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)

## Abstract

**Background:** Drinking water , sanitation, and hygiene (WASH) improvements are considered cornerstones to reduce diarrheal disease in low-income countries. However, recent trials have found no or mixed effects of household- and community-level WASH interventions on child health. Assessing whether these interventions reduce enteropathogens in the environment as an intermediate variable can illuminate whether limited health effects occur because the interventions do not lead to a cleaner environment. **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 intervention effects and pooled results across studies. **Findings:** We identified and received data from five randomized or quasi-experimental studies. Most individual enteropathogens or MST markers were not associated with diarrheal disease or child growth, but **(pooled association)** and **(specific enteropathogens)**

There was no overall effect on MST markers, and no consistent differences in intervention effects by season, animal presence, urbanicity, study design, or intervention uptake.

**Interpretation:** Few trials have measured enteropathogens or host-specific fecal markers in the environment. The small effect of sanitation interventions on enteropathogens in the environment is consistent with the lack of health impact in sanitation trials. **Funding:** The Bill & Melinda Gates Foundation.

## Research in context

**Evidence before this study.** Children in areas with (sentence on env. pathogens may affect diarrheal disease). (Mention results from this field and rec water field) Quantifying the associations between general and specific enteric pathogens in environmental samples and measures of child health may help elucidate if (something on the identification of risky areas, or using env. sampling to assess the impacts of interventions or WASH programs). Most previous studies and meta-analyses to date on the effect of WASH interventions on fecal contamination in the environment have focused on fecal indicator bacteria, showing (FIB and health results, aim 1 manuscripts mentions limited association, look at references/discussion of this). We conducted a systematic review and individual participant data meta-analysis of WASH intervention studies that measured enteropathogens or microbial source tracking (MST) markers in the environment and child diarrheal disease, enteropathogen infection, and growth outcomes to see if environmental pathogens or MST markers were associated with poor child health.

**Added value of this study.** We successfully obtained data from 6 out of 7 ( **double check number** ) eligible intervention studies identified in our systematic review that measured enteropathogens and MST markers in environmental samples and child diarrheal disease and growth outcomes. ( **XXX Update the following for aim2 results: Most individual studies indicated a protective effect of interventions on the prevalence of individual pathogens and MST markers, but most estimates were not statistically significant due to small sample sizes and rare detection of some of the targets. The individual participant data meta-analysis design of our study allowed us to detect a small but significant reduction in the prevalence of any type of pathogen in any type of sample by pooling across all studies. There was no overall intervention effect on the prevalence of MST markers. This study takes advantage of recently developed diagnostic methods to enumerate enteropathogens and host-specific fecal markers in a range of environmental samples, including understudied environmental reservoirs such as soil, to provide the first synthesis of evidence on the effect of WASH interventions on these specific targets in the domestic environment.**)

**Implications of all the available science.** ( **Update based on aim 2 results**) The small reduction we observed in pathogen prevalence in the environment when pooled across all studies may explain the small effect the interventions had on child health. These findings also validate previous findings of no effect from sanitation interventions on fecal indicator bacteria in the environment, further demonstrating the insufficiency of basic sanitation solutions in reducing fecal contamination in the environment. Possibly, more intensive WASH interventions like safely managed water and sanitation are needed to reduce environmental contamination enough to improve child health. We note that only a small number of trials met our inclusion criteria and only a subset of households was sampled in each study. Pathogen targets and diagnostic methods varied by study. Future research would benefit from sampling a more diverse set of WASH interventions using a standardized set of laboratory methods to enumerate a common range of pathogen and MST targets.

## Summary

-Pathogen-specific significant associations -Overall associations are null -specific significant findings

Most study-specific estimates are null, with inconsistent direction of effects in significant associations. Estimates pooled over multiple studies were also null, except for a small and marginally significant association between any enteropathogen in any sample and lower child height-for-age Z-scores (which is significant without adjustment for confounders).

## 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. The analyses was conducted by sample type (e.g., water, hands, soil) and pooled across study types, and used data from all study arms.

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.

Given the heterogeneity in study settings (e.g., local WASH conditions, climate, urbanization, population density, region-specific infectious disease patterns, intervention designs), 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. If there was evidence for heterogeneity but there was qualitative support for combining studies, we pooled estimates using random-effects models.

## Results

* (Talk about the data availability and then results from tables on N’s. Discuss timing of measurements, ref. tables)
* talk about diarrhea prevalence by study and growth mean/stunting prev. by study (and mainly reference table). Make sure to make dynamic in R.
* Paragraph on main results on pathogen and MST associations with diarrhea and then HAZ. Make sure to use both main and sample-specific findings \* 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). \* Presence of any enteropathogen (but not any mst marker) in any environmental sample is significantly associated with lower HAZ when pooled across studies (Adjusted mean difference: (95% CI: , )) (Figure 2). This is 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. Nevertheless, water samples with any enteropathogen presence were significantly associated with lower mean HAZ in Boehm et al. 2016. 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 .
* Paragraph on pathogen specific samples and infections
  + There is 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 associations between environmental presence and child infection, but associations were significant or near significant for *Shigella*, *Ascaris*, and *Trichuris* contaminations 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.
* Paragraph on key patterns in groups of pathogens and MST markers \* Presence of types of enteropathogens in environmental samples were not associated with diarrheal disease, except any viral enteropathogen presence located on child hands in Boehm et al. 2016 (Figure 4). \* No associations between specific groups of MST markers and child diarrheal disease in any sample type (Figure 5). \* When separated out by group of enteropathogen, enteropathogen presence in any environmental sample is no longer significantly associated with lower HAZ when pooled across studies (Figure 6). However, any virus presence in water 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 trial. \* There is a general trend of presence of groups of MST markers in water samples being associated with lower mean HAZ (otherwise estimates are null?) (Figure 7).
* Paragraph on key patterns in specific pathogens \* Most associations are 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 8). \* There are inconsistent associations between specific enteropathogens and child HAZ scores, with most estimates having null effects (Figure 9). Of the statistically significant associations, half were associated with increased linear growth and half were associated with decreased linear growth. \* 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 by positive association in all three sample types from Boehm et al 2016 (Figure 10). \* There are inconsistent associations between the presence of specific MST markers and child HAZ scores, with most estimates having null effects (Figure 11). 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 diarrheal disease, with most estimates having null effects, and with significant effects occurring in both harmful and protective directions (Figure 12).
* Abundance analyses \* 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 (Figure 13). \* There are inconsistent associations between the abundance of any enteropathogen or any MST marker and child WHZ scores, with most estimates having null effects, and with significant effects occurring in both harmful and protective directions (Figure 14). \* Most associations between the abundance of any enteropathogen or any MST marker and child WHZ scores were null effects, but any MST marker in soil in Furhmeister et al. and any enteropathogen or any MST marker in stored drinking water in Boehm et al. 2016 were associated with children having a significantly lower WAZ (Figure 15).
* Paragraph on key patterns in secondary outcomes \* While most study- and sample-specific estimates were not significant, any enteropathogen presence in any sample was associated with a significant increased risk of stunting when pooled across studies (Relative risk: (95% CI: , ), Figure 16). MST marker presence in samples was not associated with stunting, except in any sample in Fuhrmeister et al. 2020 where presence was unexpectadly associated with a reduction in stunting. \* enteropathogen and MST marker presence was not associated with wasting (Figure 17). \* enteropathogen presence was not associated with child underweight status, but MST marker presence in samples was unexpectadly associated with with child underweight status in Boehm et al. 2016 and in any sample and in samples from mothers’ hands in Fuhrmeister et al. 2020. (Figure 18).
* 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 19). There was no association with MST markers 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 20). There was no association with MST markers. Note that there were insufficient cases of diarrheal disease in household without animals (rare) to conduct a similiar subgroup analysis for the diarrheal disease outcome. \* **NOTE: MOVE THIS TO METHODS** 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. 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 (Figure 21). Most associations were null regardless of the time window.
* 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 (Figure 22). \* Associations between adjustment covariates and the presence of different enteropathogen and MST markers in different environmental samples (Figure 24). 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 adequetly adjusted for measured confounding, but unmeasured confounding may bias the results. \* The average associations were slightly larger in magnitude after covariate adjustment (Figure 25-26). 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.
* XXXX

Adjusted diarrheal disease prevalence ratio for any enteropathogen presence in any environmental sample: (95% CI: , )

Adjusted HAZ difference for any enteropathogen presence in any environmental sample:

(95% CI: , )

Significant stunting: (95% CI: , )

(95% CI: , ) (95% CI: , )

Interquartile range for the ratio of adjusted and unadjusted prevalence rations: [0.97, 1.05] Interquartile range for the difference of adjusted and unadjusted prevalence rations: [-0.014, 0.061]

-percent significant of all estimates

Percent of diarrhea estimates significant 6.5

Percent of diarrhea estimates significant and harmful 5.9

Percent of HAZ estimates significant 11.2

Percent of HAZ estimates significant and harmful 7.2

-look at wasting/whz under 4 months -look stratified by animal presence.

## Discussion

* Make sure to look at if the WAZ outcomes (which are in odigari instead of HAZ) are sig for odigari

## Overall notes on data availability:

* Odagiri et al. 2016 only measured weight, so we only have WAZ, and Reese only measures/shared height, so we only have HAZ.
* The tables at the bottom of the report show the number of samples and number of health outcomes by study the column for both positive sample and diarrhea measure is likely the limiting factor for sparse analyses. (Give specific numbers for these tables)

## Notes on analysis

* The analysis included baseline (pre-intervention) measurements.
* All primary estimates are adjusted for intervention arm and child and household covariates.
* Only child health measurements taken after environmental samples were used.
* Diarrhea measurements must have occured after environmental samples, but within 4 months of environmental sample collection.
* Environmental samples were matched to the most proximate child health outcome, without using multiple measurements. For example, environmental samples at baseline were matched to child anthropometry and midline, but not endline.

## Notes on time ordering of environmental samples and child health outcomes, and data merging by study

#### WASH Benefits Bangladesh

* Endline (year 2) anthropometry and diarrhea was used for Kwong et al. 2021 (STH samples)
* World Bank substudy diarrhea and anthropometry was used for Boehm et al. 2016
* R01 substudy diarrhea and anthropometry was used for Fuhrmeister et al. 2020. The substudy was conducted over 8 rounds taken around 3 months apart, with environmental sampling occurring in rounds 3 and 4. Environmental samples were merged to diarrhea from the subsequent round and anthropometry from the main trial endline (year two) sampling.

#### WASH Benefits Kenya

* Endline (year 2) anthropometry and diarrhea was used.

#### Mapsan

* The Mapsan trial had three sampling rounds, baseline, midline, and endline, each 12 months
* The Mapsan trial environmental sampling data is divided into three studies which had differences in samples, microbial targets, and sampling times, Holcomb et al 2020 (baseline and midline), Capone et al 2021 (baseline and endline), and Capone et al 2021 in prep. (baseline and midline).
* Diarrhea was used from concurrent rounds, while anthropometry was used from subsequent rounds, except for endline environmental samples, where concurrent anthropometry was used.

#### Odisha

* Environmental samples were shared already merged with child health data, but samples outside of the specified time range for diarrhea or taken before environmental

#### Gram Vikas

* Sampling rounds were approximately 4 months apart, so anthropometry data was taken from subsequent round, and diarrhea data was taken from either the current or subsequent round, based on which sample was taken after but closer to the environmental sampling, and within 4 months.

## Limitations

* Unmeasured confounding
* Small sample size
* Timing of measurements

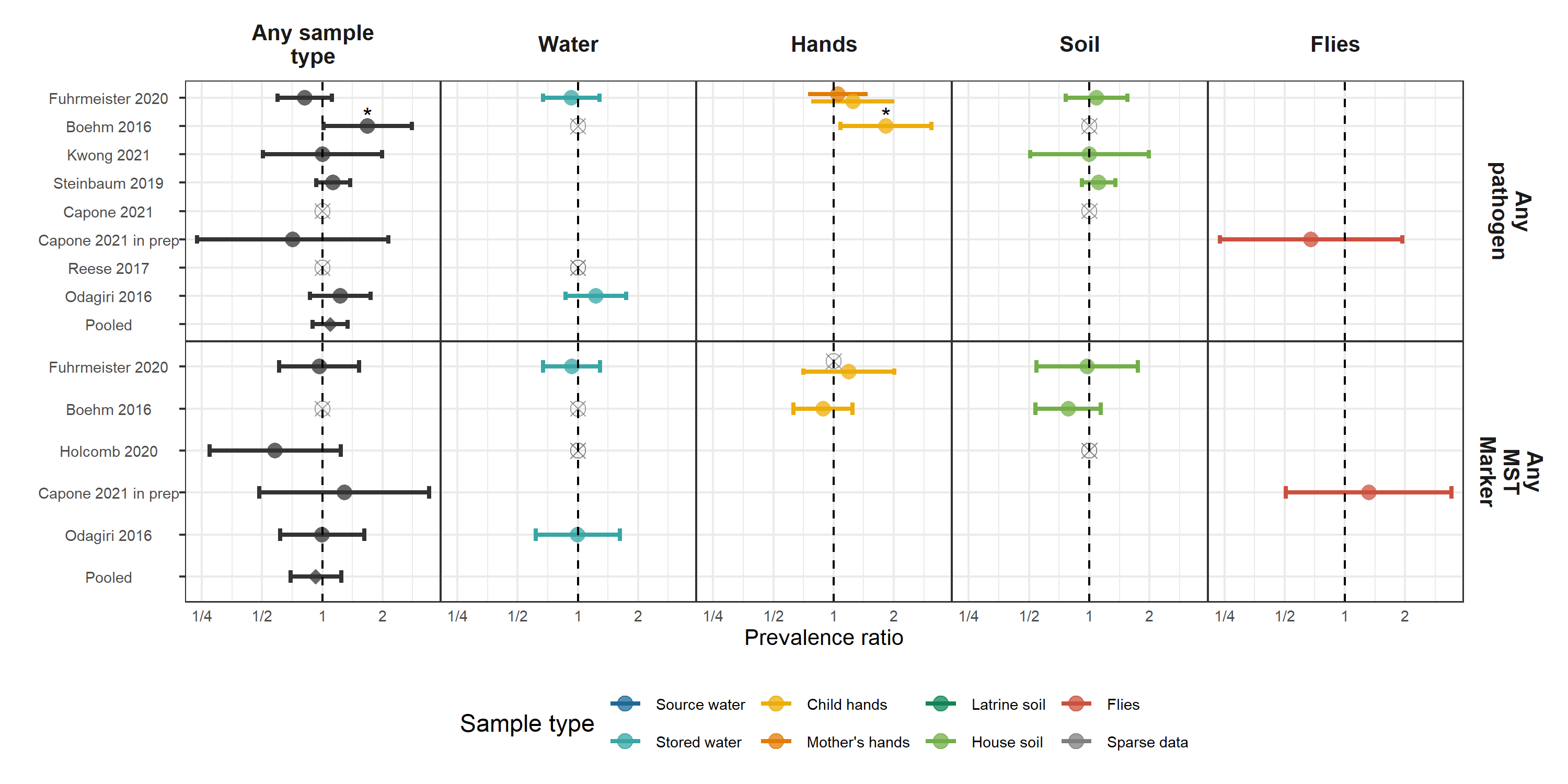
-give mean of number of confounders adjusted for

plot out difference by number of confounders.

-we didn’t see estimates move that much with many variables to adjust for.

