

PRIMARY FIGURES

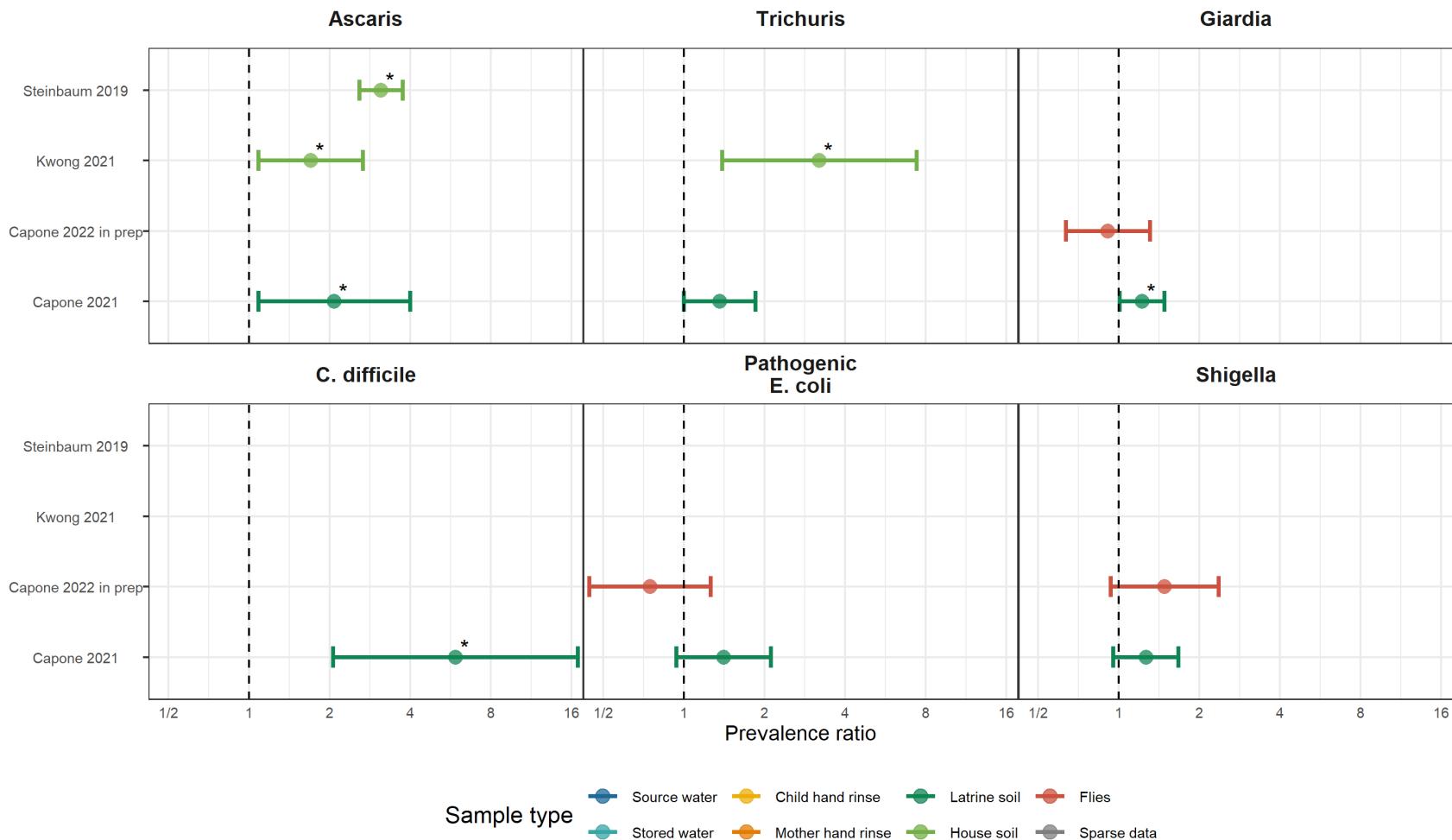


Figure 1. Forest plots of associations between specific enteropathogens in environmental samples and child infections with the same enteropathogens. The presented prevalence ratios compare the detection prevalence of a pathogen in stool between children from compounds where the pathogen was detected 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.

