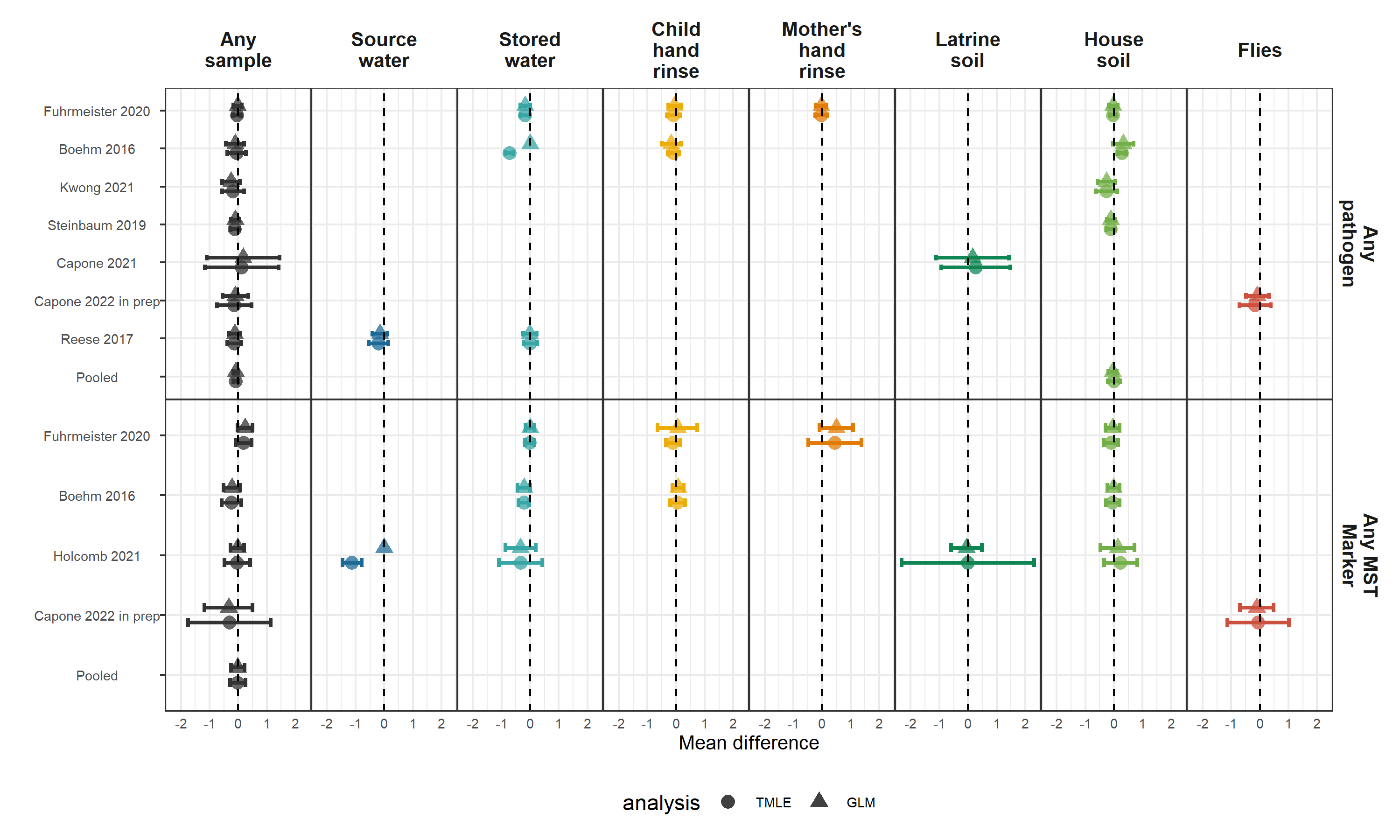
**Figure S16.** Comparison between associations estimated with and without including potential confounders for the binary diarrhoea and growth outcomes. Points mark the ratio of relative risks estimated using adjusted and unadjusted generalized linear models. The blue line shows the average ratio between adjusted estimates and unadjusted estimates, fitted using a cubic spline.