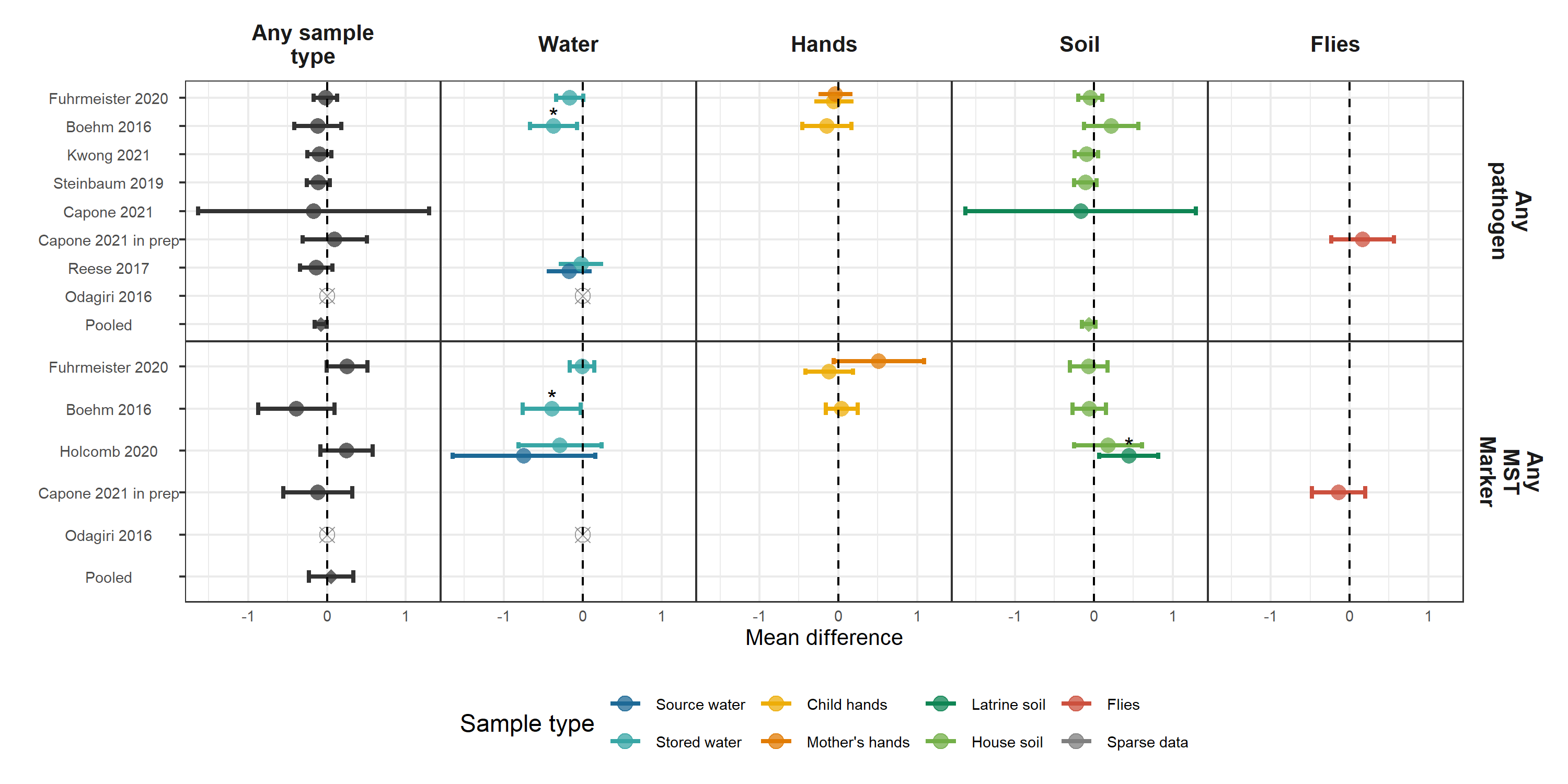
# Primary figures

## Adjusted associations between diarrhea and any enteropathogen or MST marker



**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.

**Interpretation:** 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.

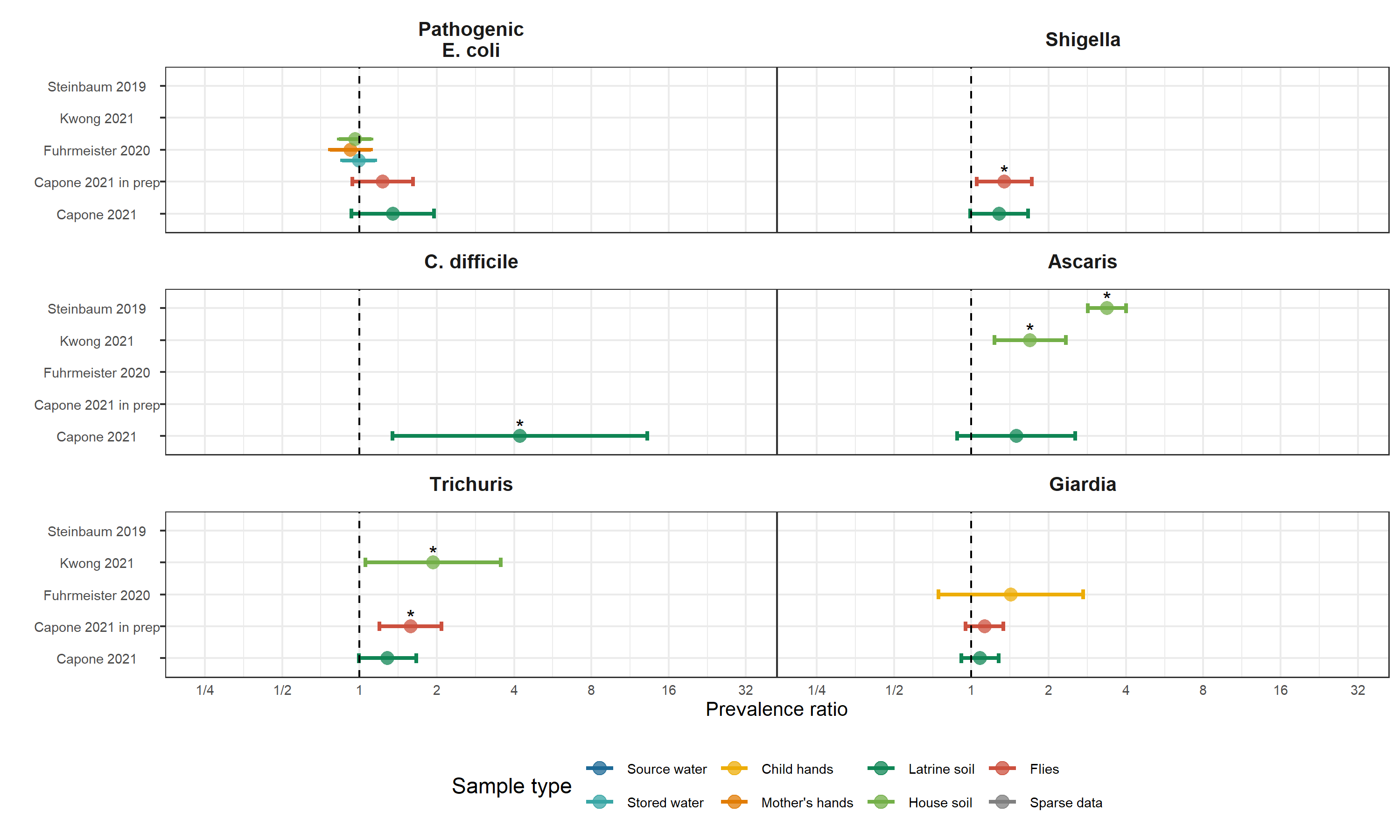


## Adjusted associations between HAZ and any enteropathogen or MST marker

**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. 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.

**Interpretation:** Presence of any enteropathogen (but not any mst marker) in any environmental sample is significantly associated with lower HAZ when pooled across studies (Adjusted mean difference: (95% CI: , )). This is 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. Nevertheless, water samples with any enteropathogen presence were significantly associated with lower mean HAZ in Boehm et al. 2016. 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.

## Adjusted associations between enteropathogen-specific presence in environmental samples and enteropathogen-specific infections

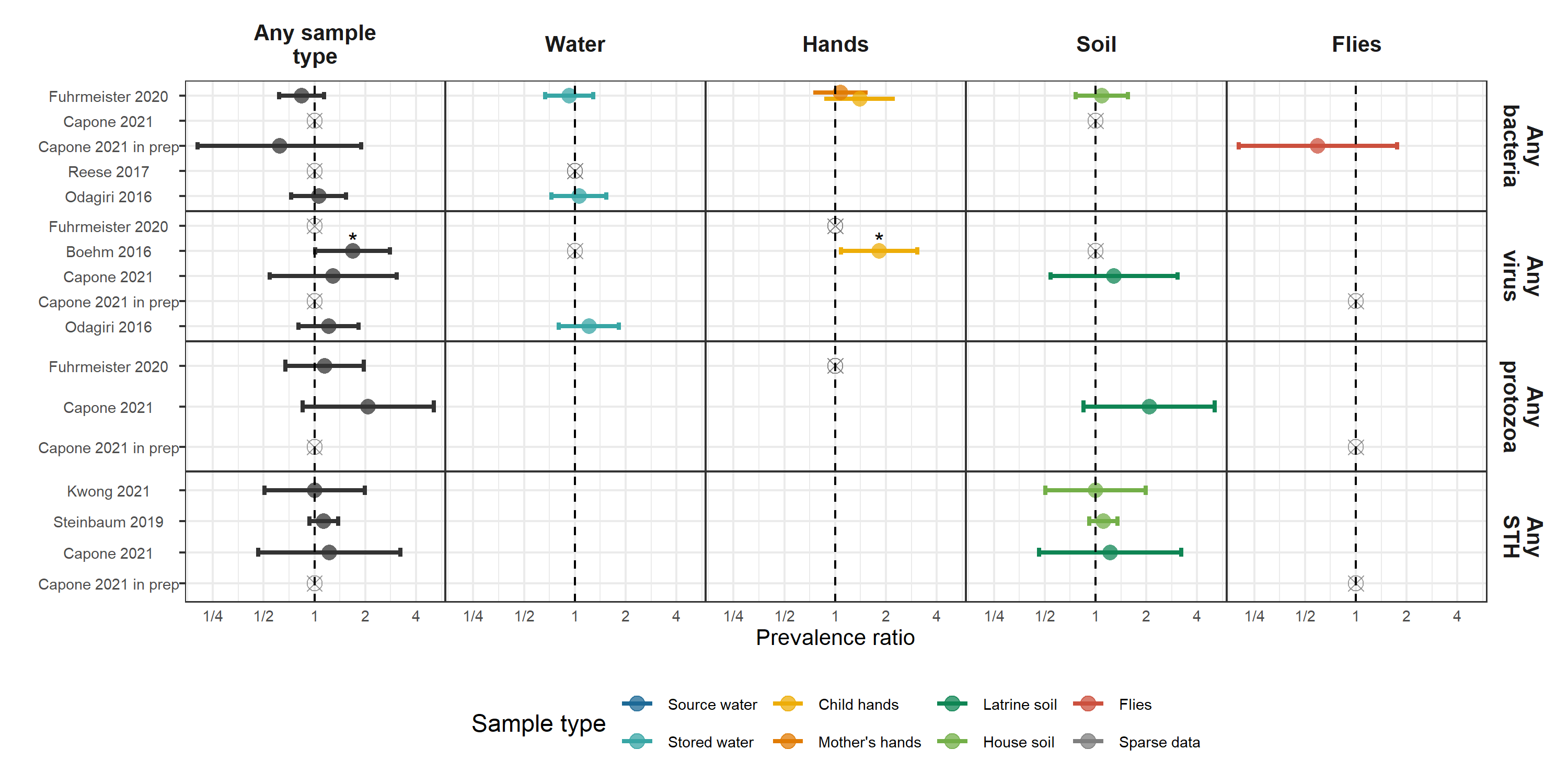


**Figure 3.** Forest plots of associations between specific enteropathogens in environmental samples and child infections with the same enteropathogens. Samples of the same type from different locations (source vs. stored water, flies in kitchen vs. latrine, soil from courtyard vs. latrine) or different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

**Interpretation:** There is 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. *Giardia* and enteropathogenic *E. coli* were two enteropathogens without associations between environmental presence and child infection, but associations were significant or near significant for *Shigella*, *Ascaris*, and *Trichuris* contaminations 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.

# Supplimentary figures

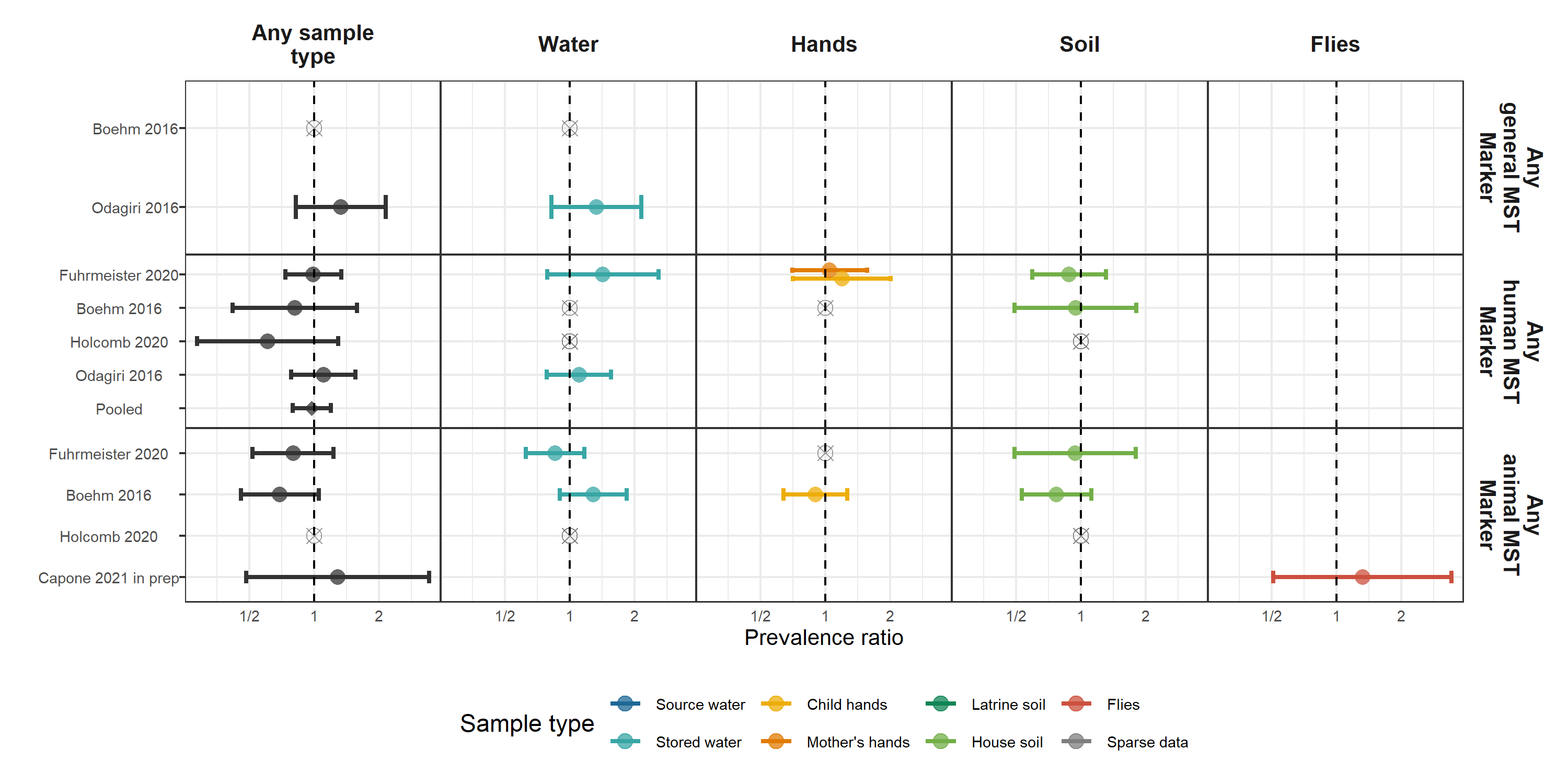
## Adjusted associations between diarrhea and types of enteropathogens



**Figure 4.** Forest plots of associations between child diarrheal disease and the prevalence of any virus, any bacteria, any protozoa and any STH 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., <5 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.

**Interpretation:** Presence of types of enteropathogens in environmental samples were not associated with diarrheal disease, except any viral enteropathogen presence located on child hands in Boehm et al. 2016.

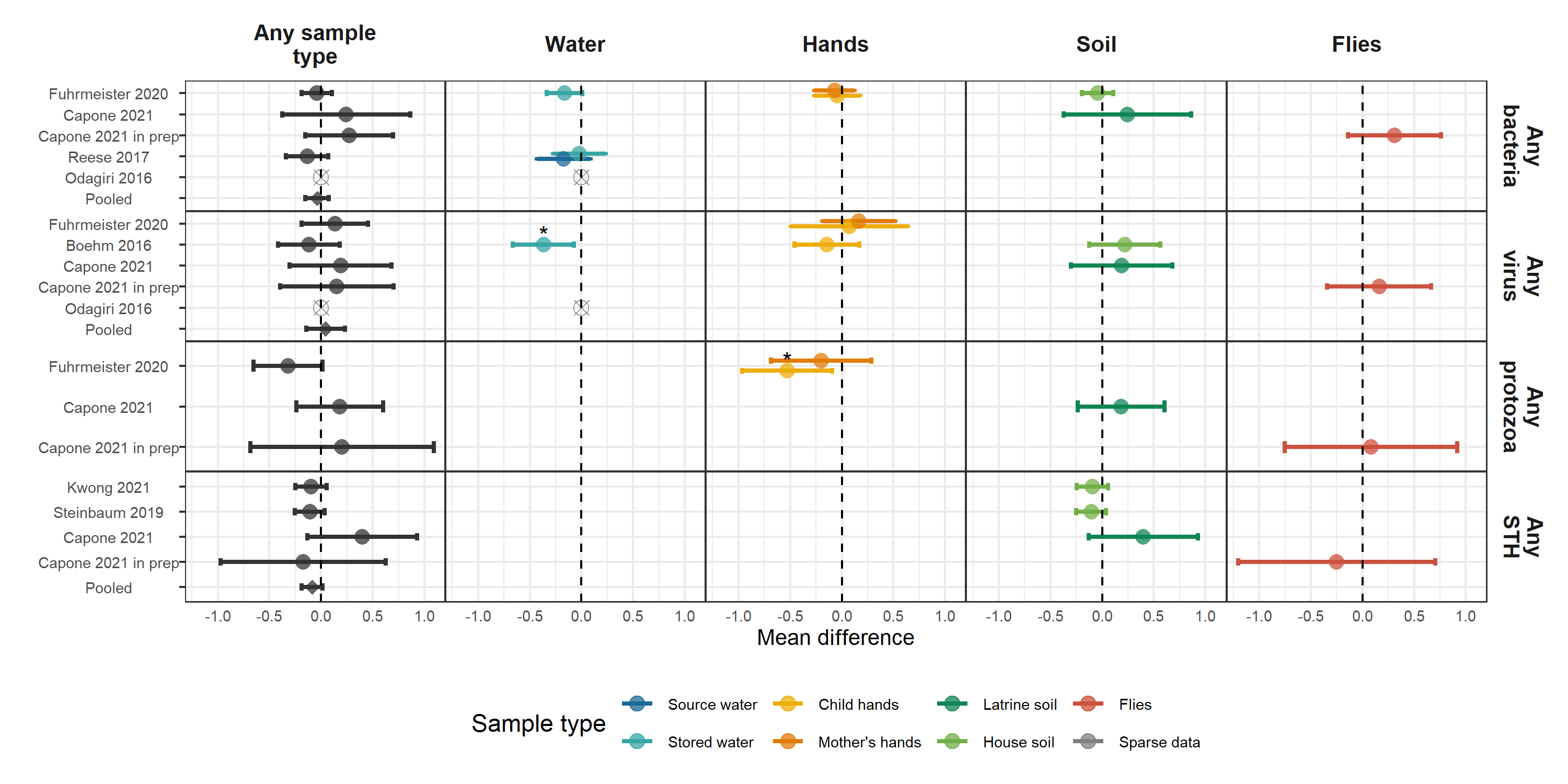
## Adjusted associations between diarrhea and types of MST markers



**Figure 5.** Forest plots of associations between child diarrheal disease and the prevalence of any general, human, or animal MST 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., <5 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.

**Interpretation:** No associations between specific groups of MST markers and child diarrheal disease in any sample type.