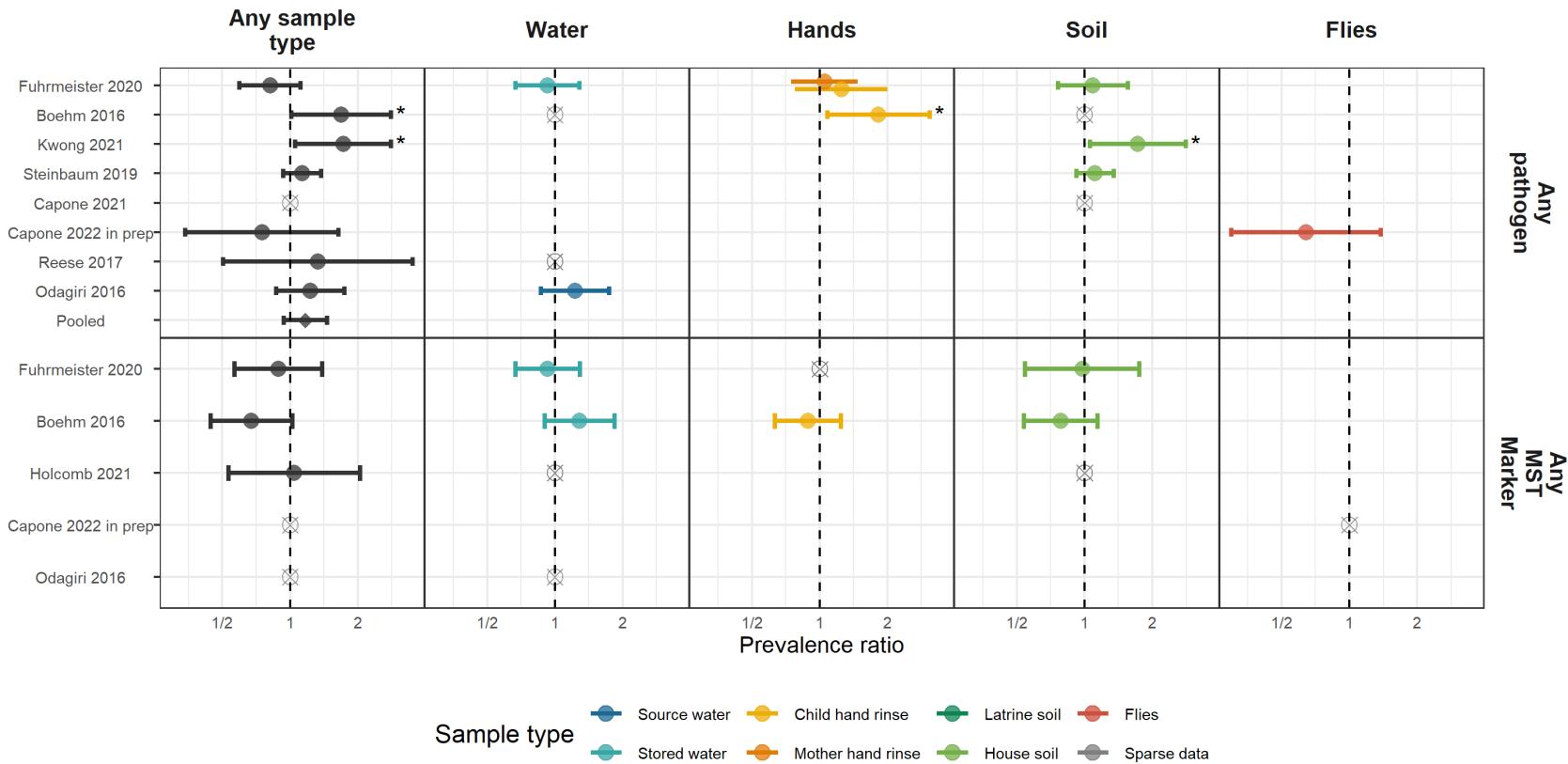


Figure 2. Forest plots of associations between the prevalence of any enteropathogen or any MST markers in different types of environmental samples and child diarrheal disease. The presented prevalence ratios compare diarrhea prevalence between children from compounds where any pathogen/MST marker was 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.