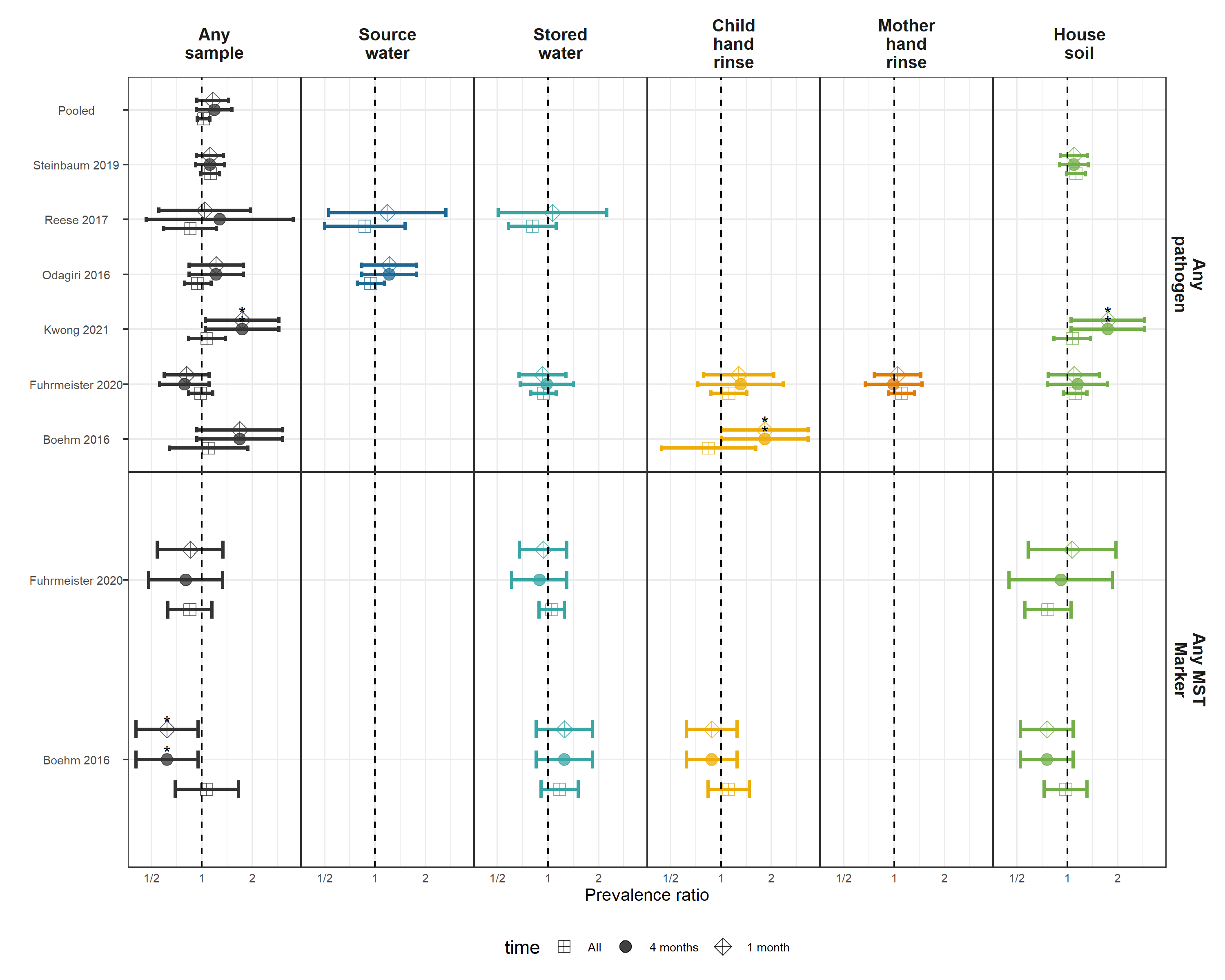
**Figure S17.** Comparison between associations estimated with and without including potential confounders for the continuous growth outcomes. Points mark the differences between mean differences estimated using adjusted and unadjusted generalized linear models. The blue line shows the average difference in differences between adjusted estimates and unadjusted estimates, fitted using a cubic spline.



**Figure S18.** Comparison between associations estimated with generalized linear models (GLM) and machine-learning based targeted likelihood estimation models (TMLE) for the diarrhoea outcome.



**Figure S19.** Comparison between associations estimated with generalized linear models (GLM) and machine-learning based targeted likelihood estimation models (TMLE) for the height-for-age (HAZ) Z-score outcome.



**Figure S20.** Comparison between associations estimated in the primary diarrhoea analysis (diarrhoeal disease occurring after environmental sampling, but no more than 4 months later with associations estimated only using diarrhoeal disease cases within 1 month, or occurring at any time). For the analysis of all diarrhoea, it included diarrhoeal cases, even cases occurring prior to sampling, under the hypothesis that enteropathogen presence at one time is a surrogate variable for general environmental contamination.