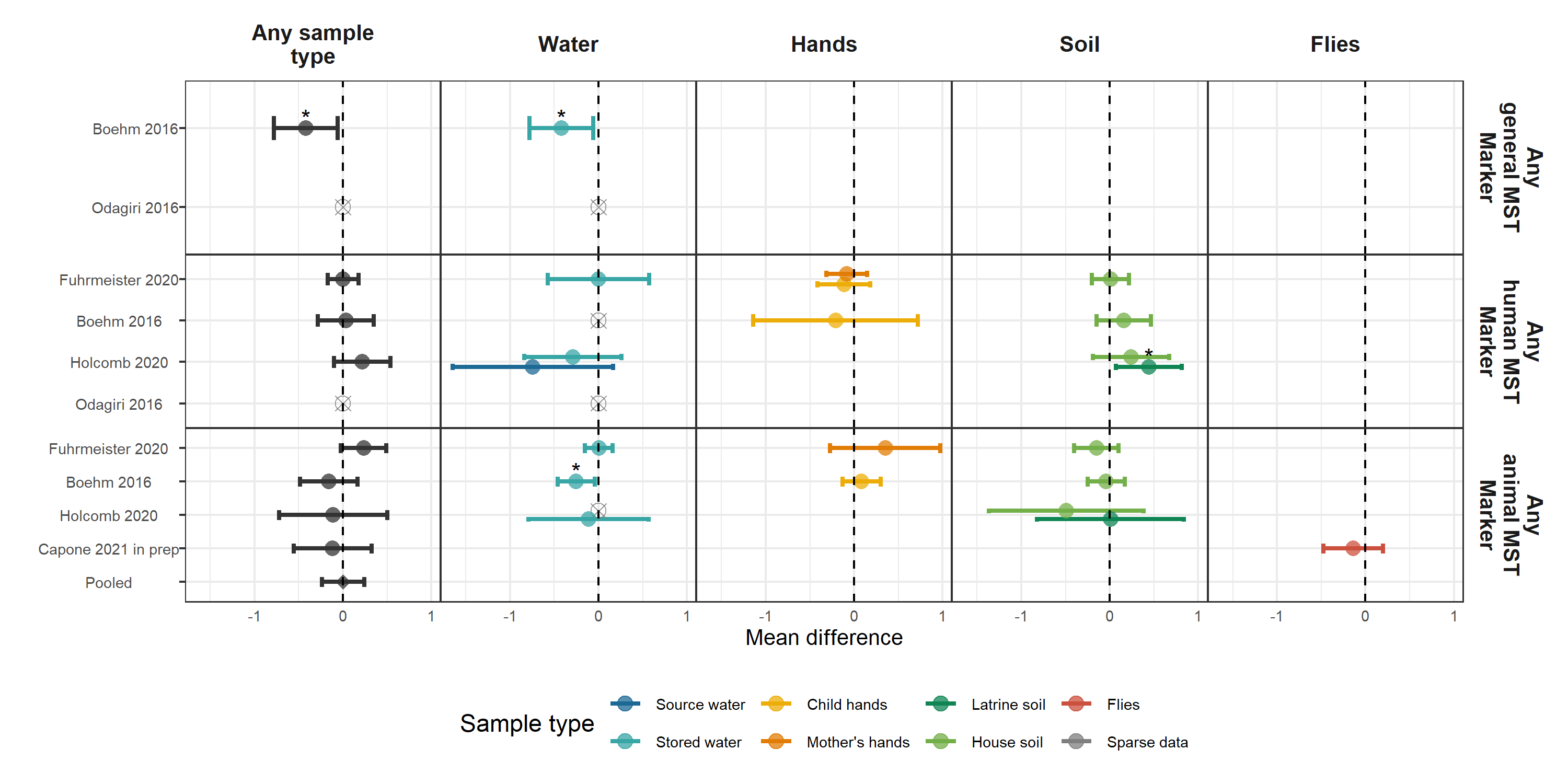
## Adjusted associations between HAZ and types of enteropathogens



**Figure 6.** Forest plots of associations between child HAZ and the prevalence of groups of enteropathogens 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.

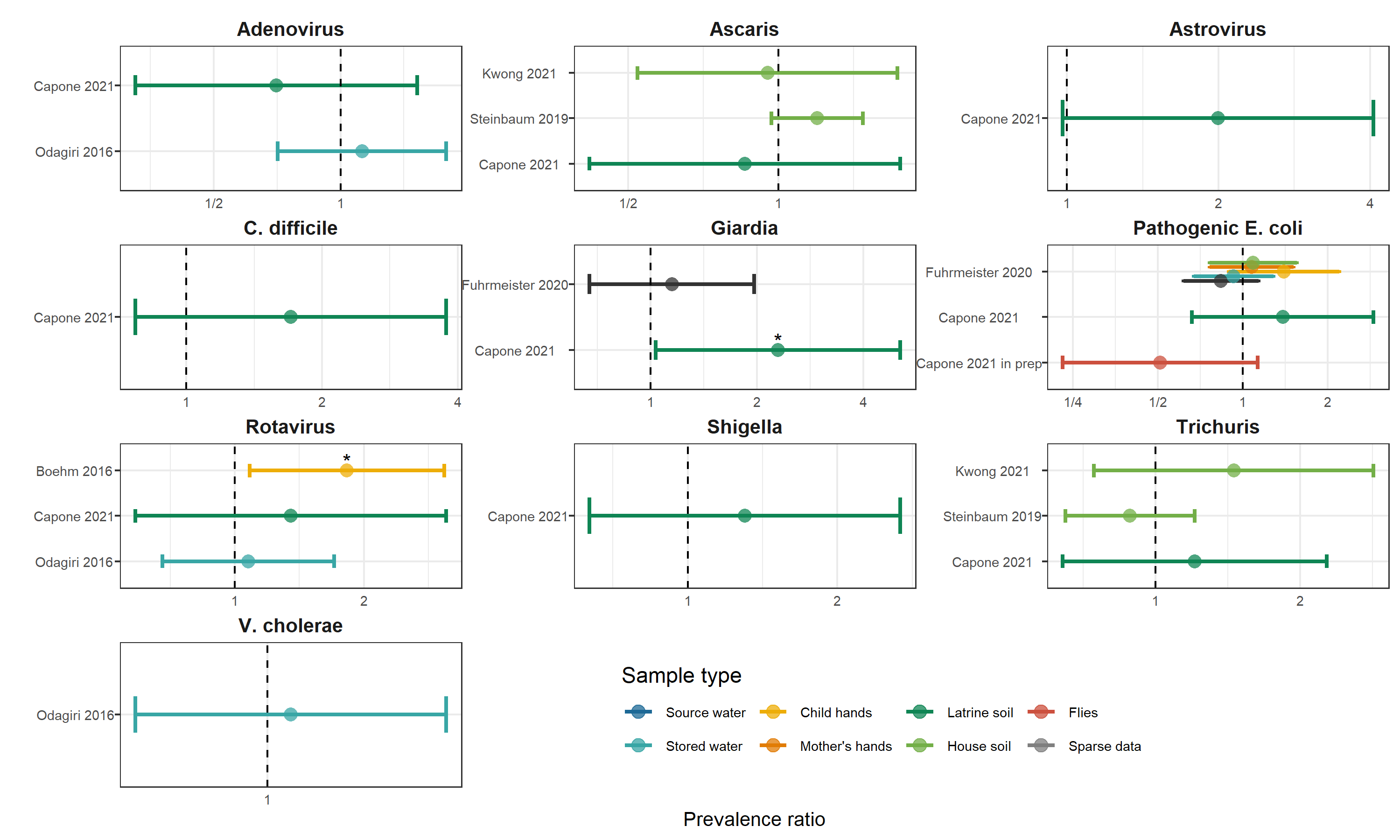
**Interpretation:** When separated out by group of enteropathogen, enteropathogen presence in any environmental sample is no longer significantly associated with lower HAZ when pooled across studies. However, any virus presence in water was significantly associated with lower mean HAZ in Boehm et al. 2016 and any protozoa in water was significantly associated with lower mean HAZ in Furhmeister et al. 2020, both from the WASH Benefits trial.

## Adjusted associations between HAZ and types of MST Markers



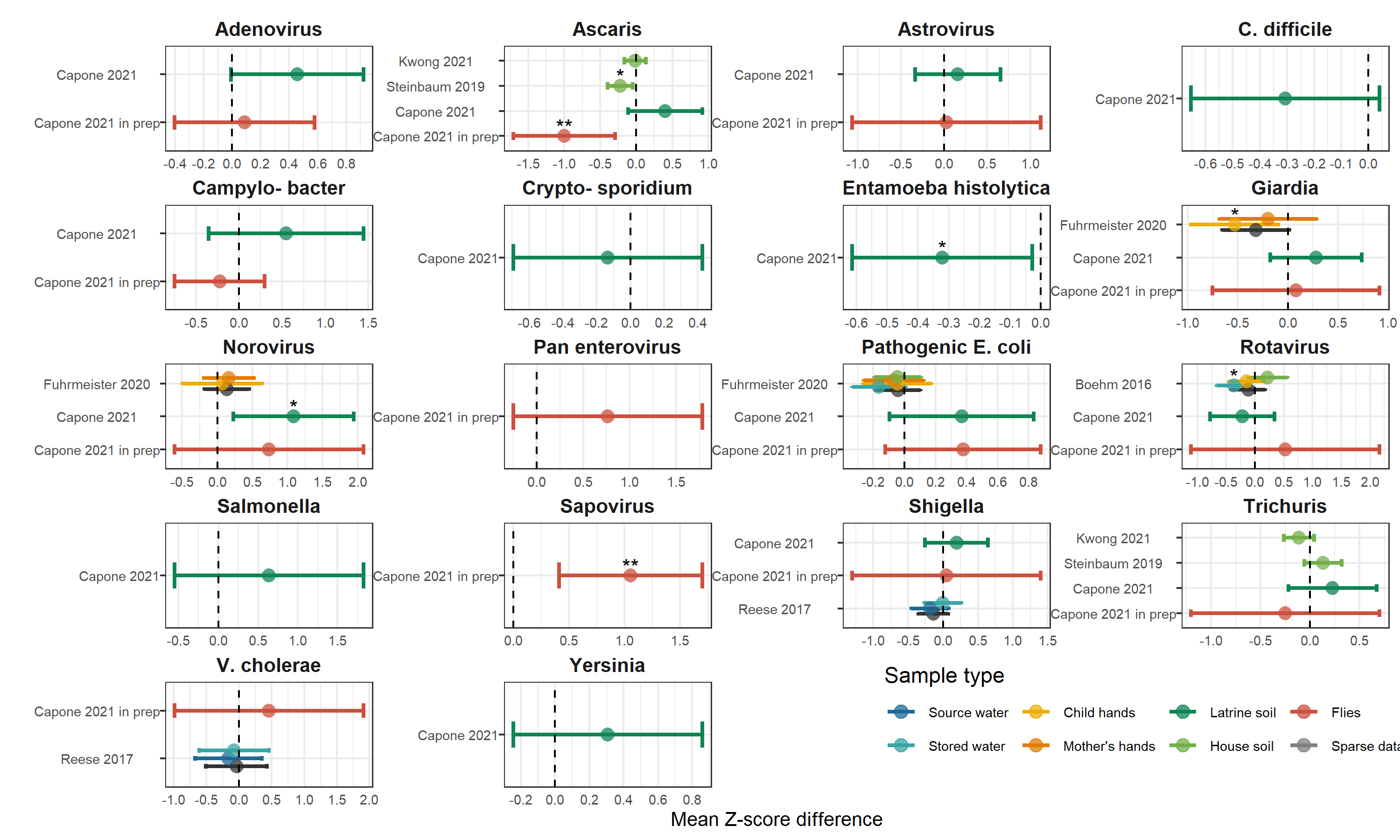
**Figure 7.** Forest plots of associations between child HAZ and the prevalence of groups of MST markers in different types of environmental samples. Pooled estimates are presented when there are four or more study-specific estimates for a specific sample type and target combination and are denoted with 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.

**Interpretation:** There is a general trend of presence of groups of MST markers in water samples being associated with lower mean HAZ.



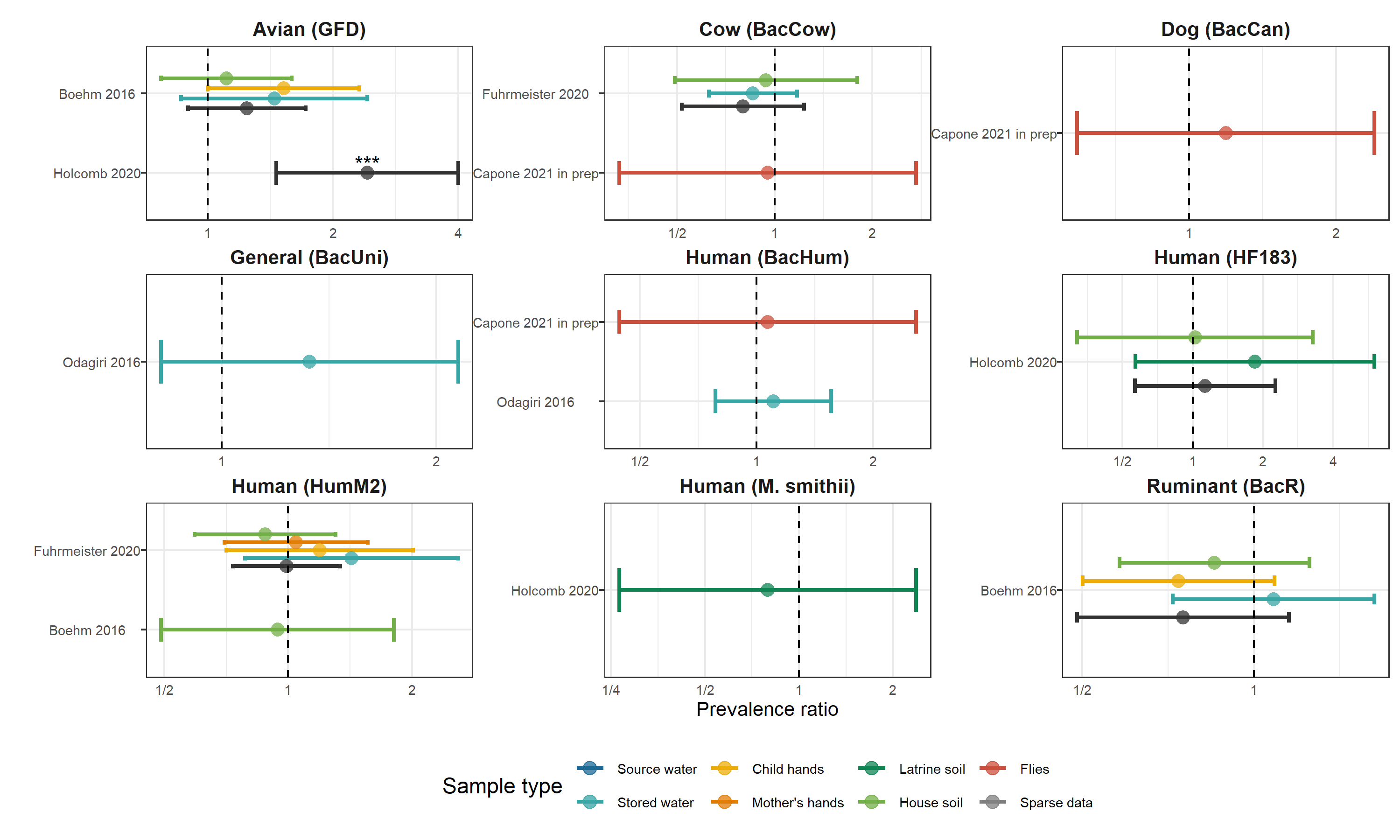
**Figure 8.** Forest plots of associations between child diarrheal disease and the prevalence of specific enteropathogens in different types of environmental samples. Samples of the same type from different locations (source vs. stored water, flies in kitchen vs. latrine, soil from courtyard vs. latrine) or different individuals (child vs. mother’s hands) are plotted separately and colored differently. 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.

**Interpretation:** Most associations are 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.



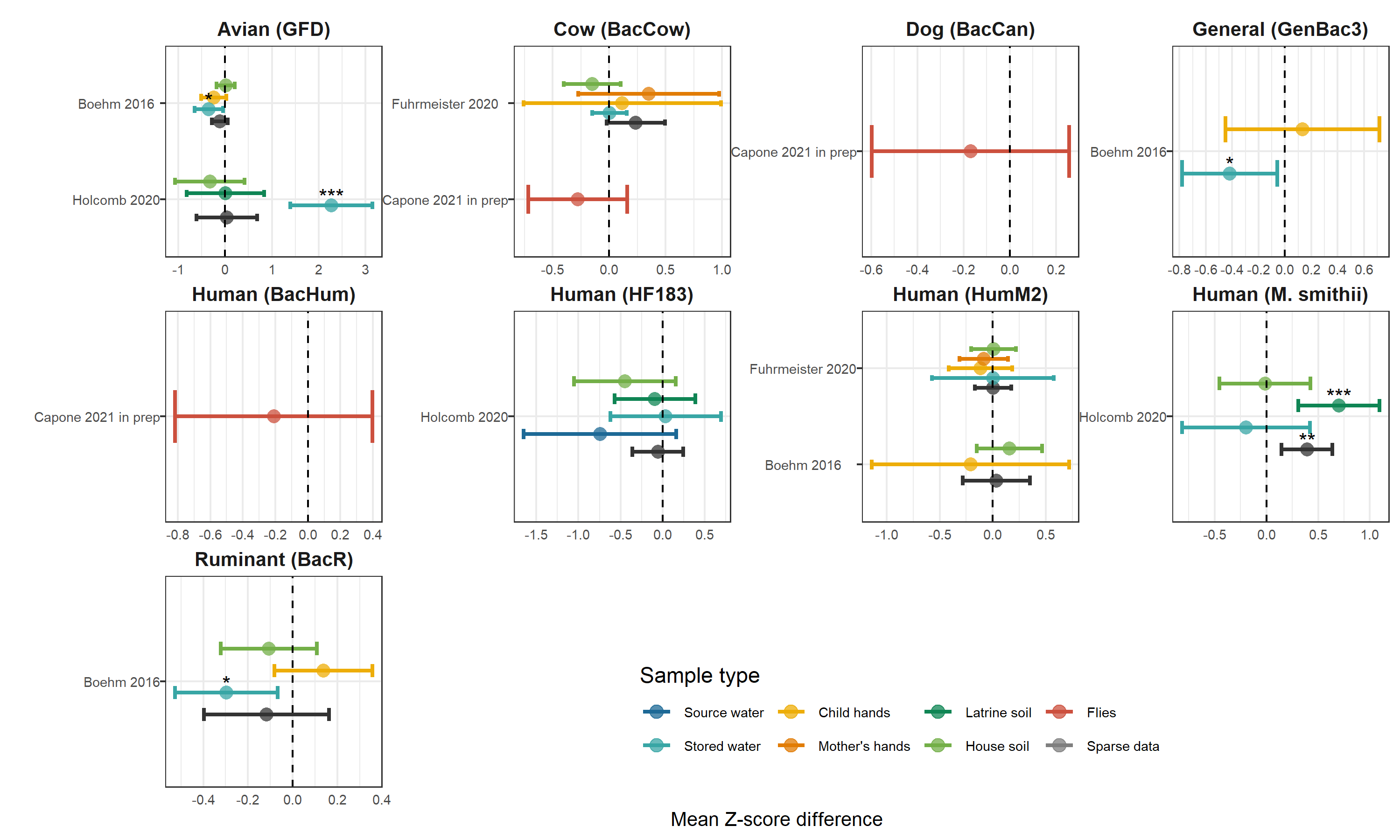
**Figure 9.** Forest plots of associations between child HAZ and the prevalence of specific enteropathogens in different types of environmental samples. Samples of the same type from different locations (source vs. stored water, flies in kitchen vs. latrine, soil from courtyard vs. latrine) or different individuals (child vs. mother’s hands) are plotted separately and colored differently. 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.