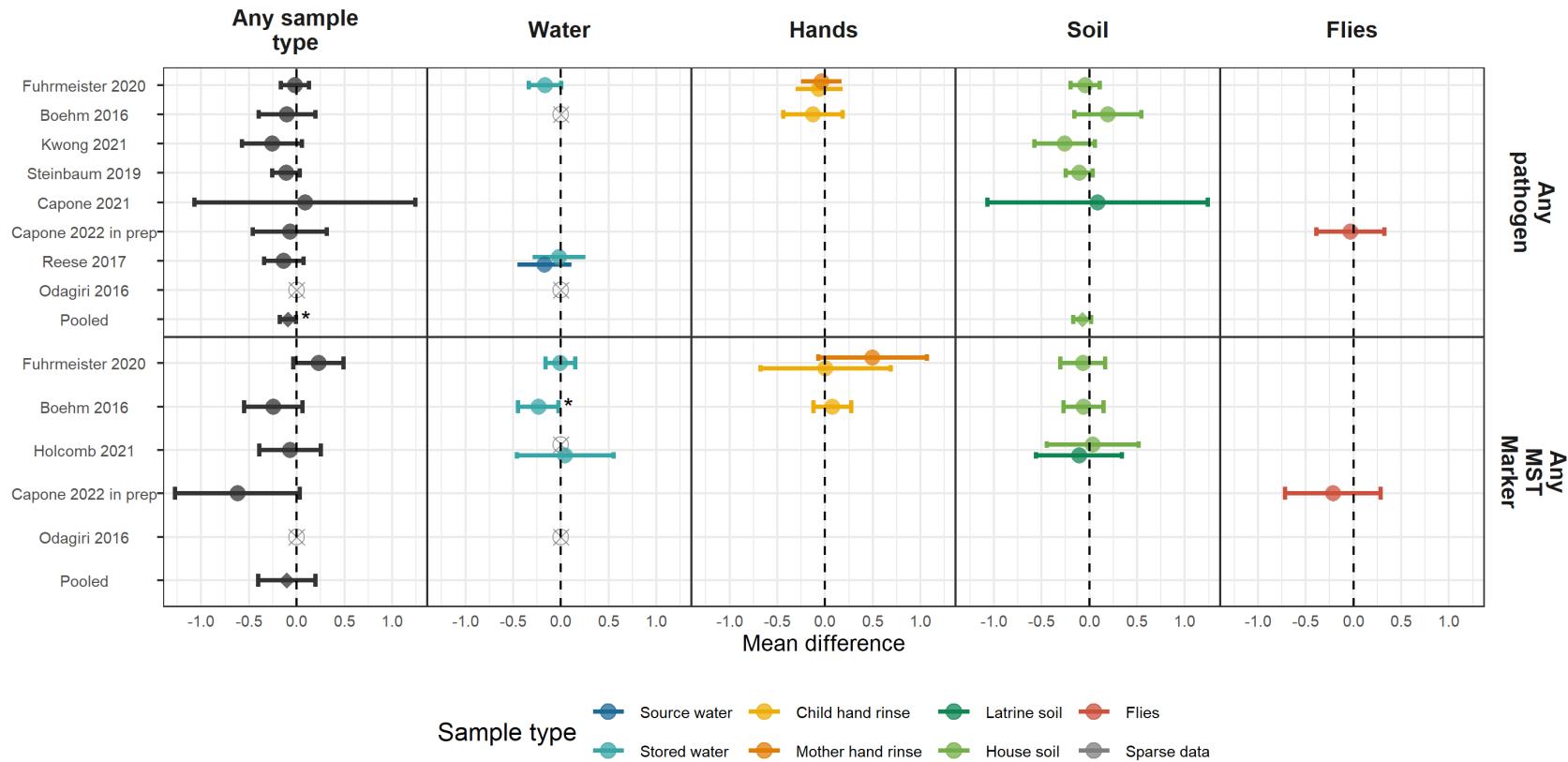
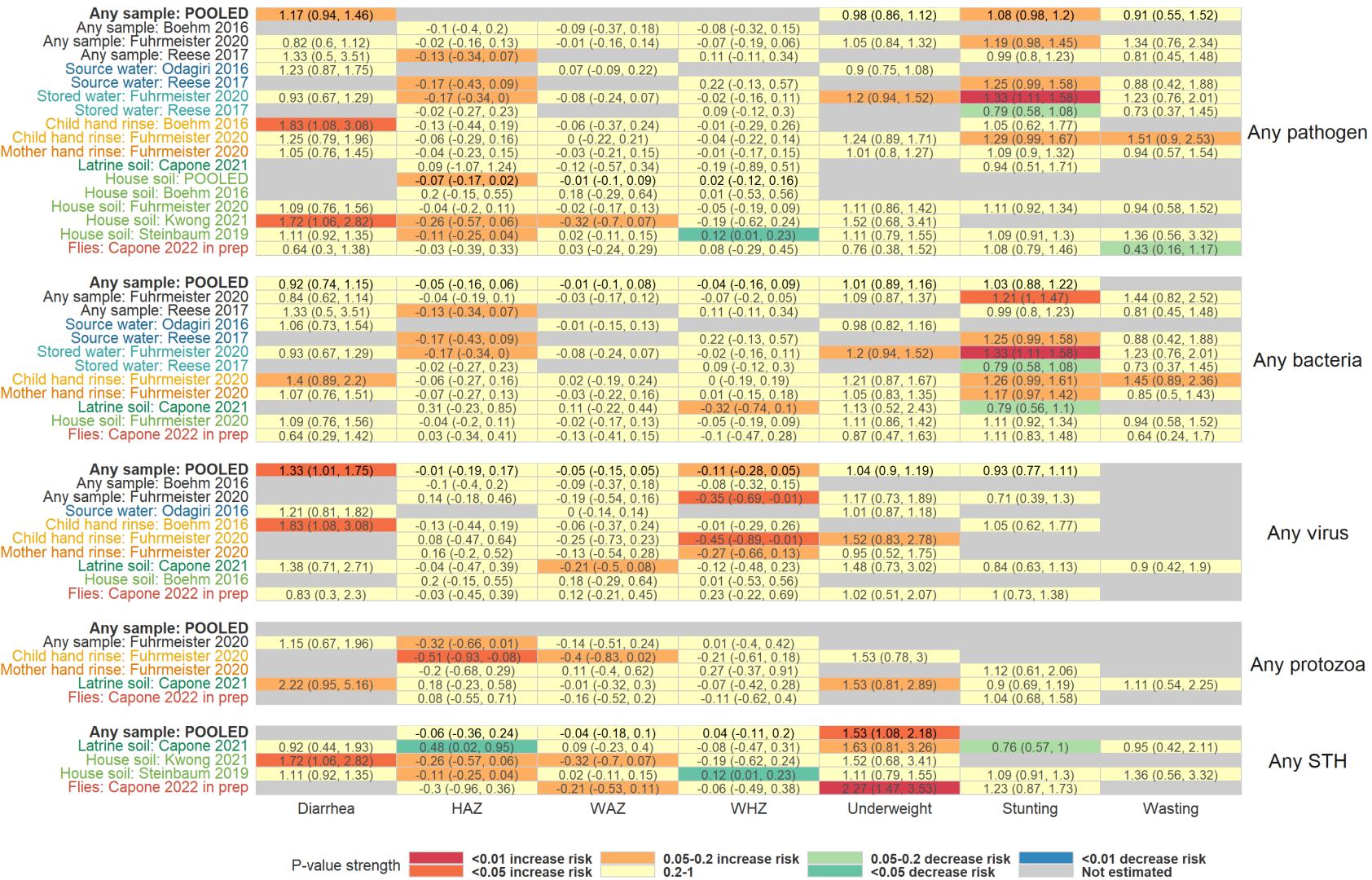
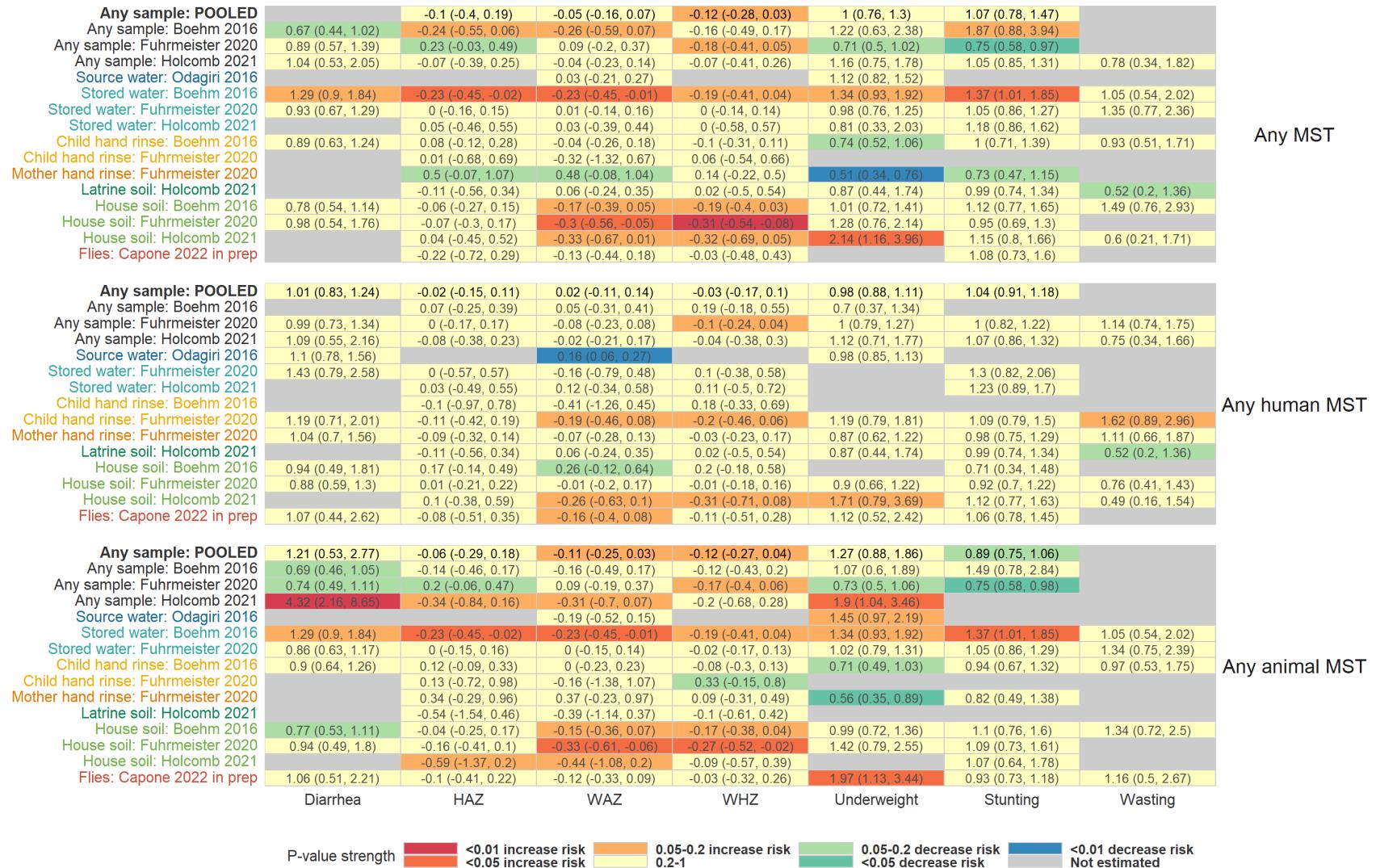
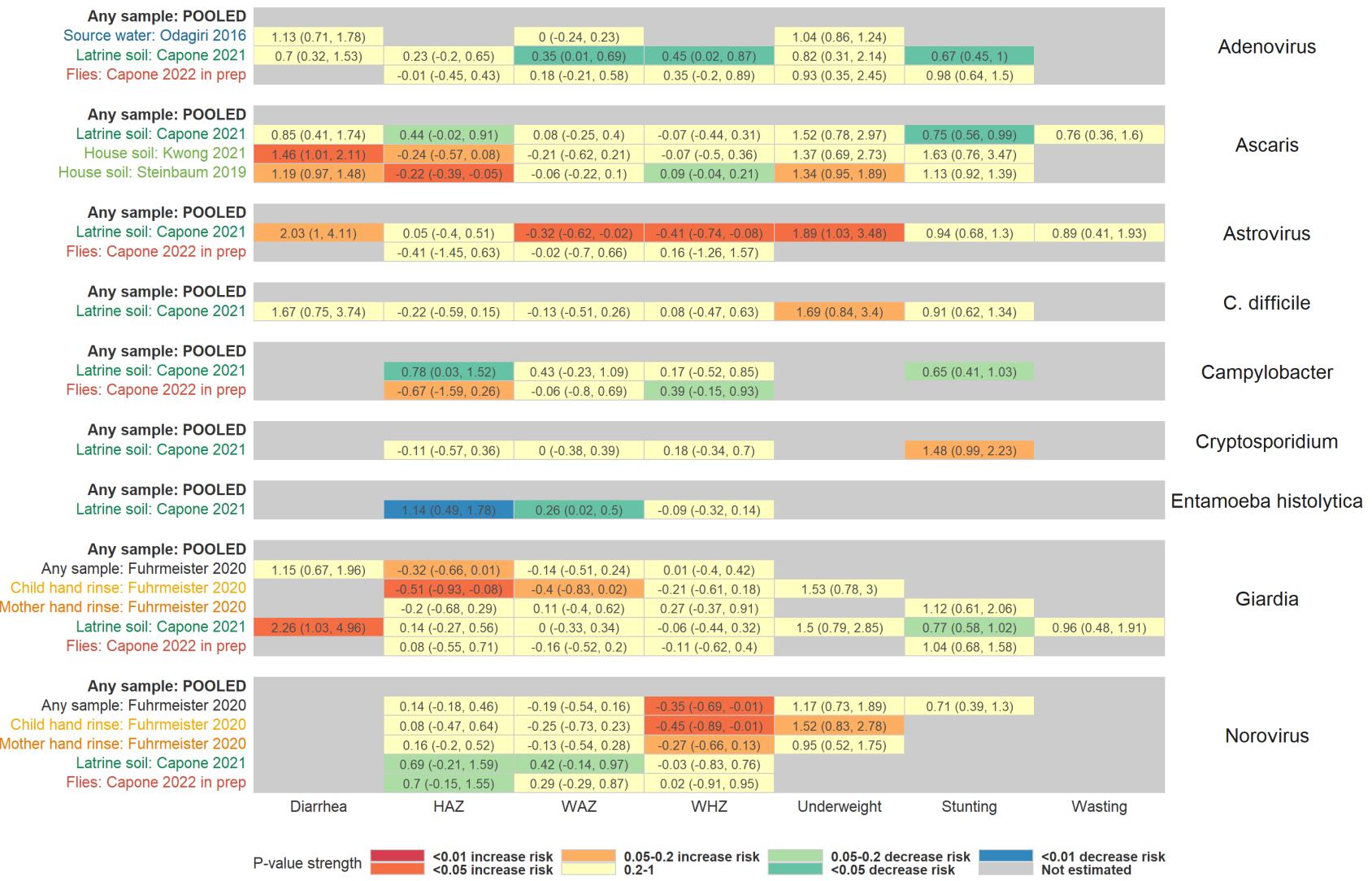