**Interpretation:** There are inconsistent associations between specific enteropathogens and child HAZ scores, with most estimates having null effects. Of the statistically significant associations, half were associated with increased linear growth and half were associated with decreased linear growth.



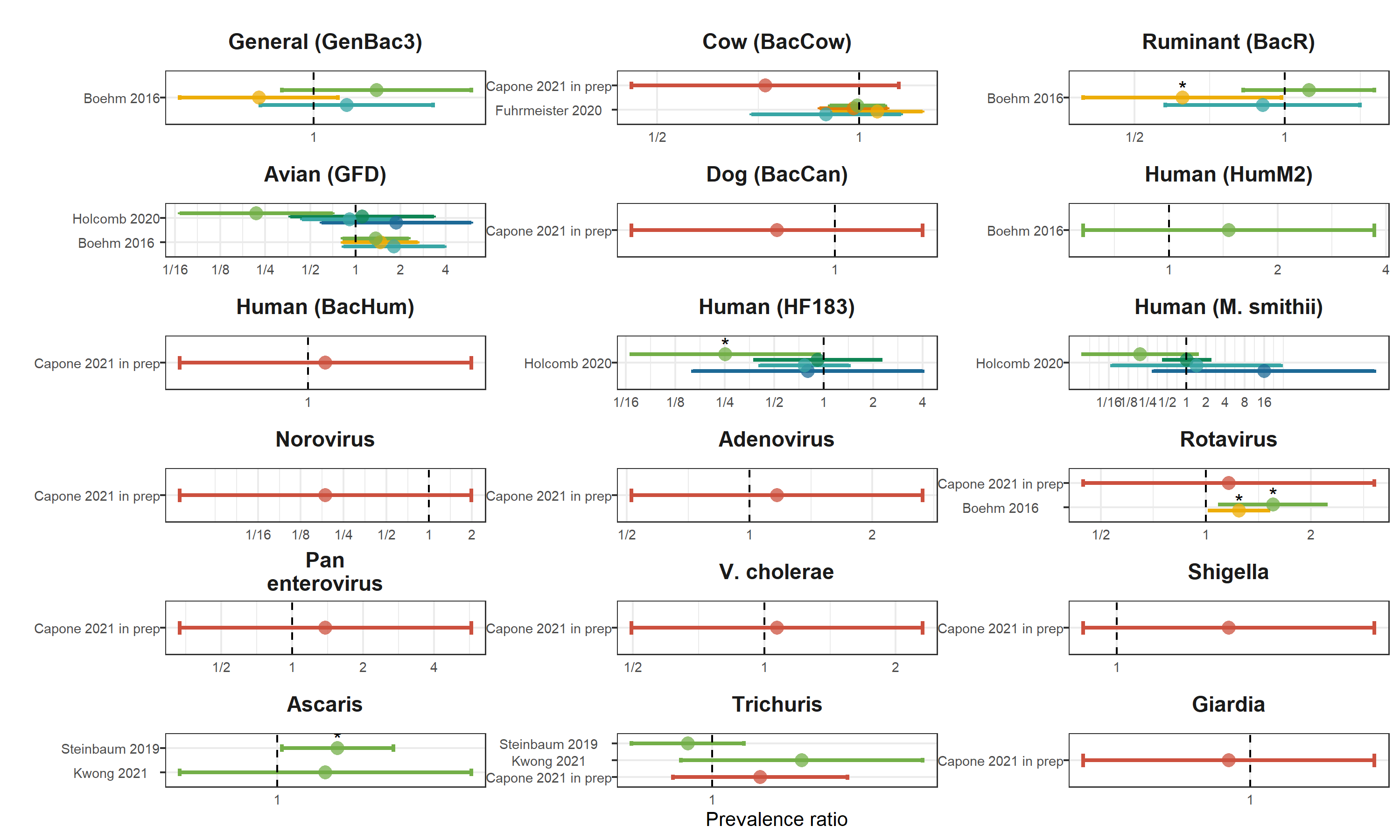
**Figure 10.** Forest plots of associations between child diarrheal disease and the prevalence of specific MST markers in different types of environmental samples. Samples of the same type from different locations (source vs. stored water, flies in kitchen vs. latrine, soil from courtyard vs. latrine) or different individuals (child vs. mother’s hands) are plotted separately and colored differently. 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.

**Interpretation:** 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 by positive association in all three sample types from Boehm et al 2016.



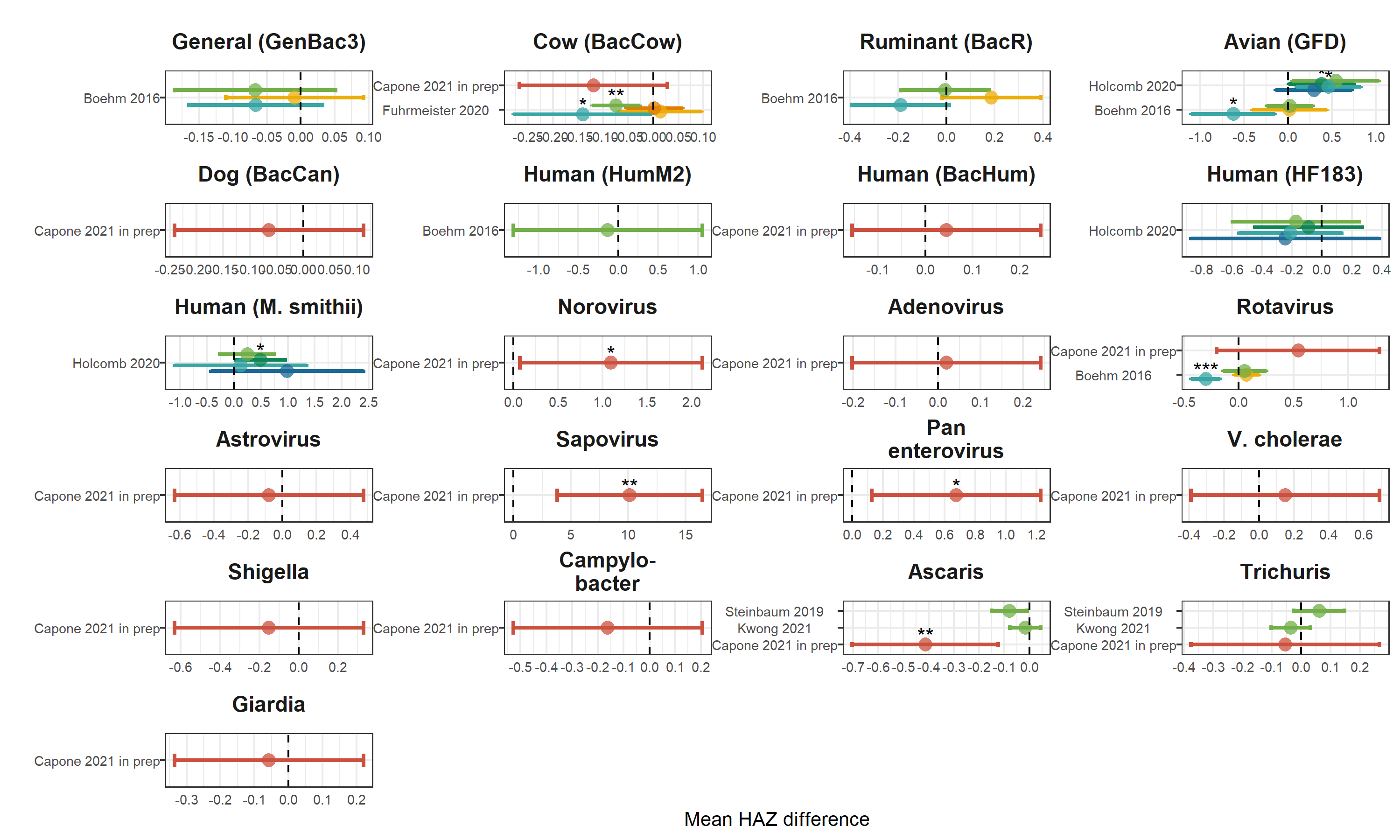
**Figure 11.** Forest plots of associations between child HAZ and the prevalence of specific MST markers in different types of environmental samples. Samples of the same type from different locations (source vs. stored water, flies in kitchen vs. latrine, soil from courtyard vs. latrine) or different individuals (child vs. mother’s hands) are plotted separately and colored differently. 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.

**Interpretation:** There are inconsistent associations between the presence of specific MST markers and child HAZ scores, with most estimates having null effects. 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.



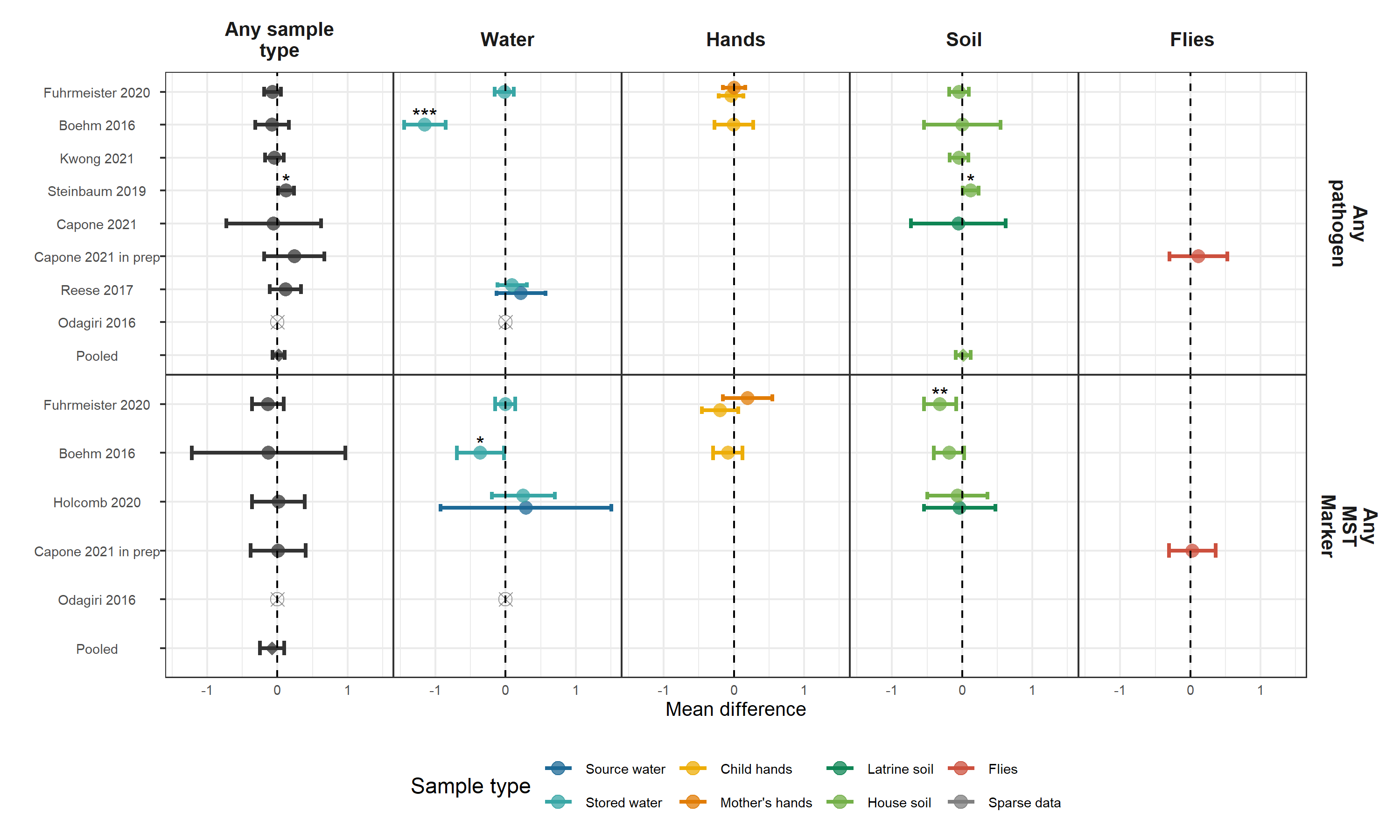
**Figure 12.** Forest plots of associations between child diarrheal disease and the abundance of specific enteropathogens and MST markers in different types of environmental samples. Samples of the same type from different locations (source vs. stored water, flies in kitchen vs. latrine, soil from courtyard vs. latrine) or different individuals (child vs. mother’s hands) are plotted separately and colored differently. 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.

**Interpretation:** 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.



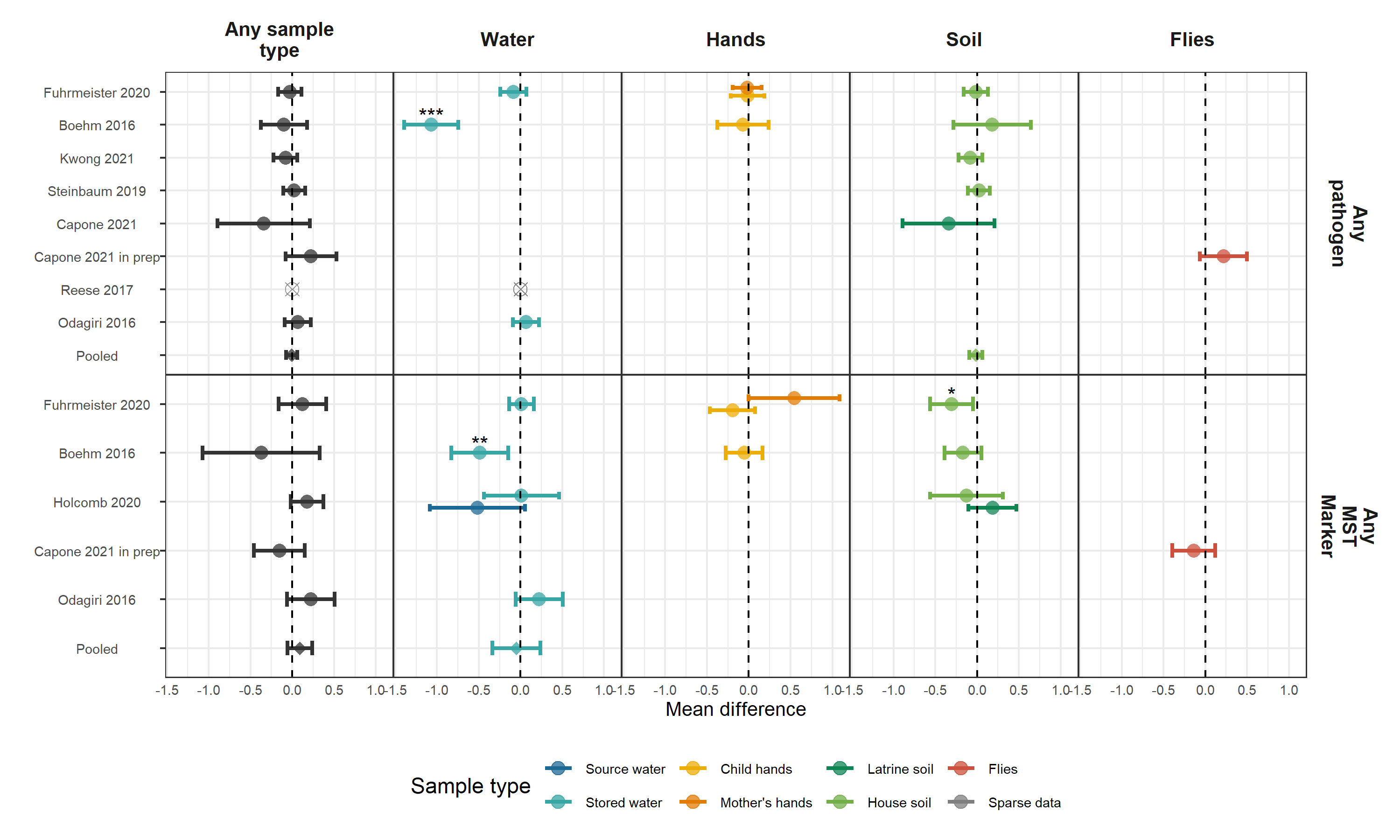
**Figure 13.** Forest plots of associations between child height-for-age Z-scores and the log-10 transformed abundance of specific enteropathogens and MST markers in different types of environmental samples. Samples of the same type from different locations (source vs. stored water, flies in kitchen vs. latrine, soil from courtyard vs. latrine) or different individuals (child vs. mother’s hands) are plotted separately and colored differently. 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.

**Interpretation:** 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.



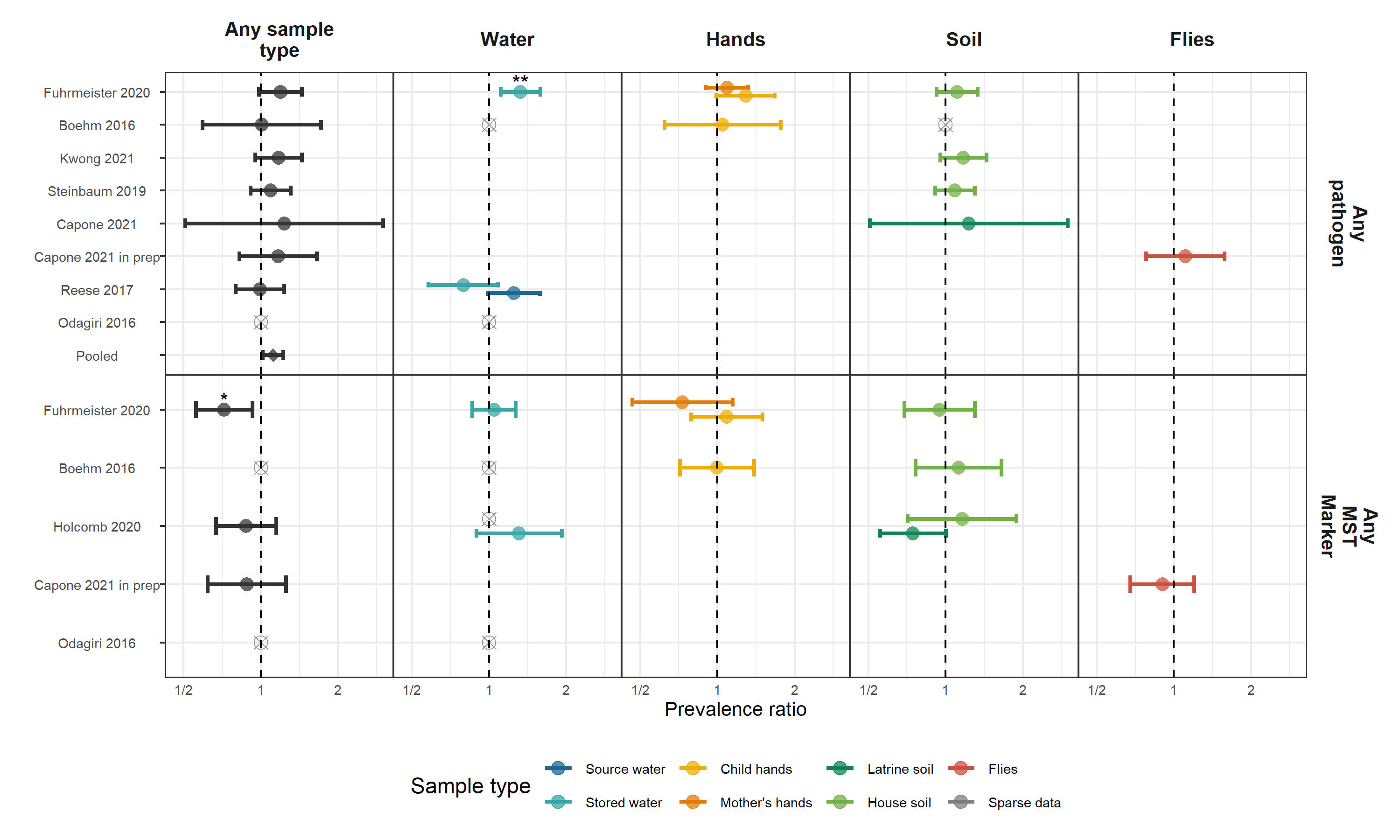
**Figure 14.** Forest plots of associations between child WHZ and the prevalence of any enteropathogen or any MST markers in different types of environmental samples. Pooled estimates are presented when there are four or more study-specific estimates for a specific sample type and target combination 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.

**Interpretation:** There are inconsistent associations between the abundance of any enteropathogen or any MST marker and child WHZ scores, with most estimates having null effects, and with significant effects occurring in both harmful and protective directions.



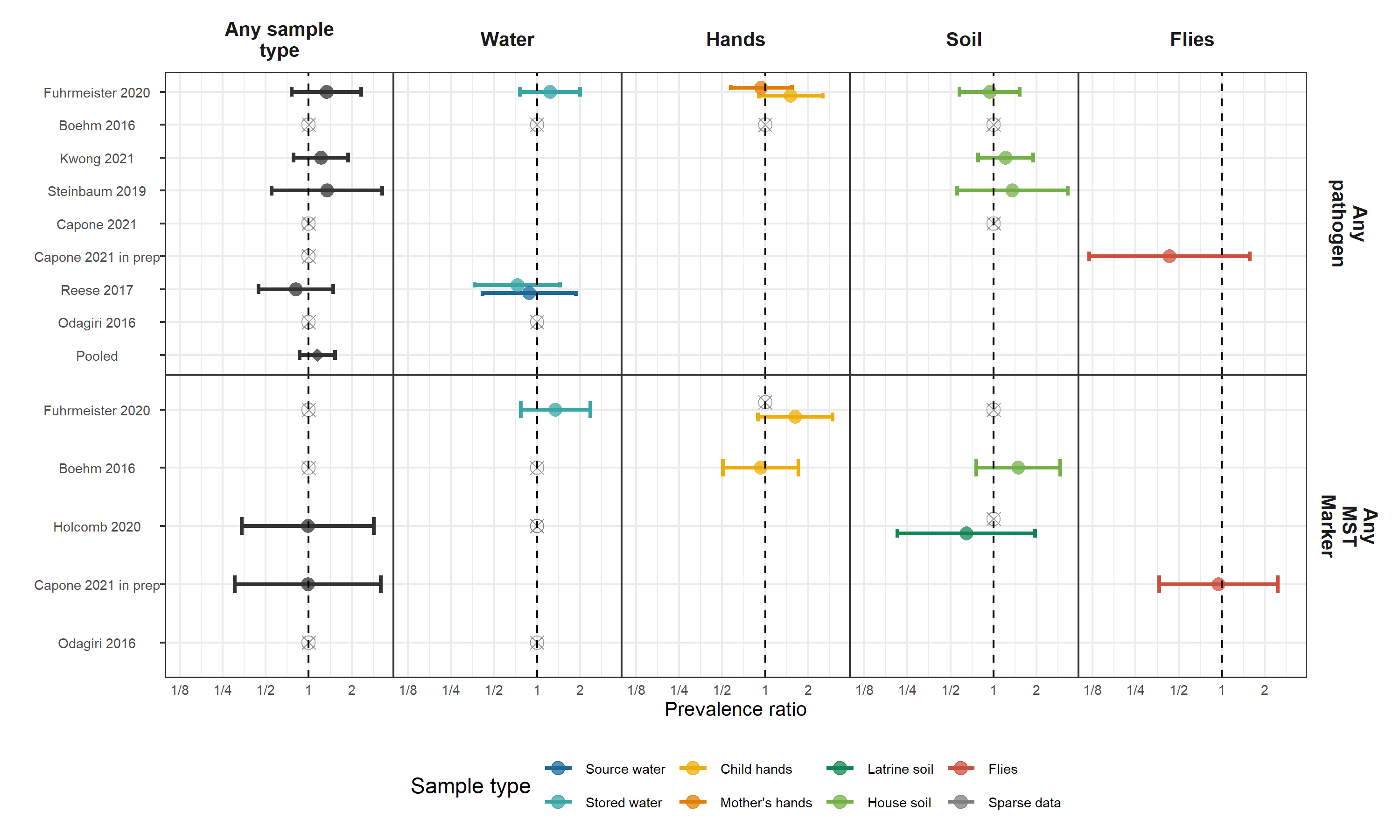
**Figure 15.** Forest plots of associations between child WAZ and the prevalence of any enteropathogen or any MST markers in different types of environmental samples. Pooled estimates are presented when there are four or more study-specific estimates for a specific sample type and target combination 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.

**Interpretation:** Most associations between the abundance of any enteropathogen or any MST marker and child WHZ scores were null effects, but any MST marker in soil in Furhmeister et al. and any enteropathogen or any MST marker in stored drinking water in Boehm et al. 2016 were associated with children having a significantly lower WAZ.



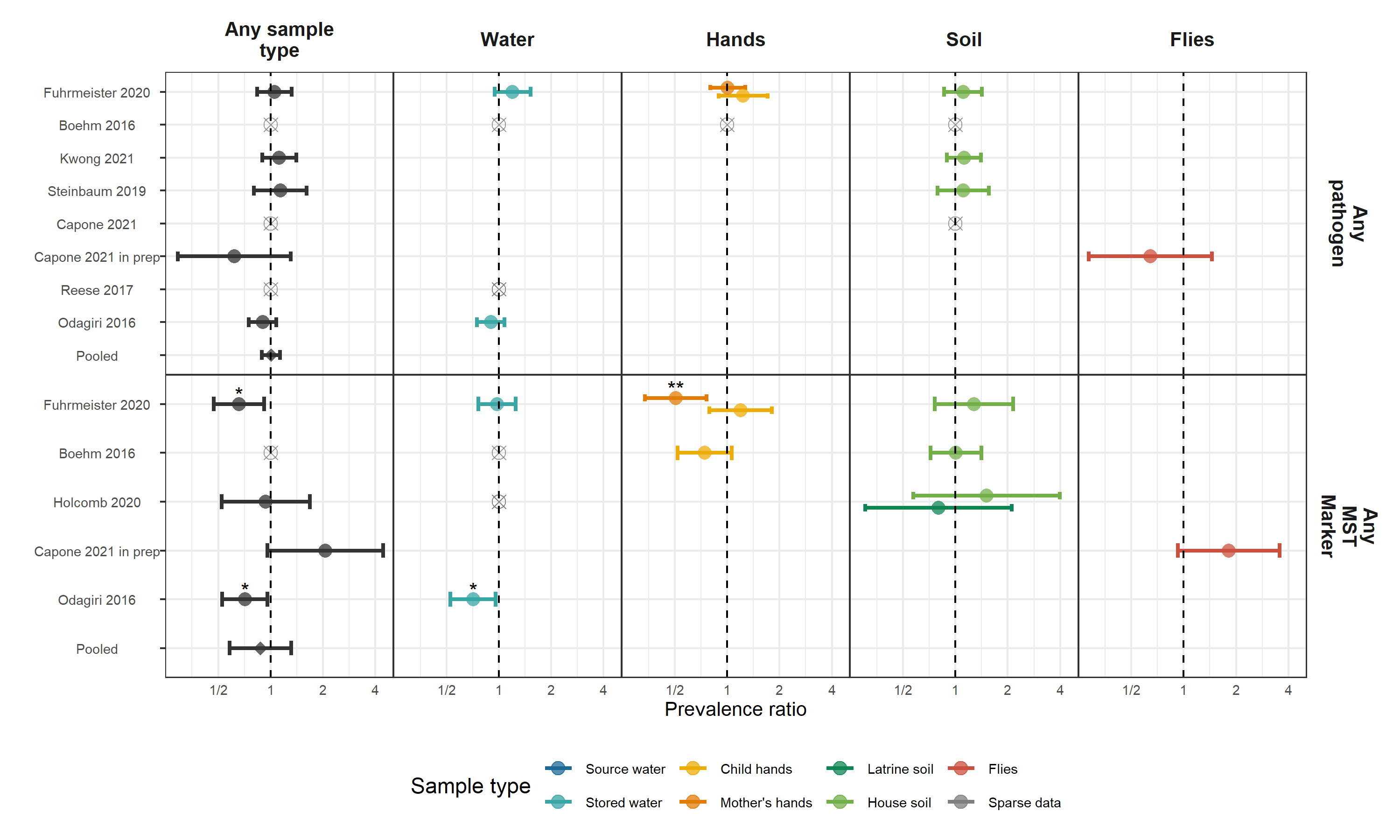
**Figure 16.** Forest plots of associations between child stunting (HAZ < -2) and the prevalence of any enteropathogen or any MST markers in different types of environmental samples. Pooled estimates are presented when there are four or more study-specific estimates for a specific sample type and target 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.

**Interpretation:** While most study- and sample-specific estimates were not significant, any enteropathogen presence in any sample was associated with a significant increased risk of stunting when pooled across studies (Relative risk: (95% CI: , )). MST marker presence in samples was not associated with stunting, except in any sample in Fuhrmeister et al. 2020 where presence was unexpectadly associated with a reduction in stunting.



**Figure 17.** Forest plots of associations between child wasting (WHZ < -2) 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.

**Interpretation:** enteropathogen and MST marker presence was not associated with wasting.

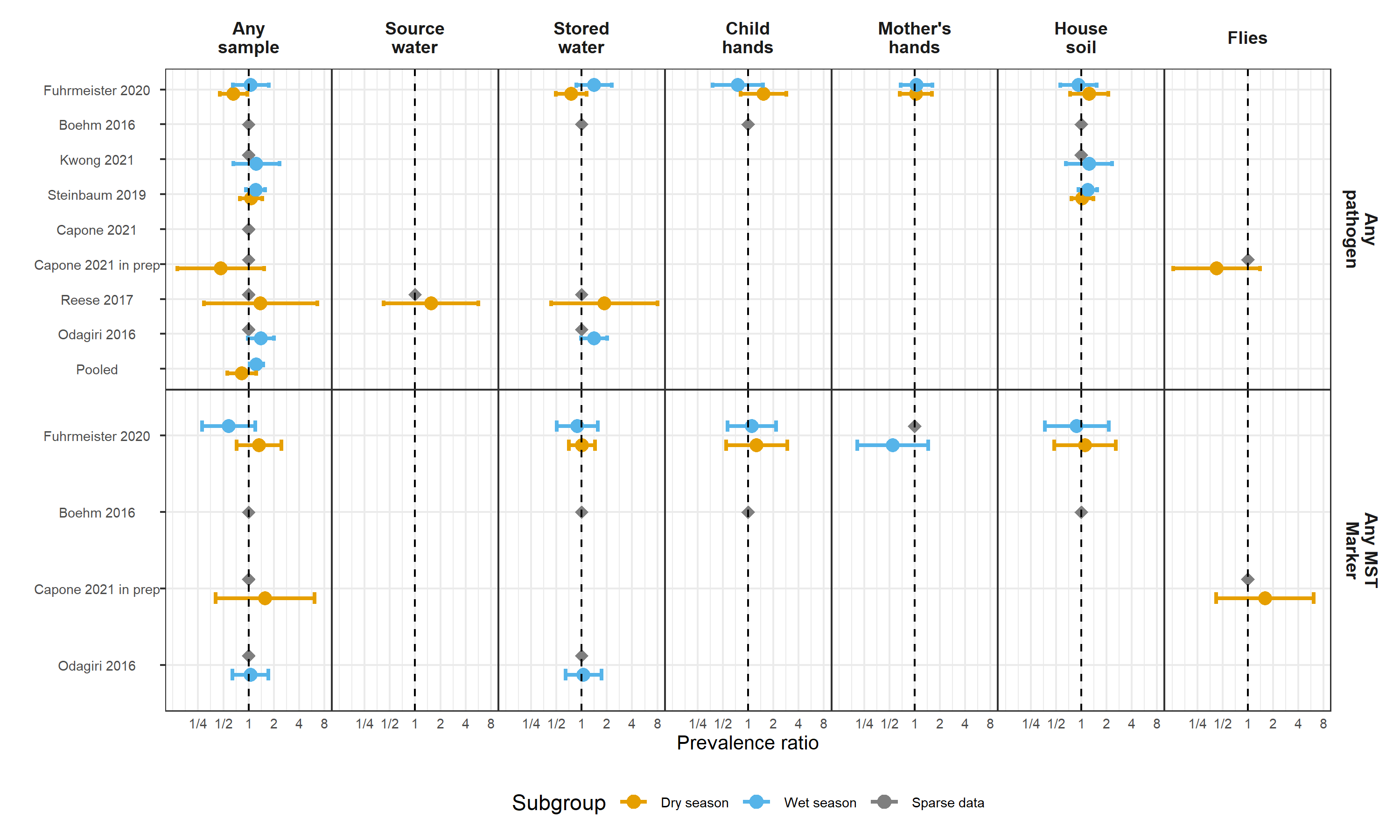


**Figure 18.** Forest plots of associations between child underweight (WAZ < -2) and the prevalence of any enteropathogen or any MST markers in different types of environmental samples. Pooled estimates are presented when there are four or more study-specific estimates for a specific sample type and 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.

**Interpretation:** enteropathogen presence was not associated with child underweight status, but MST marker presence in samples was unexpectadly associated with with child underweight status in Boehm et al. 2016 and in any sample and in samples from mothers’ hands in Fuhrmeister et al. 2020.