Figure 3. Forest plots of associations between the prevalence of any enteropathogen or any MST markers in different types of environmental samples and heigh-for-age Z-scores (HAZ). The presented differences compare HAZ between children from compounds where any pathogen/MST marker was detected vs. not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







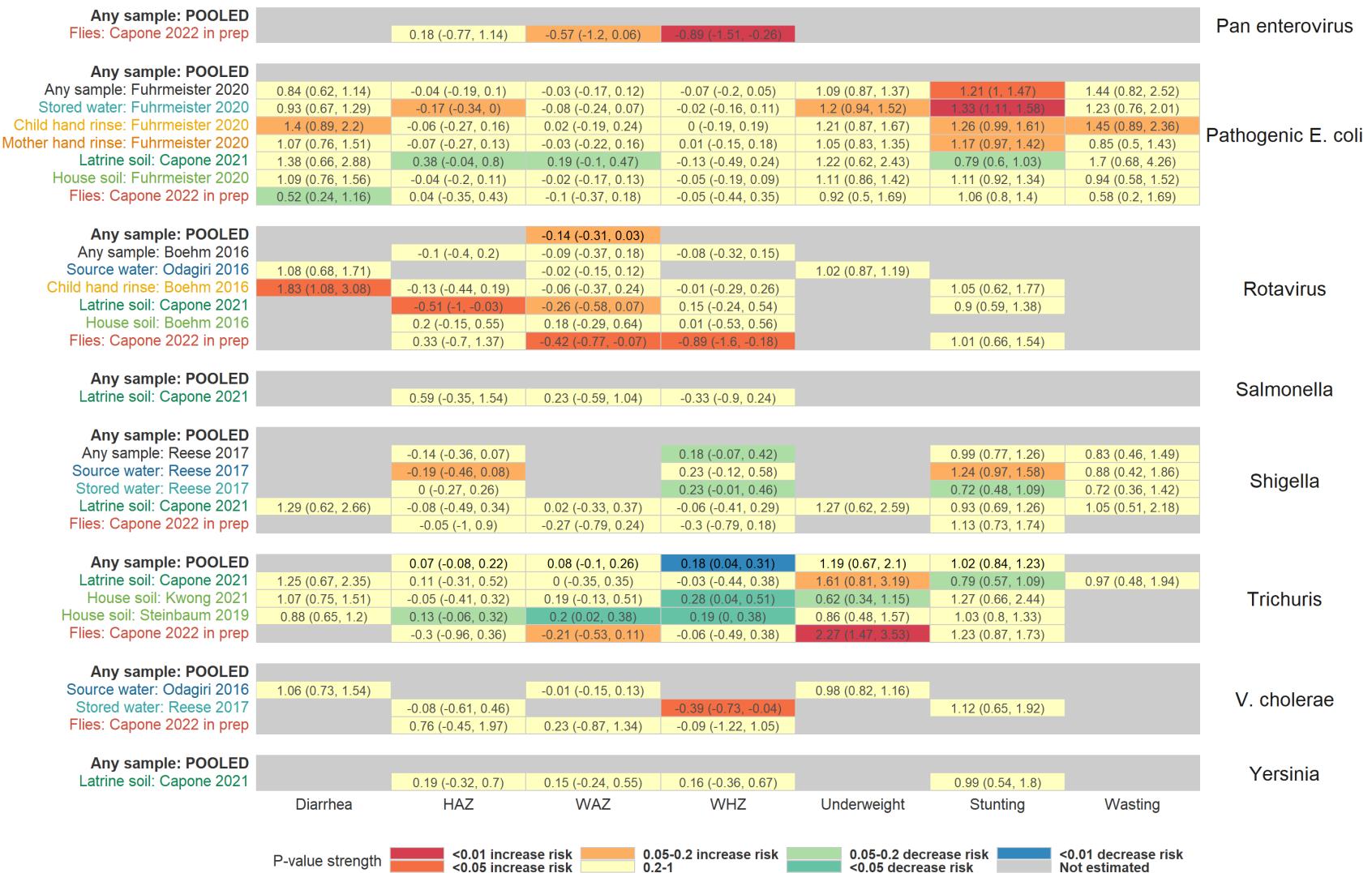


Figure S2. Heatmap of significance and direction of associations between specific pathogens in environmental samples and child diarrhea and growth outcomes. Cells are colored by the strength of significance and direction of association, and the point estimate and confidence intervals are printed within 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.