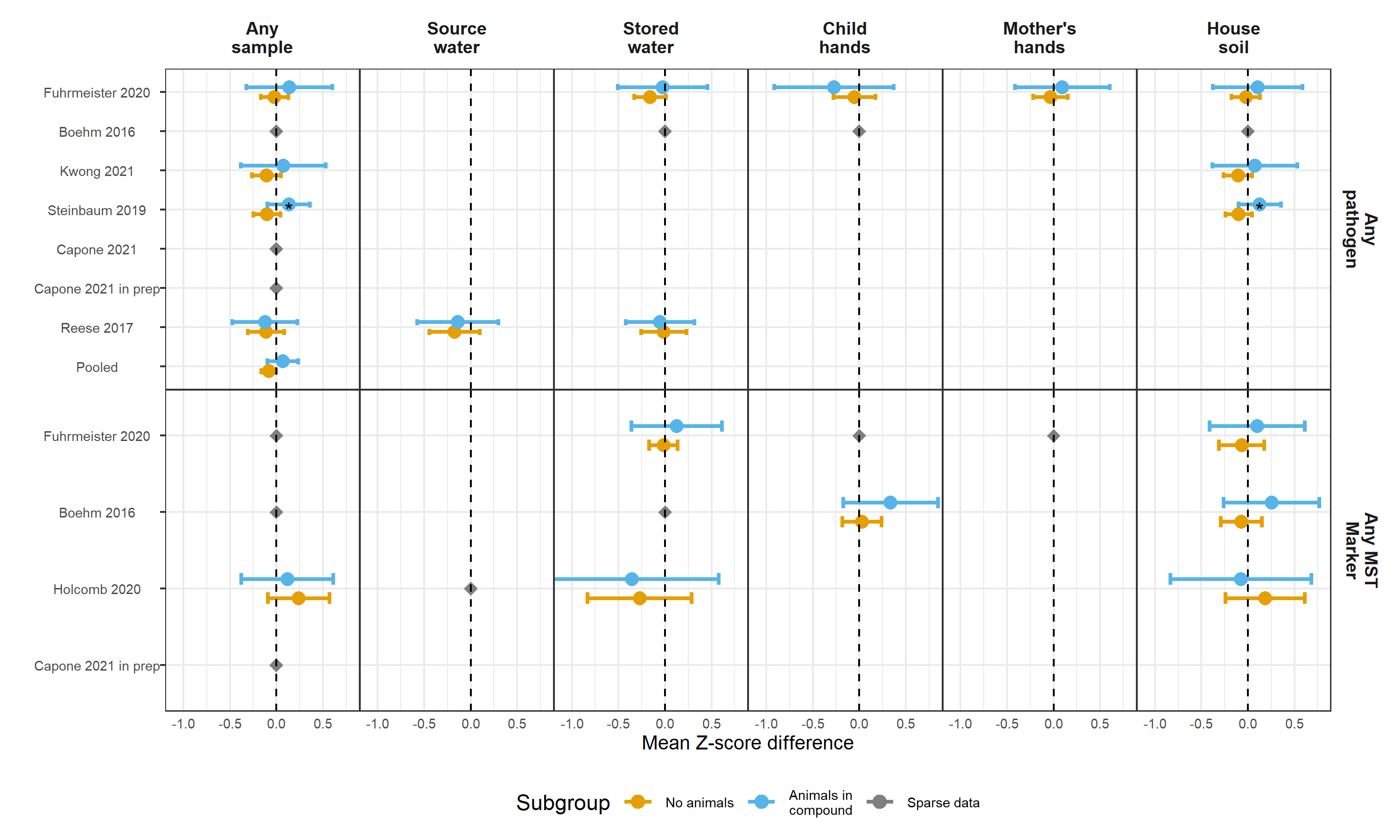
# Sensitivity analysis figures

## Subgroup analysis



**Figure 19.** 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 = “\*\*\*”).

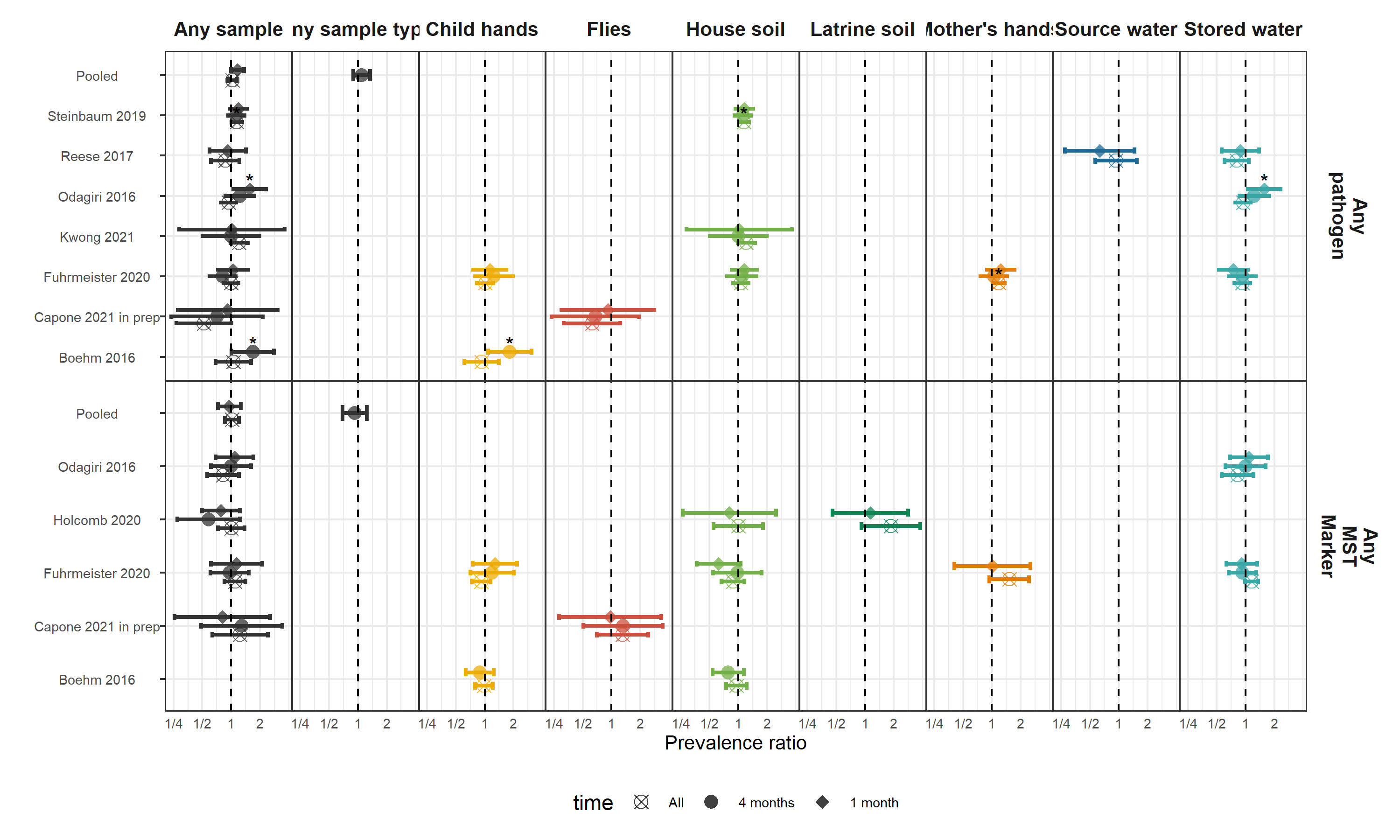
**Interpretation:** Pooled across studies, there was a significant increase in child diarrheal disease risk in compounds with any sample with any enteropathogen detected when the child diarrheal disease occurred during the wet season. There was no association with MST markers in either season.



**Figure 20.** 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 = “\*\*\*”).

**Interpretation:** 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. There was no association with MST markers. Note that there were insufficient cases of diarrheal disease in household without animals (rare) to conduct a similiar subgroup analysis for the diarrheal disease outcome.

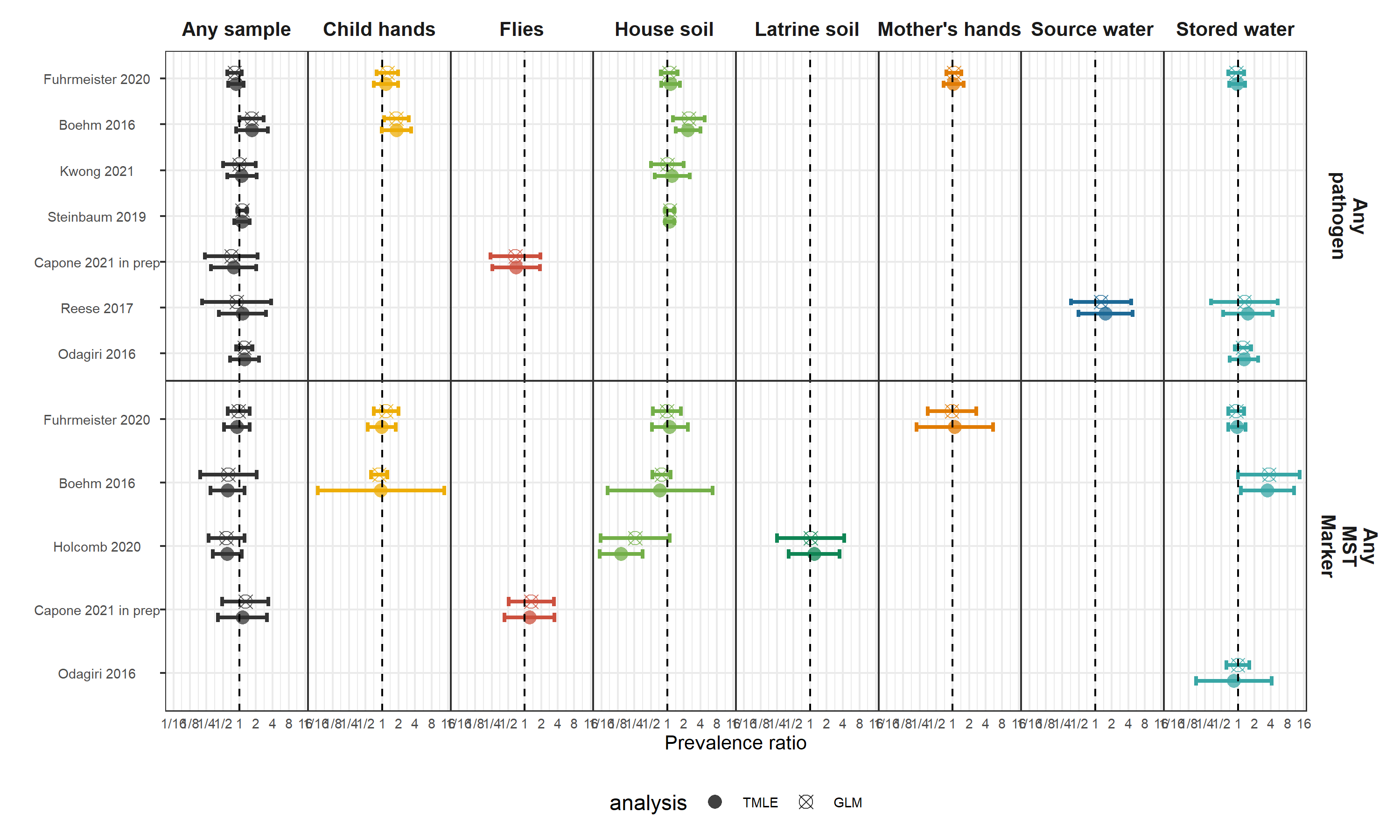
## Timing of diarrheal disease



**Figure 21.** 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.

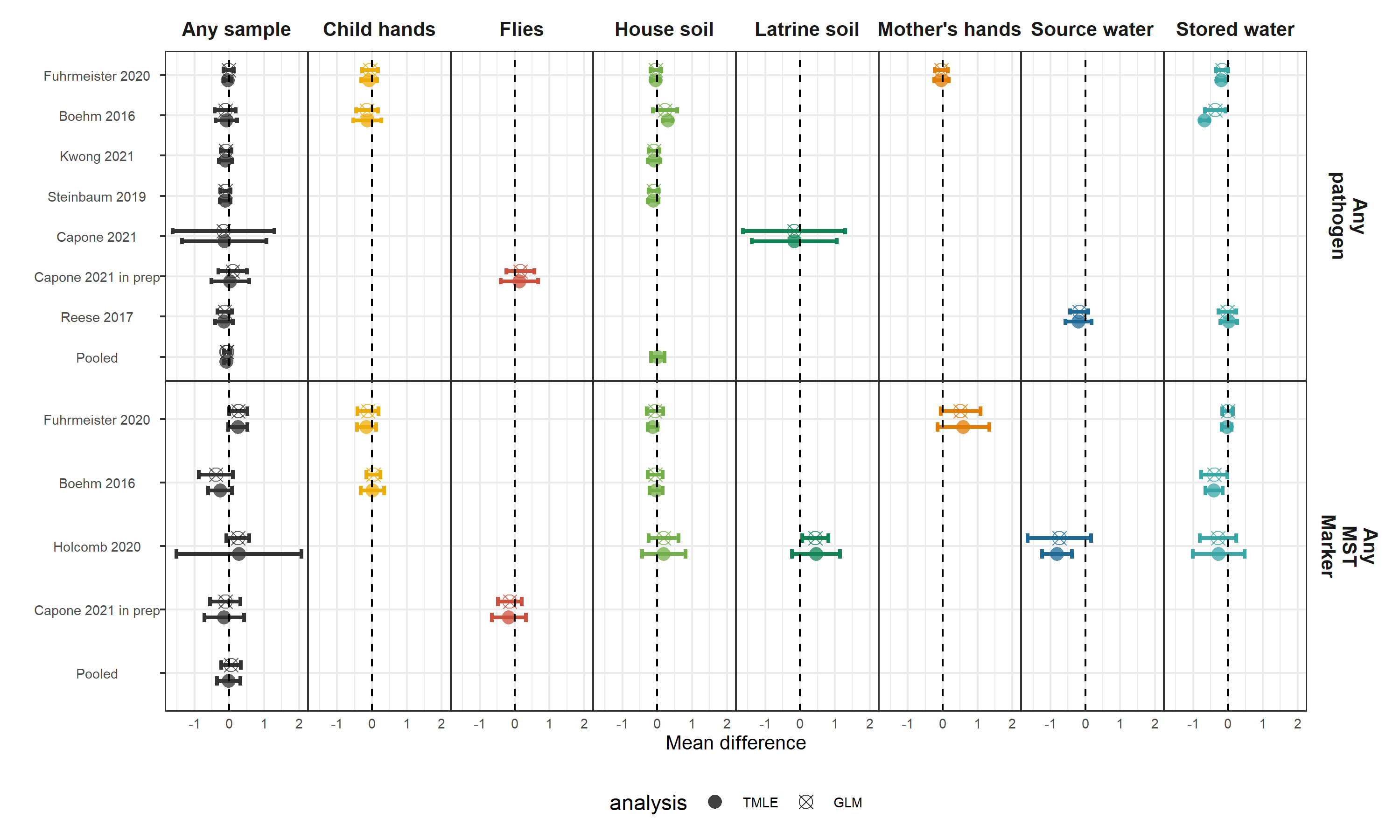
**Interpretation:** The choice to use a maximum of four months as a time window between sampling and diarrheal disease did not obscure major trends in the associations between any enteropathogen or any MST marker in environment and diarrheal disease. Most associations were null regardless of the time window.

## TMLE comparison



**Figure 22.** Comparison between associations estimated with generalized linear models (GLM) and machine-learning based targeted likelihood estimation models (TMLE) for the diarrhea outcome.