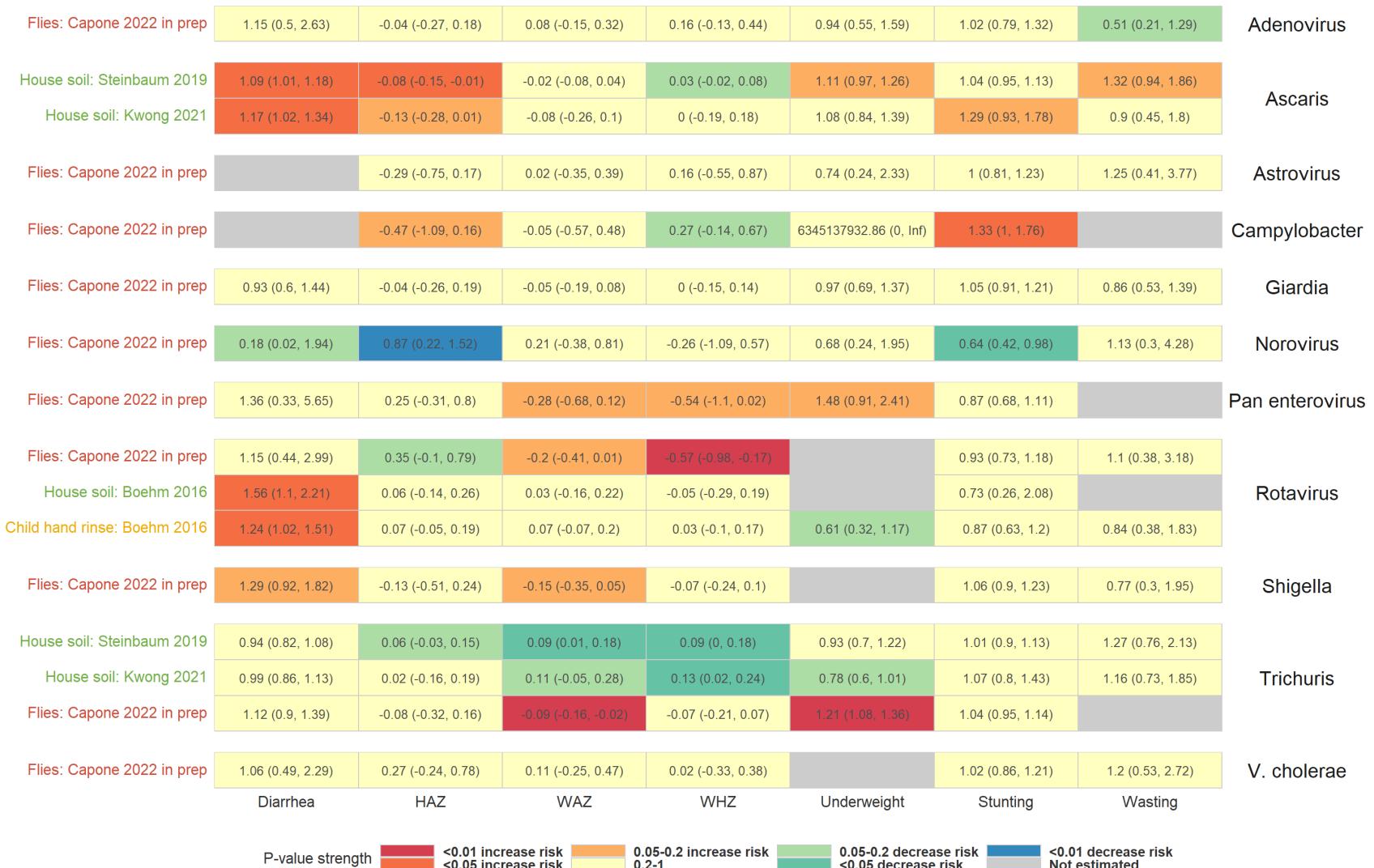


Figure S3. Heatmap of significance and direction of associations between the abundance of specific pathogens in environmental samples and child diarrhea and growth outcomes. Cells are colored by the strength of significance and direction of association, and the point estimate and confidence intervals 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.

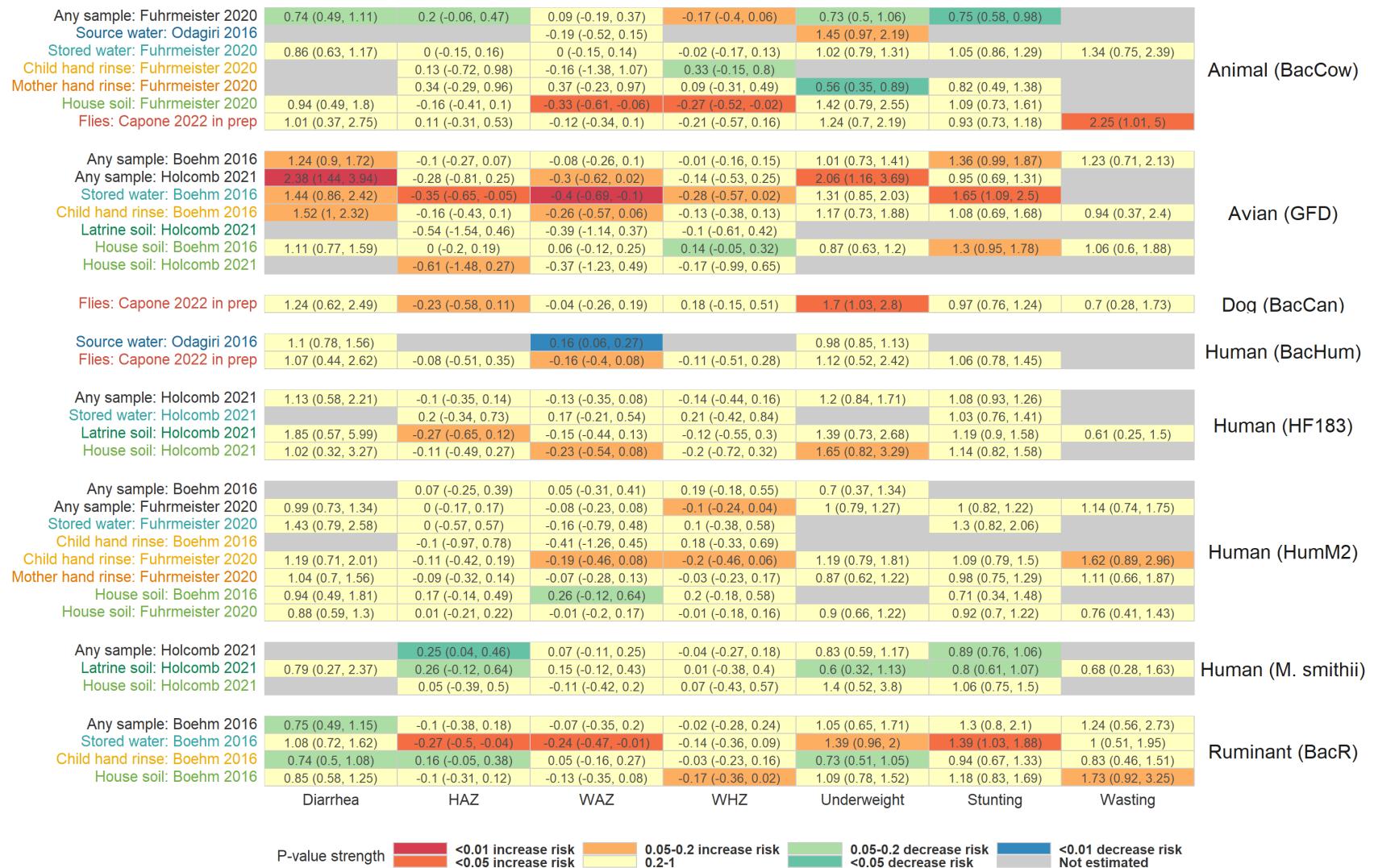


Figure S4. Heatmap of significance and direction of associations between specific microbial source tracking markers in environmental samples and child diarrhea and growth outcomes. Cells are colored by the strength of significance and direction of association, and the point estimate and confidence intervals 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.

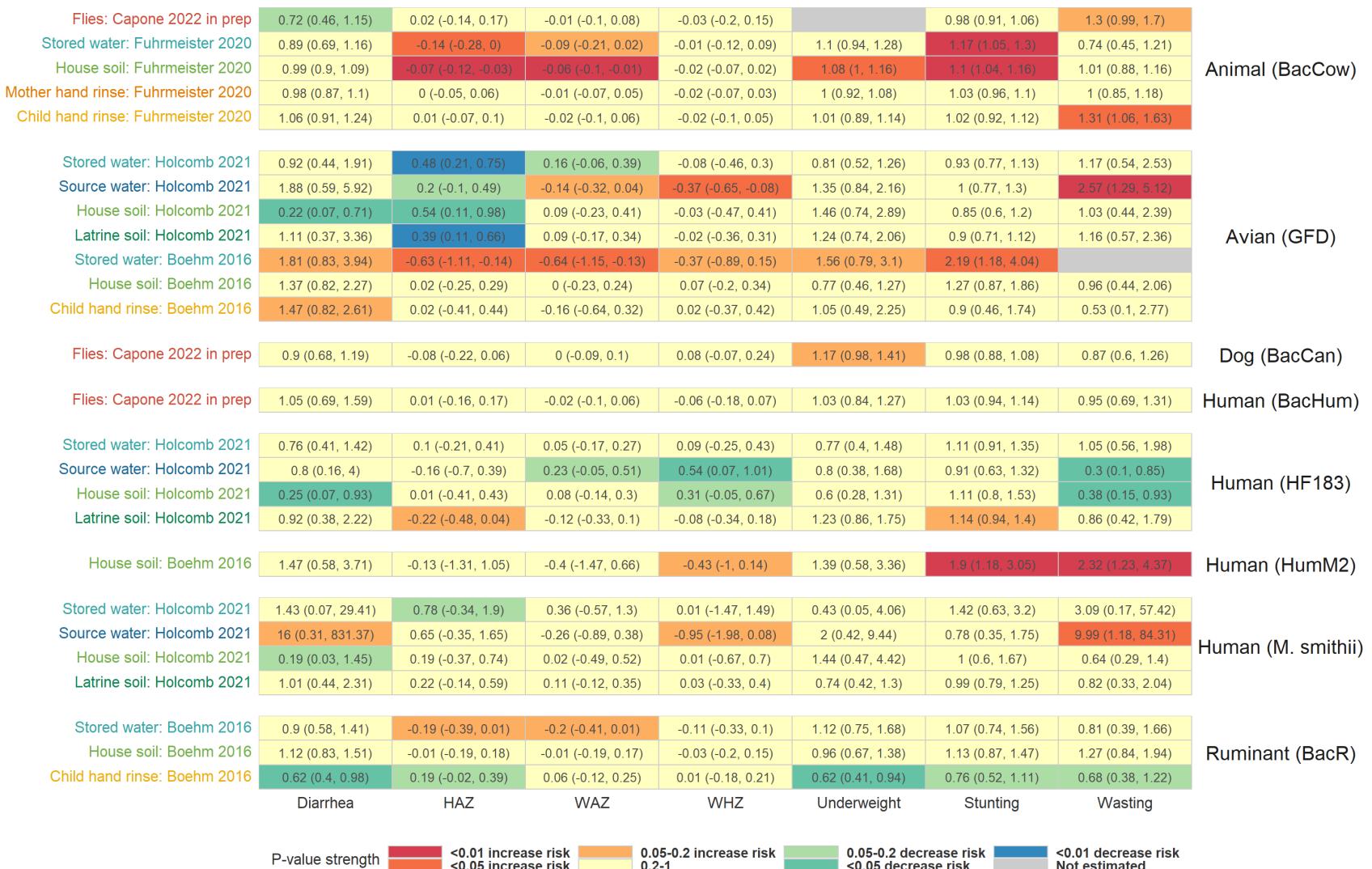


Figure S5. Heatmap of significance and direction of associations between the abundance of specific microbial source tracking markers in environmental samples and child diarrhea and growth outcomes. Cells are colored by the strength of significance and direction of association, and the point estimate 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.