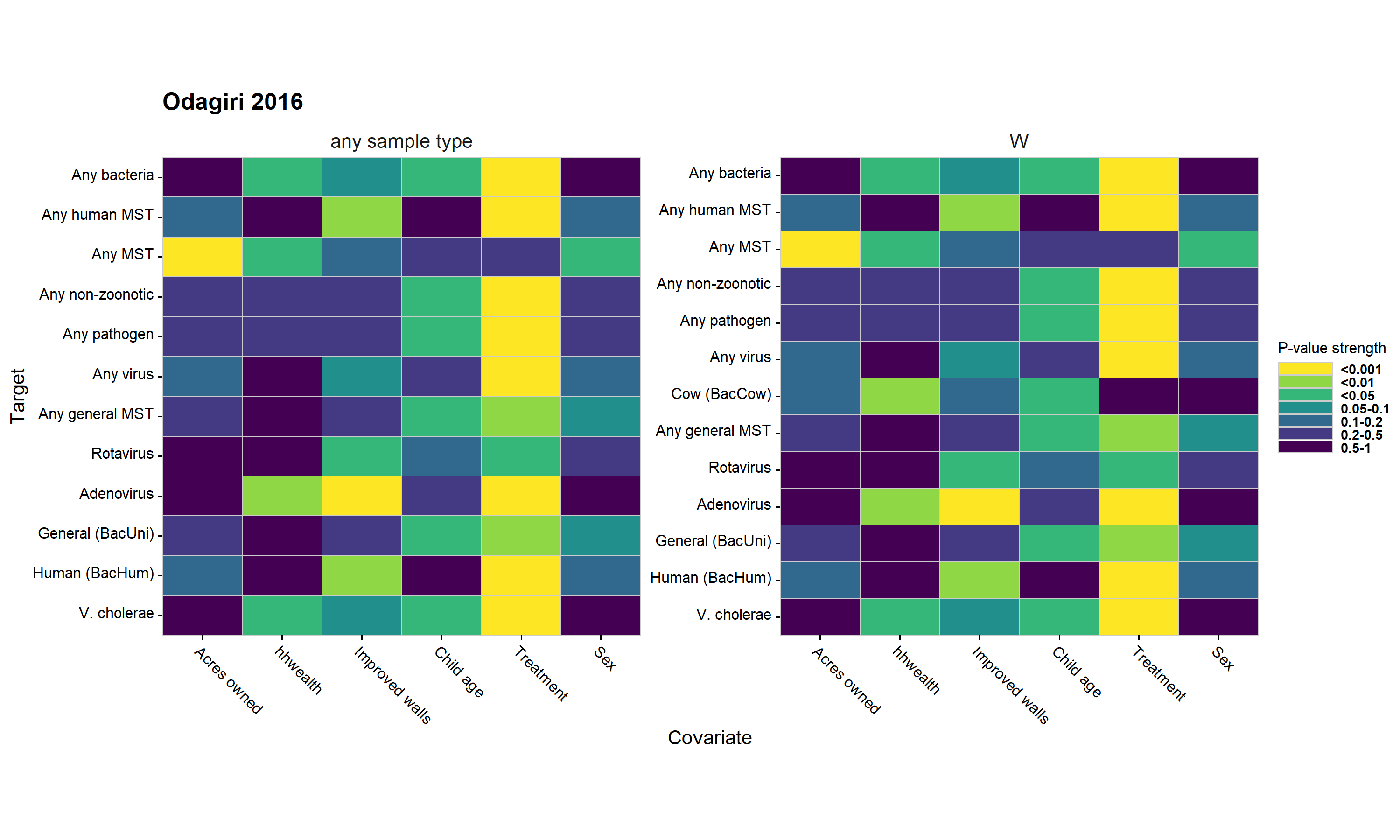
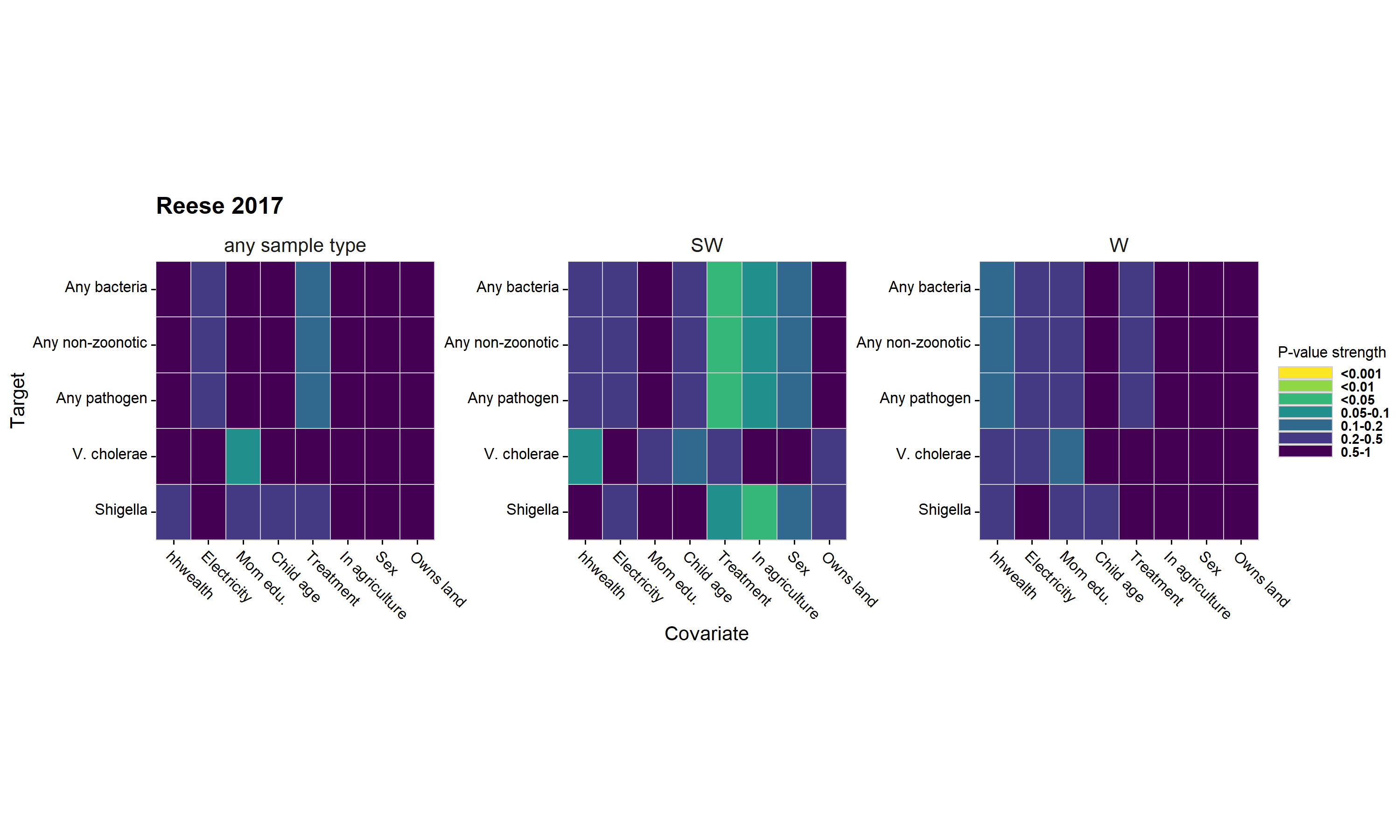
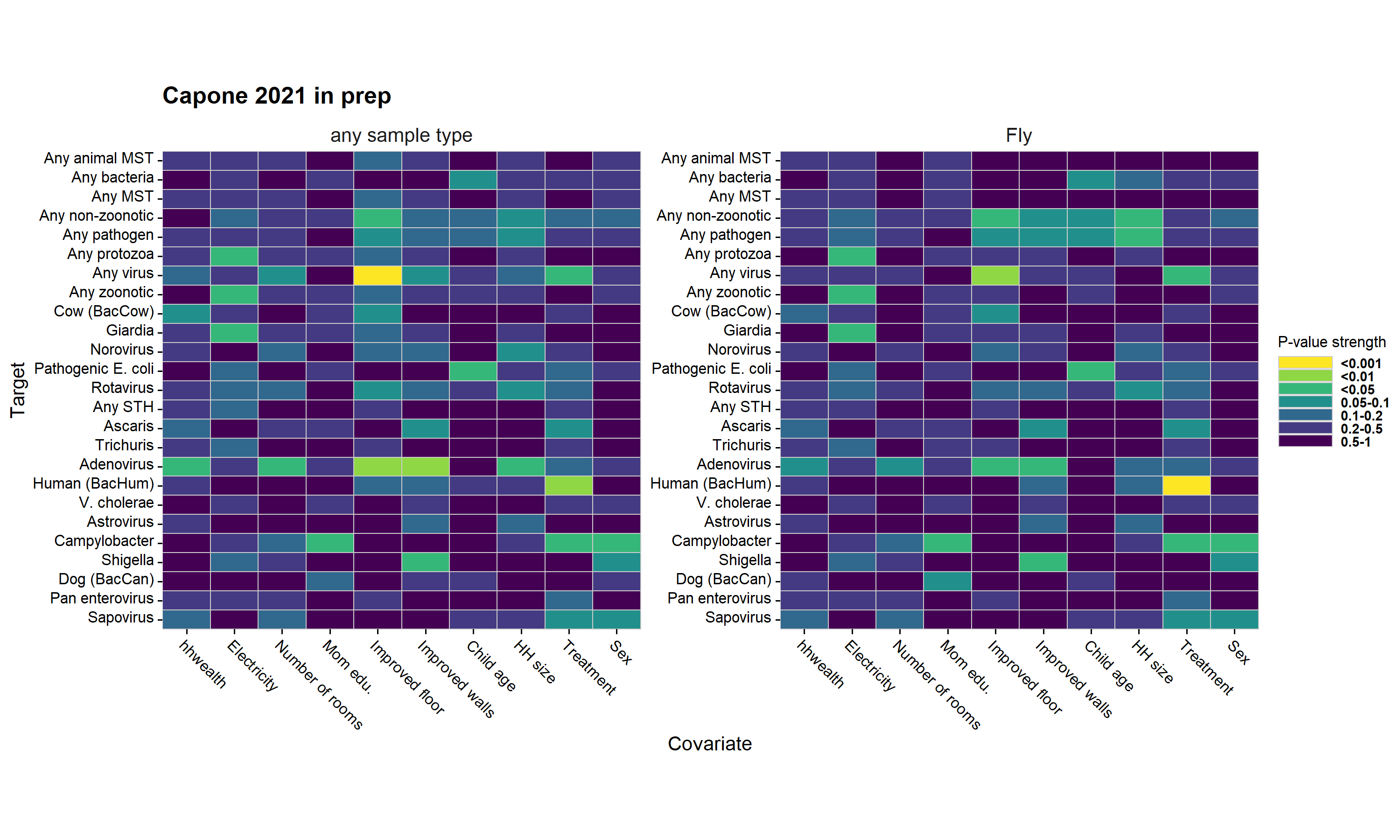
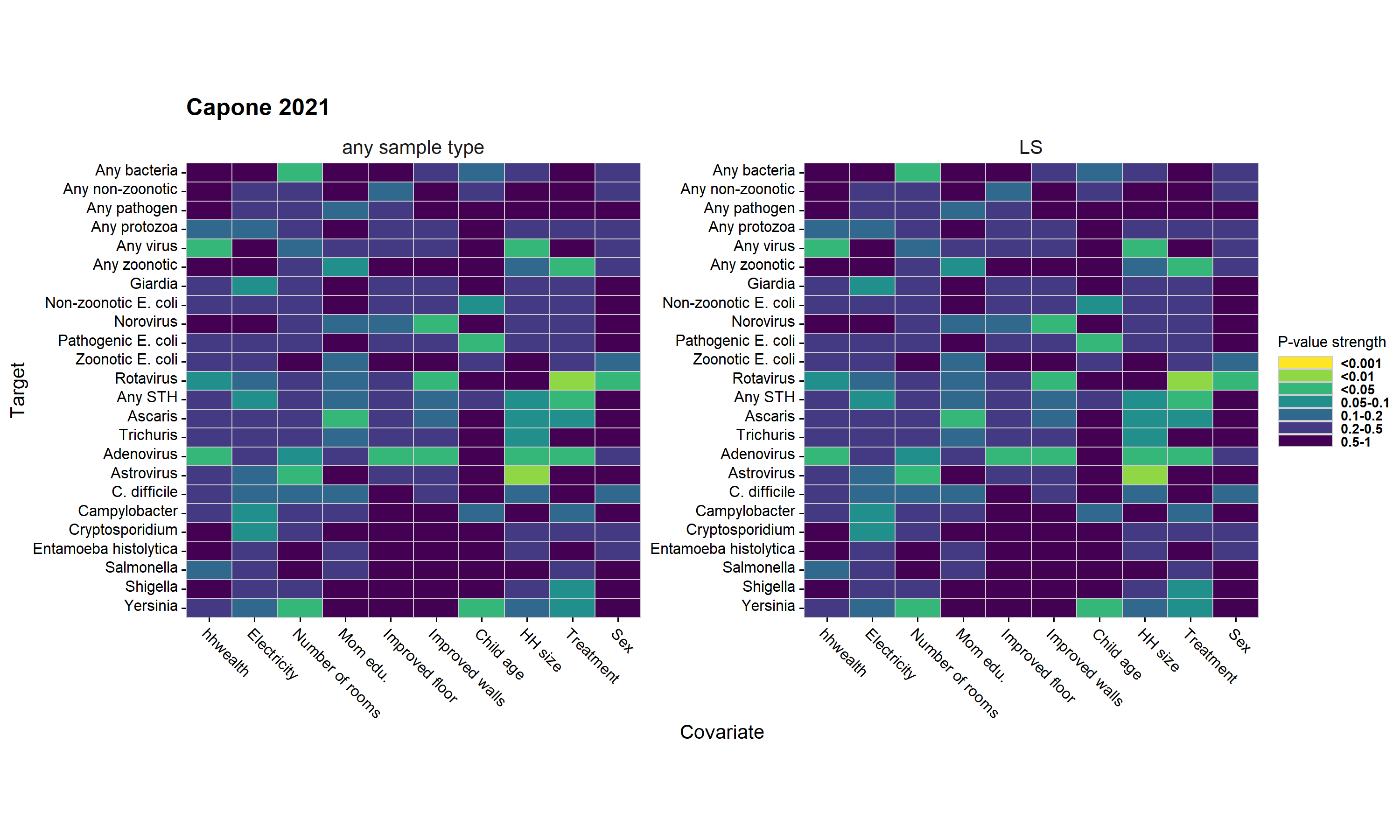
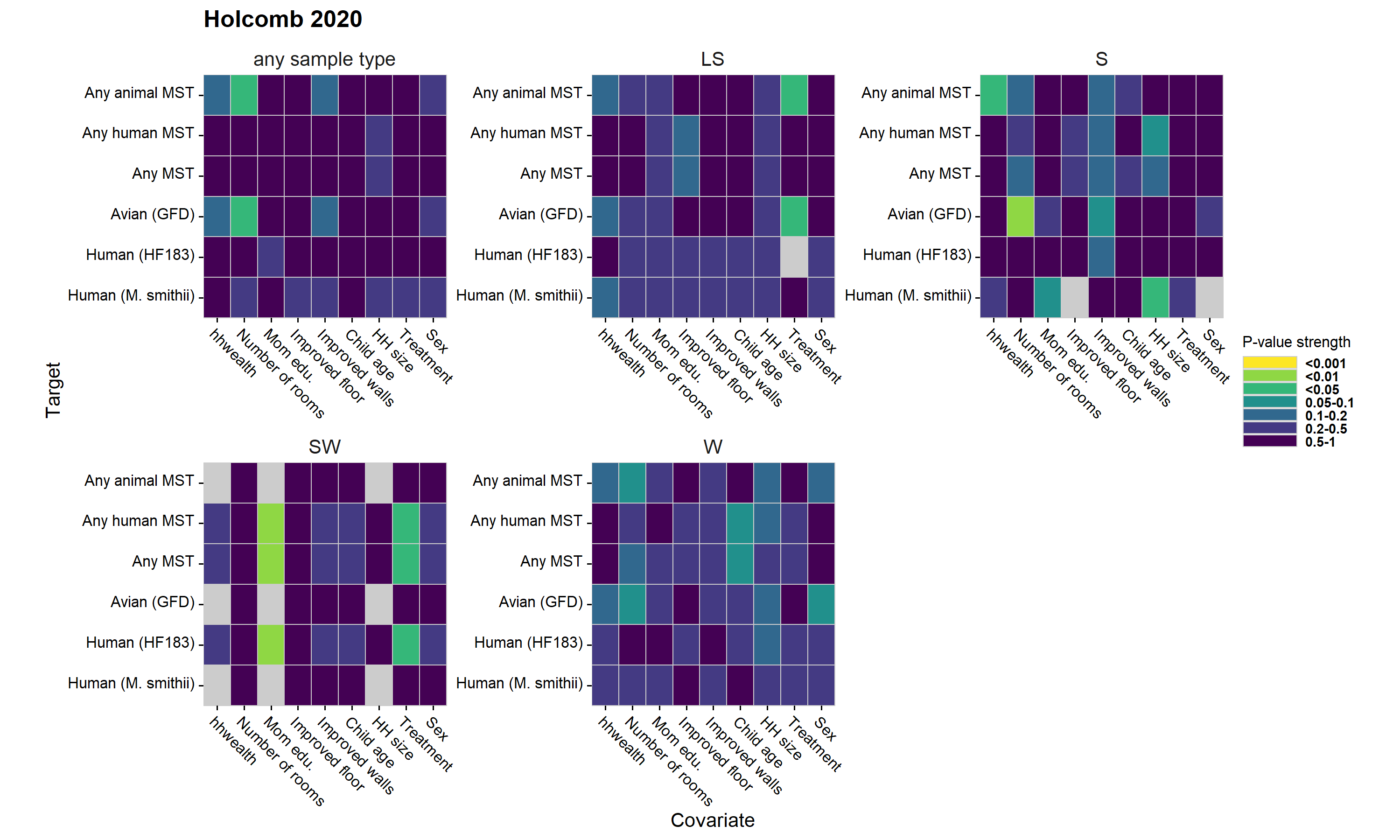
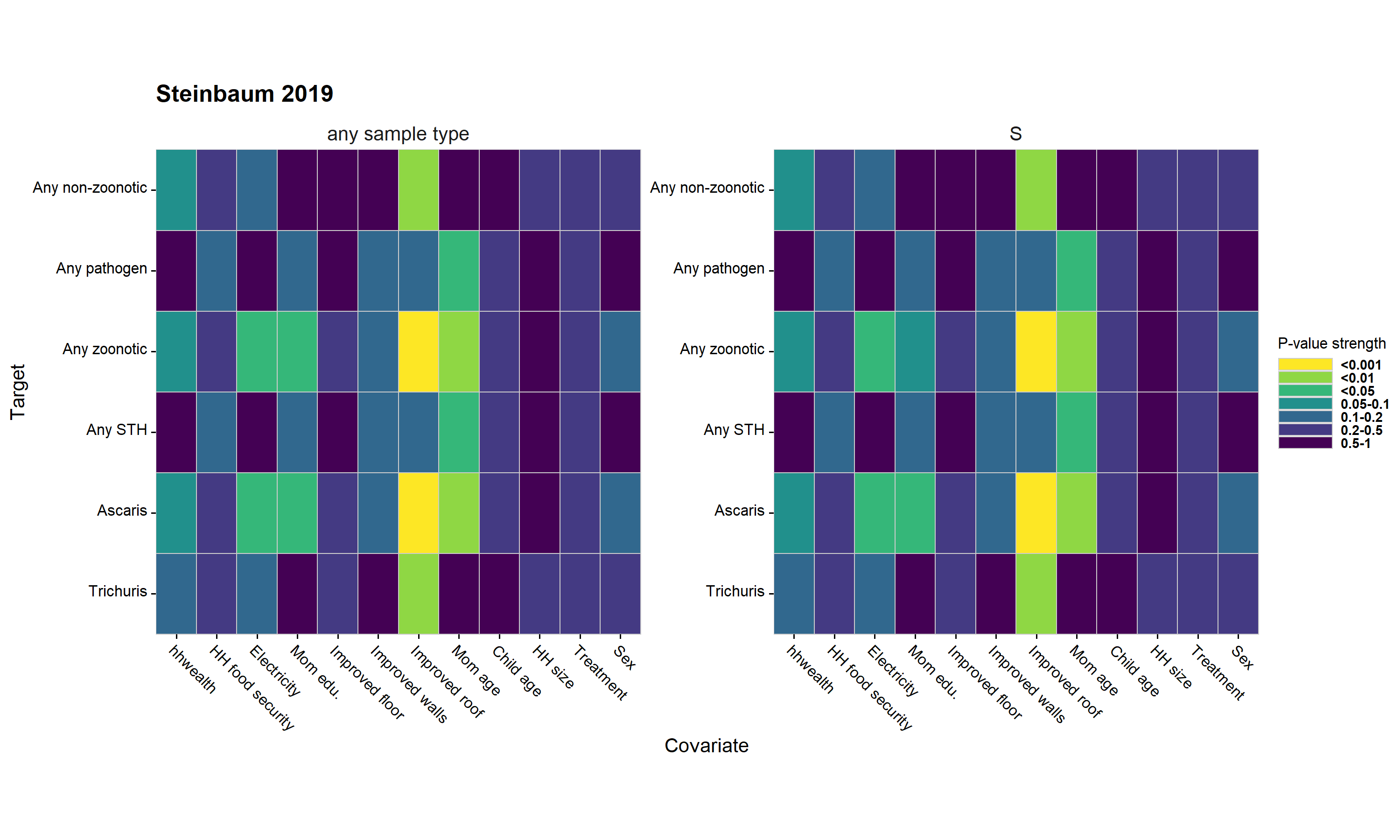
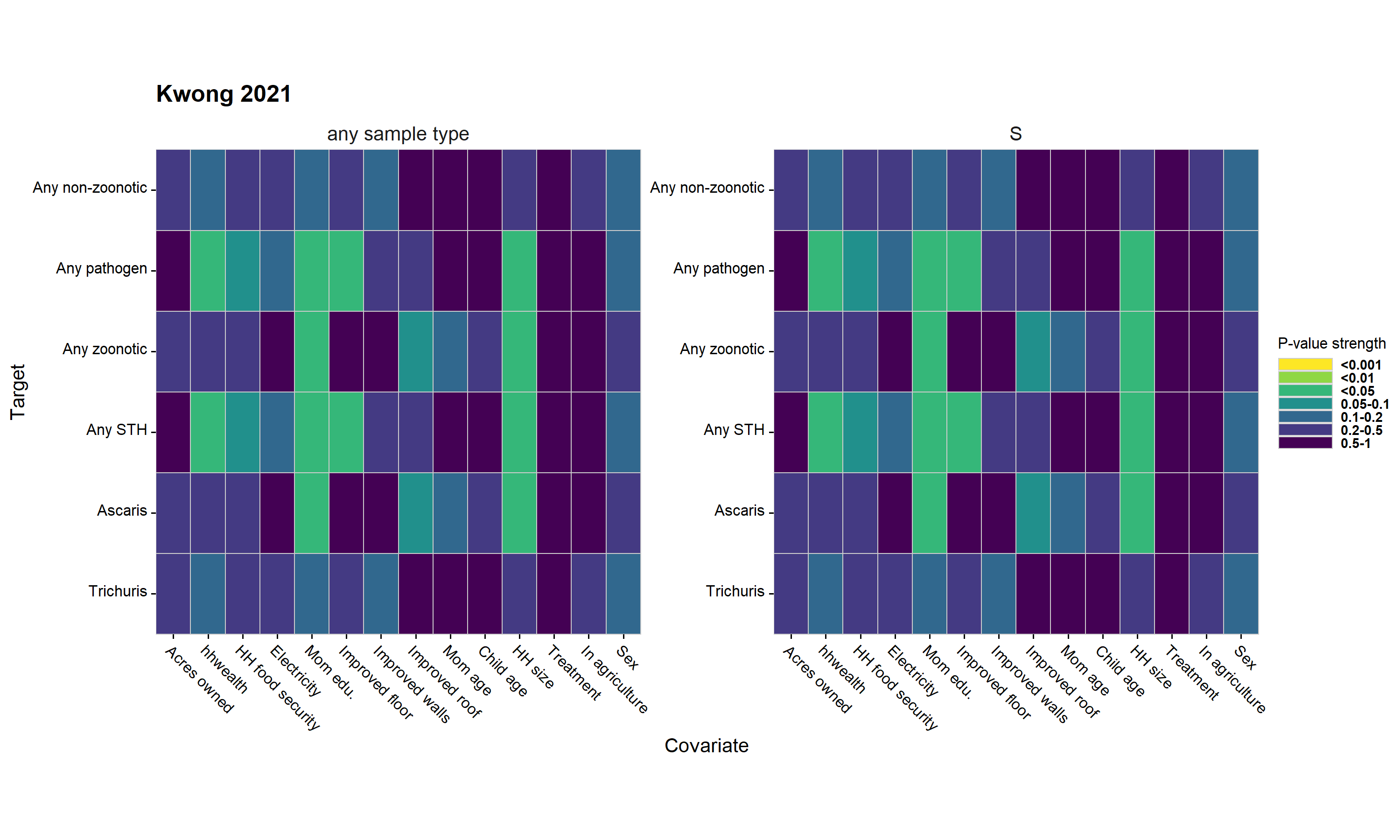
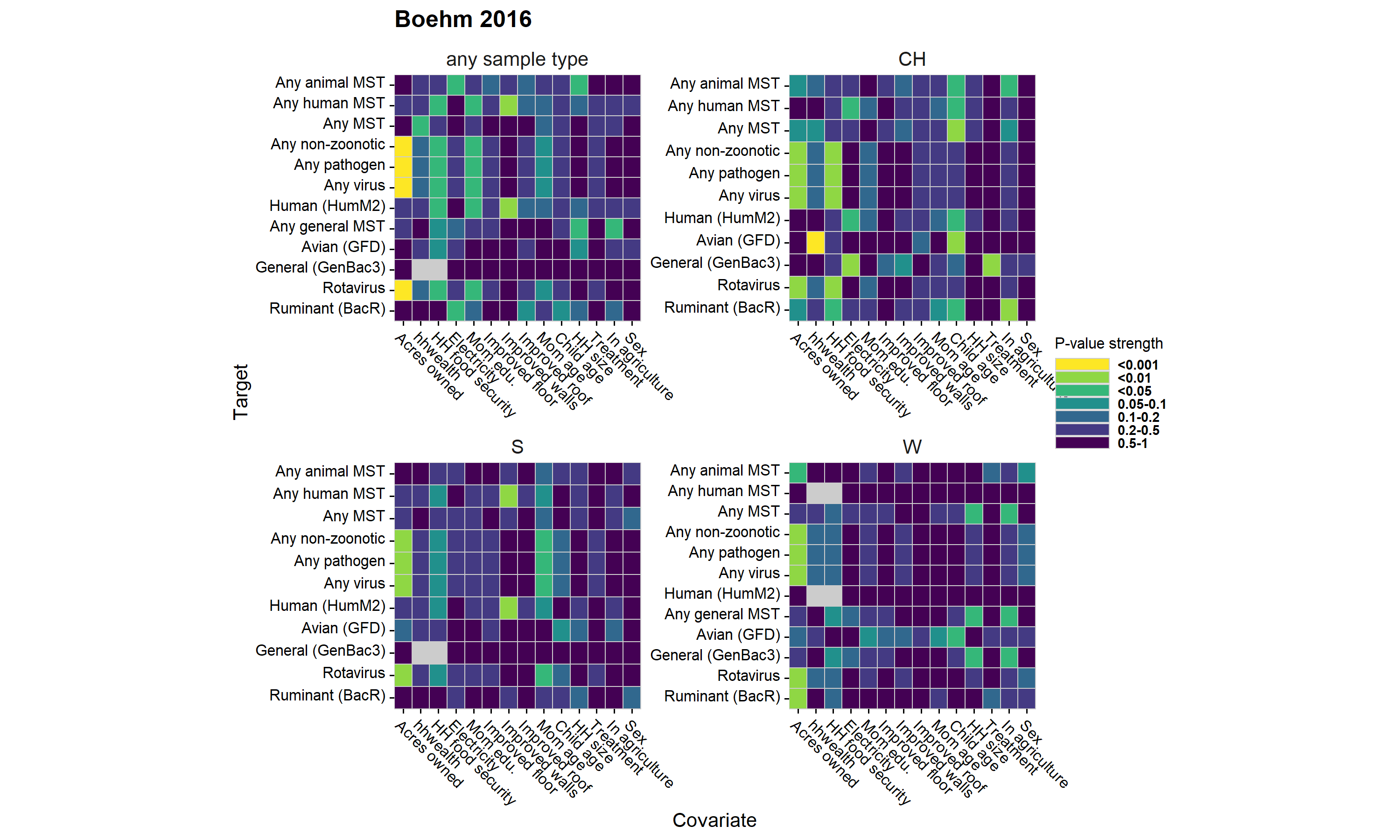
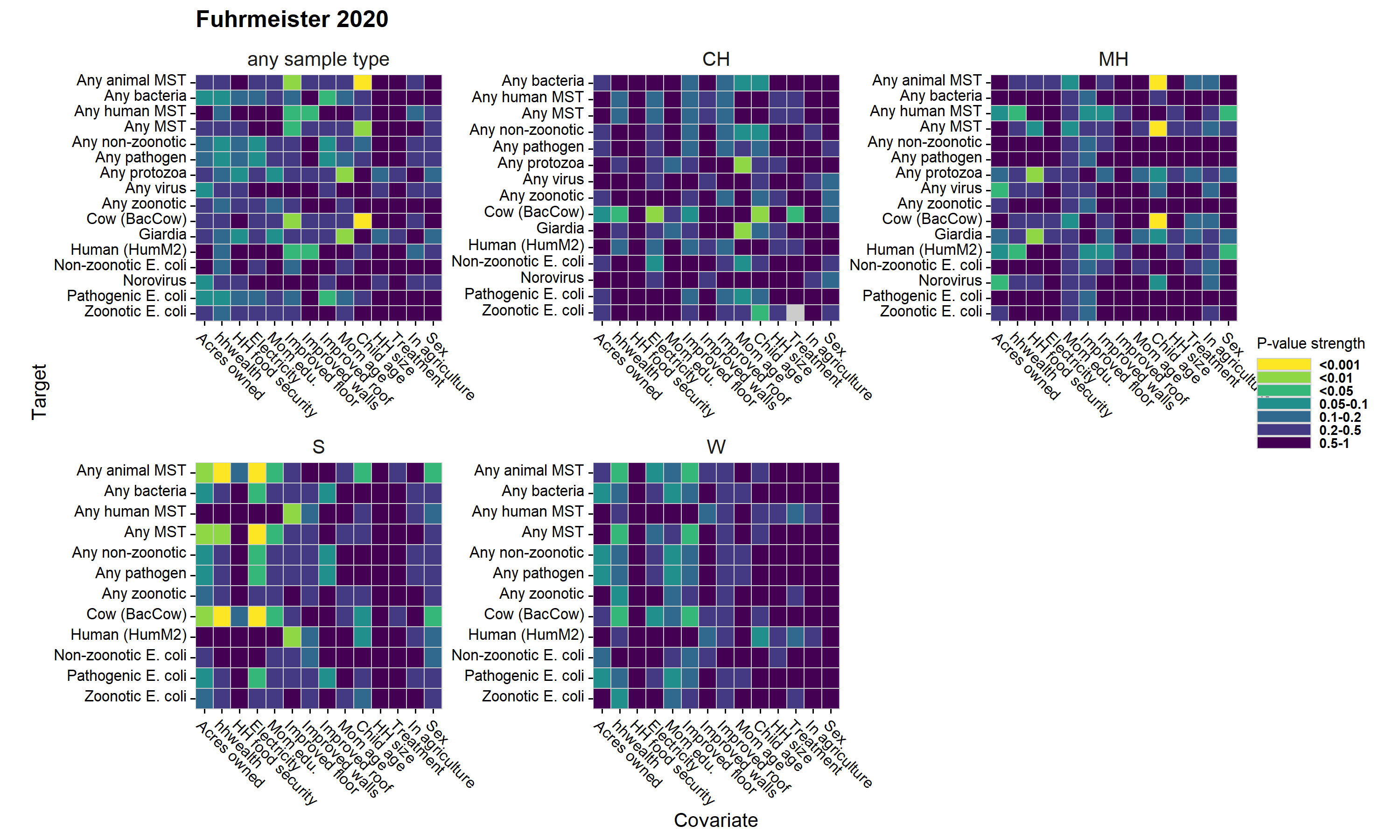
**Interpretation:** The estimation approach chosen did not affect our conclusions about associations between environmental contamination and diarrheal disease.