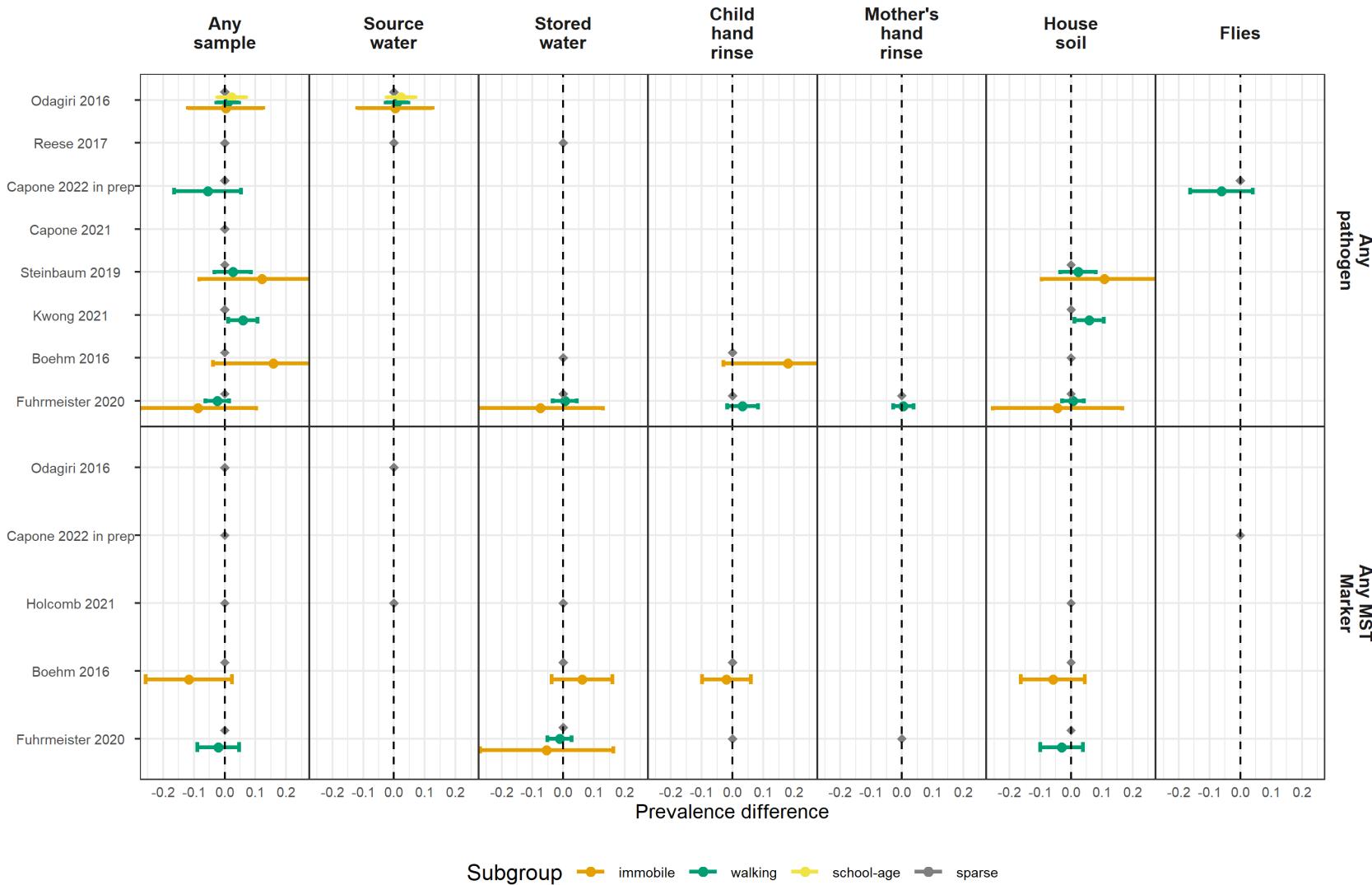


Figure S6. Forest plots of child diarrheal disease prevalence differences between environmental samples with and without any enteropathogen or any MST marker detected, stratified by child age. Grey points mark sparse age strata without estimated relative risks. Significant effect modification, as determined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

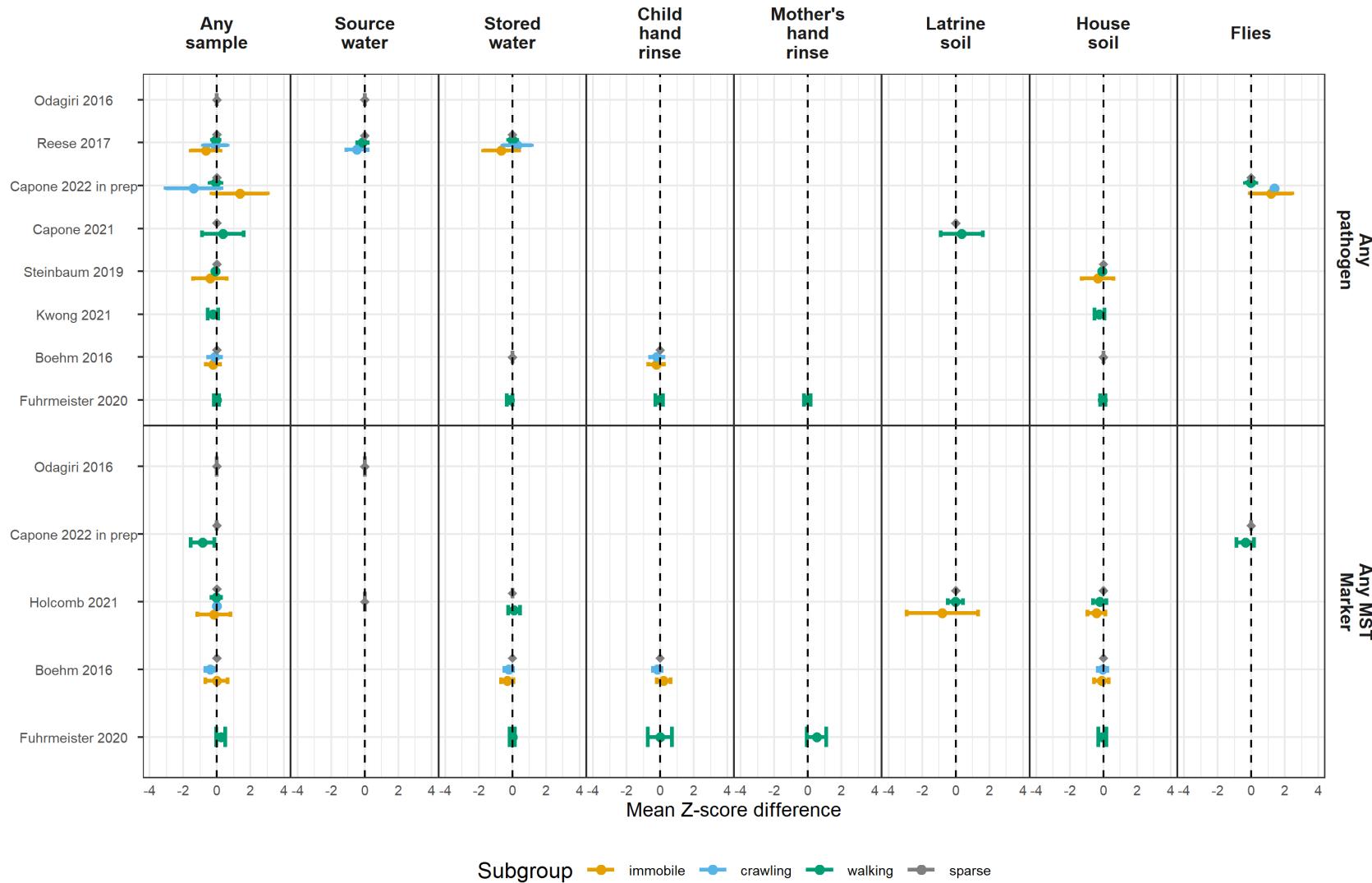


Figure S7. Forest plots of associations between any enteropathogen/any MST markers in different types of environmental samples and child height-for-age Z-score (HAZ), stratified by child age. Grey points mark sparse age strata without estimated mean differences. Significant effect modification, as determined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