**Figure 23.** 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.

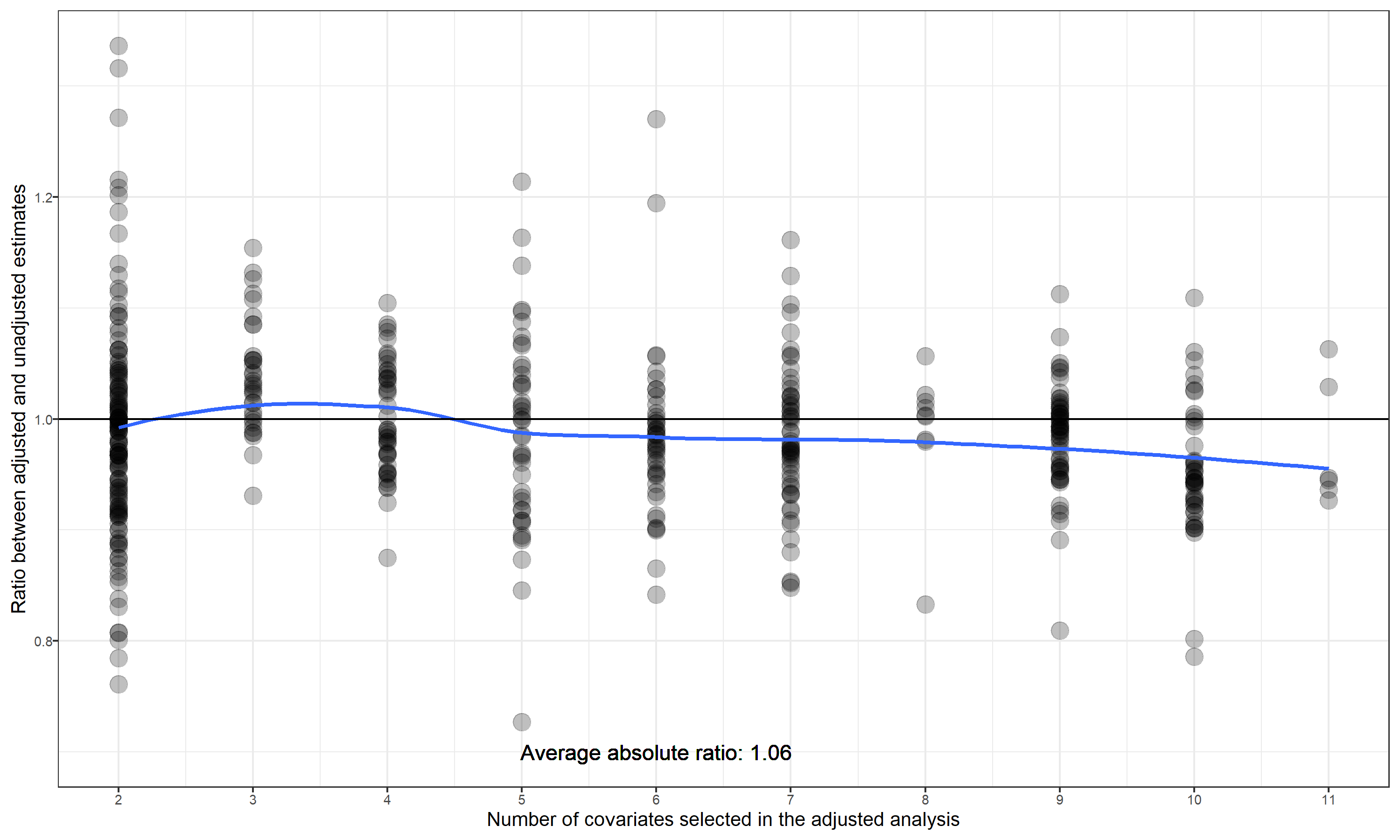
**Interpretation:** The estimation approach chosen did not affect our conclusions about associations between environmental contamination and child linear growth.

## Covariate tables



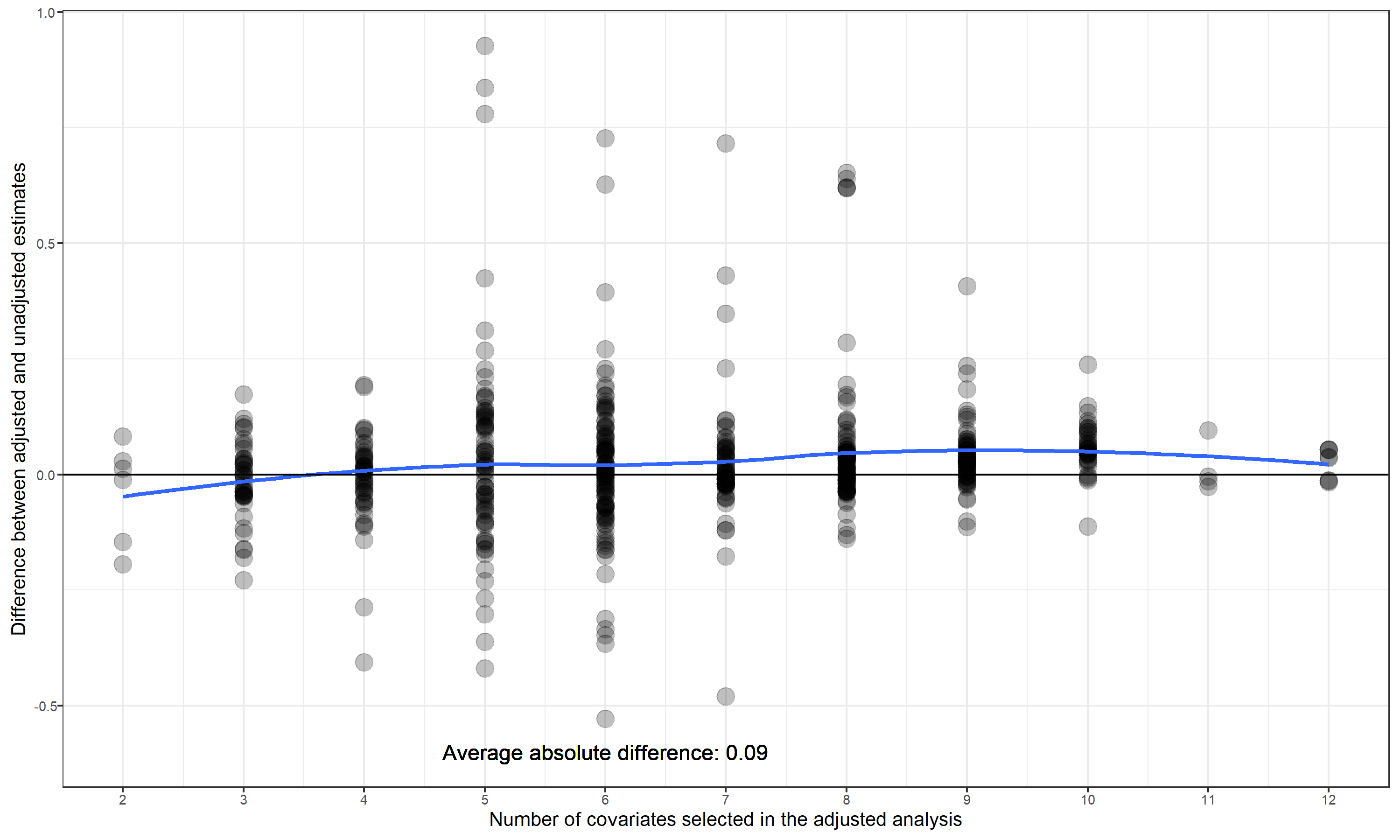
**Figure 24.** Associations between adjustment covariates and the presence of different enteropathogen and MST markers in different environmental samples. The columns are different pre-screened confounders, and the rows are specific enteropathogens and MST markers. Cells of the heatmaps are colored by P-values of bivariate likelihood ratio tests, and heatmaps are stratified by sample type.

**Interpretation:** 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 adequetly adjusted for measured confounding, but unmeasured confounding may bias the results.



**Figure 25.** 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.

**Interpretation:** The average ratio in binary outcomes between environmental samples positive for enteropathogens or MST markers and those negative for enteropathogens or MST markers is 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.



**Figure 26.** 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.

**Interpretation:** The average difference in Z-scores between environmental samples positive for enteropathogens or MST markers and those negative for enteropathogens or MST markers is slightly larger in magnitude after covariate adjustment, but on average the covariate adjustment had small effects on the results and did not vary by number of covariates used for adjustment.

# Tables

## Data availability tables

## Outcome table

| **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 |

## Data availability tables

## Any sample type, any enteropathogen

| **study** | **sample** | **target** | **N samples** | **N pos. samples** | **N diar. meas.** | **N diar. pos.** | **N sample. and diar. pos.** | **N haz** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Fuhrmeister 2020 | any sample type | Any pathogen | 1,643 | 1,180 | 1,590 | 188 | 128 | 857 |
| Boehm 2016 | any sample type | Any pathogen | 497 | 34 | 412 | 99 | 11 | 411 |
| Kwong 2021 | any sample type | Any pathogen | 2,543 | 1,901 | 703 | 43 | 29 | 758 |
| Steinbaum 2019 | any sample type | Any pathogen | 2,234 | 414 | 1,874 | 485 | 97 | 1,761 |
| Capone 2021 | any sample type | Any pathogen | 566 | 531 | 167 | 21 | 20 | 253 |
| Capone 2021 in prep | any sample type | Any pathogen | 487 | 285 | 195 | 19 | 10 | 213 |
| Reese 2017 | any sample type | Any pathogen | 1,044 | 274 | 84 | 9 | 4 | 578 |
| Odagiri 2016 | any sample type | Any pathogen | 4,825 | 3,787 | 2,038 | 188 | 117 | 0 |

## Any sample type, any MST

| **study** | **sample** | **target** | **N samples** | **N pos. samples** | **N diar. meas.** | **N diar. pos.** | **N sample. and diar. pos.** | **N haz** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Fuhrmeister 2020 | any sample type | Any MST | 1,636 | 1,533 | 1,583 | 188 | 175 | 850 |
| Boehm 2016 | any sample type | Any MST | 497 | 490 | 412 | 99 | 97 | 411 |
| Holcomb 2020 | any sample type | Any MST | 935 | 692 | 292 | 27 | 17 | 412 |
| Capone 2021 in prep | any sample type | Any MST | 487 | 266 | 195 | 19 | 11 | 213 |
| Odagiri 2016 | any sample type | Any MST | 4,825 | 4,672 | 2,038 | 188 | 174 | 0 |

**Interpretation:** 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.

## 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).