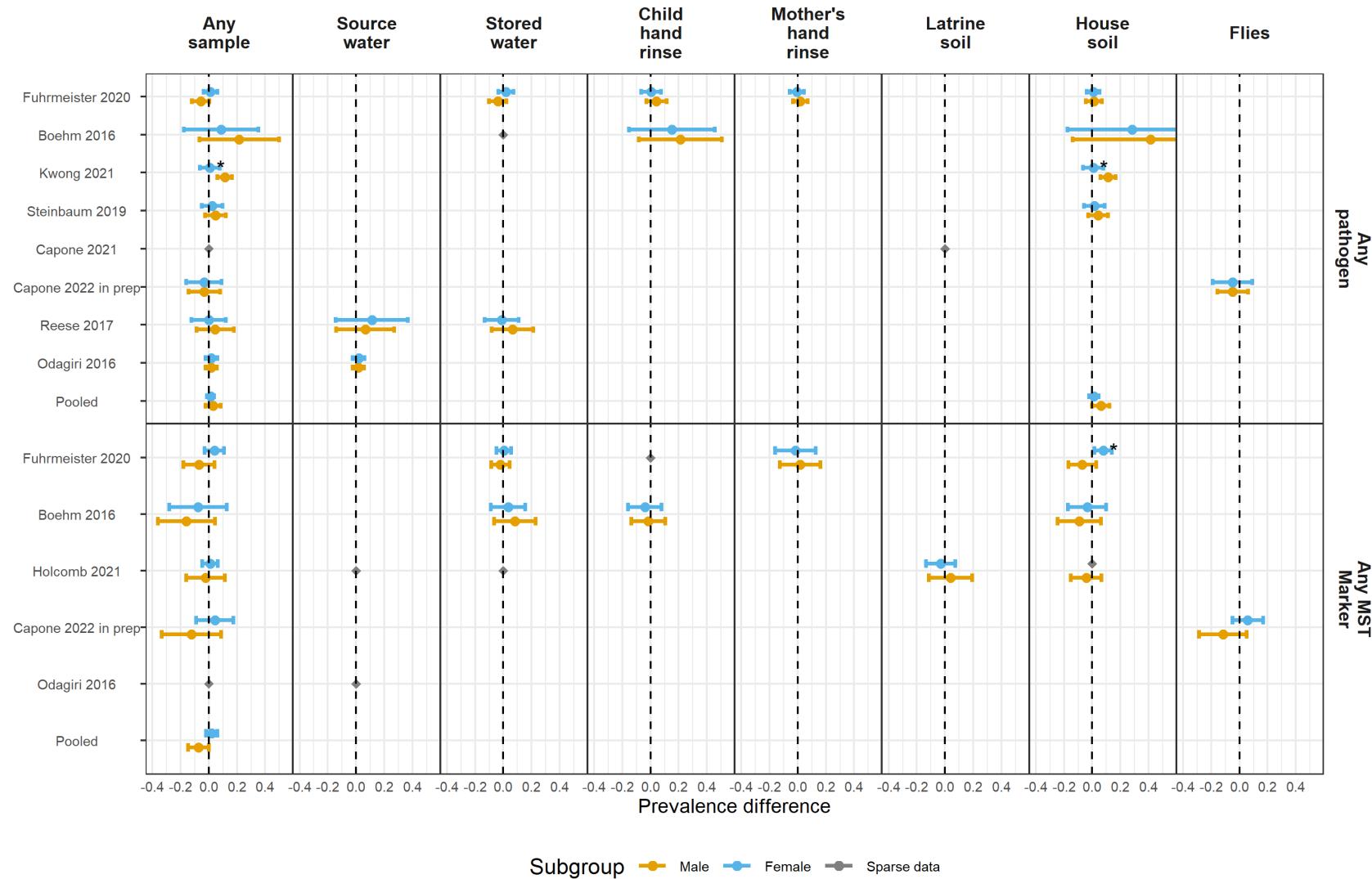


Figure S8. Forest plots of child diarrheal disease prevalence differences between environmental samples with and without any enteropathogen or any MST marker detected, stratified by child sex. Significant effect modification, as determined by the p-values on the regression model interaction term, is marked above points with asterisks ($P < 0.05 = *$, $P < 0.01 = **$, $P < 0.001 = ***$).

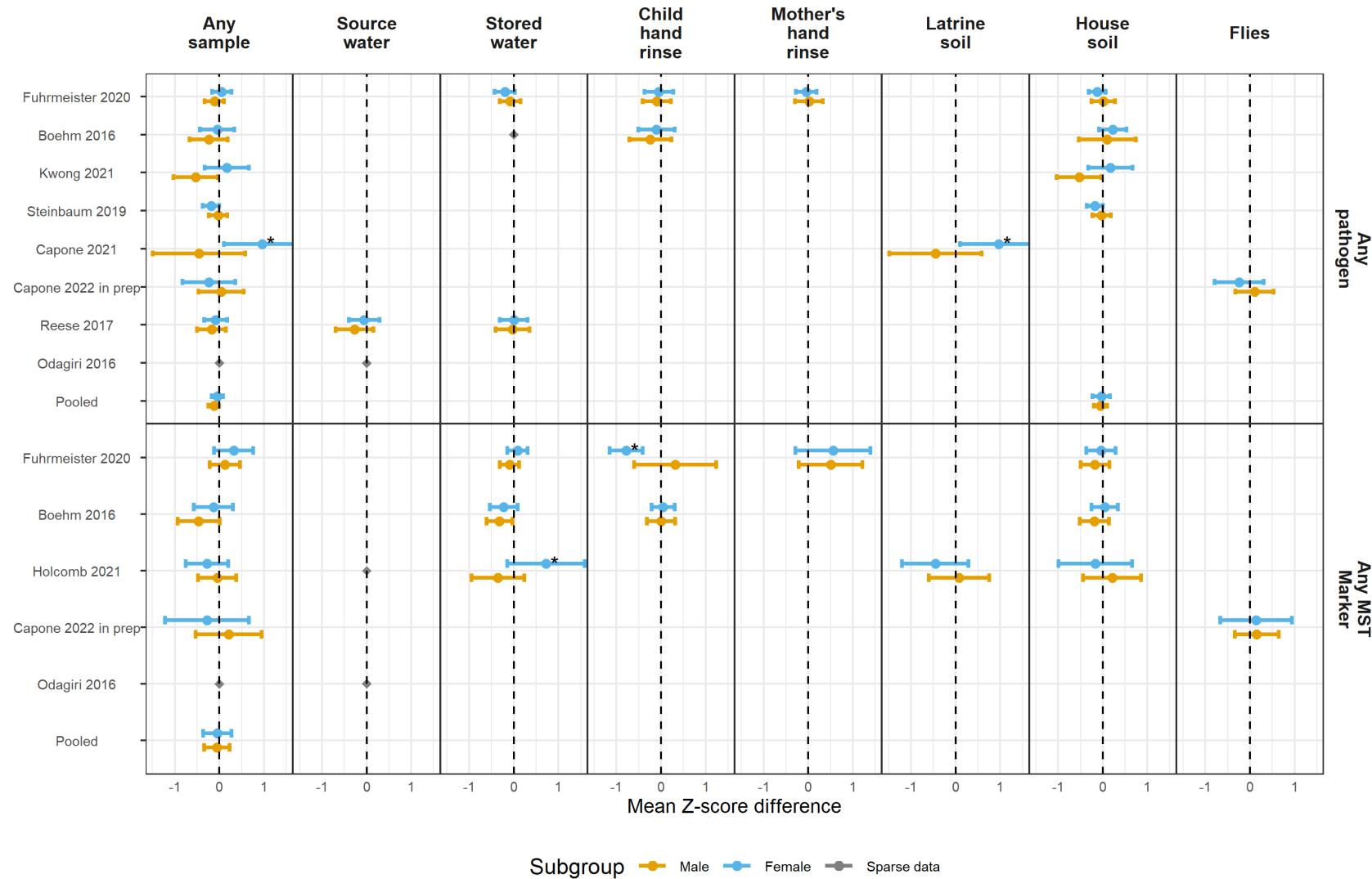


Figure S9. Forest plots of associations between any enteropathogen/any MST markers in different types of environmental samples and child height-for-age Z-scores (HAZ), stratified by child sex. Significant effect modification, as determined by the p-values on the regression model interaction term, is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

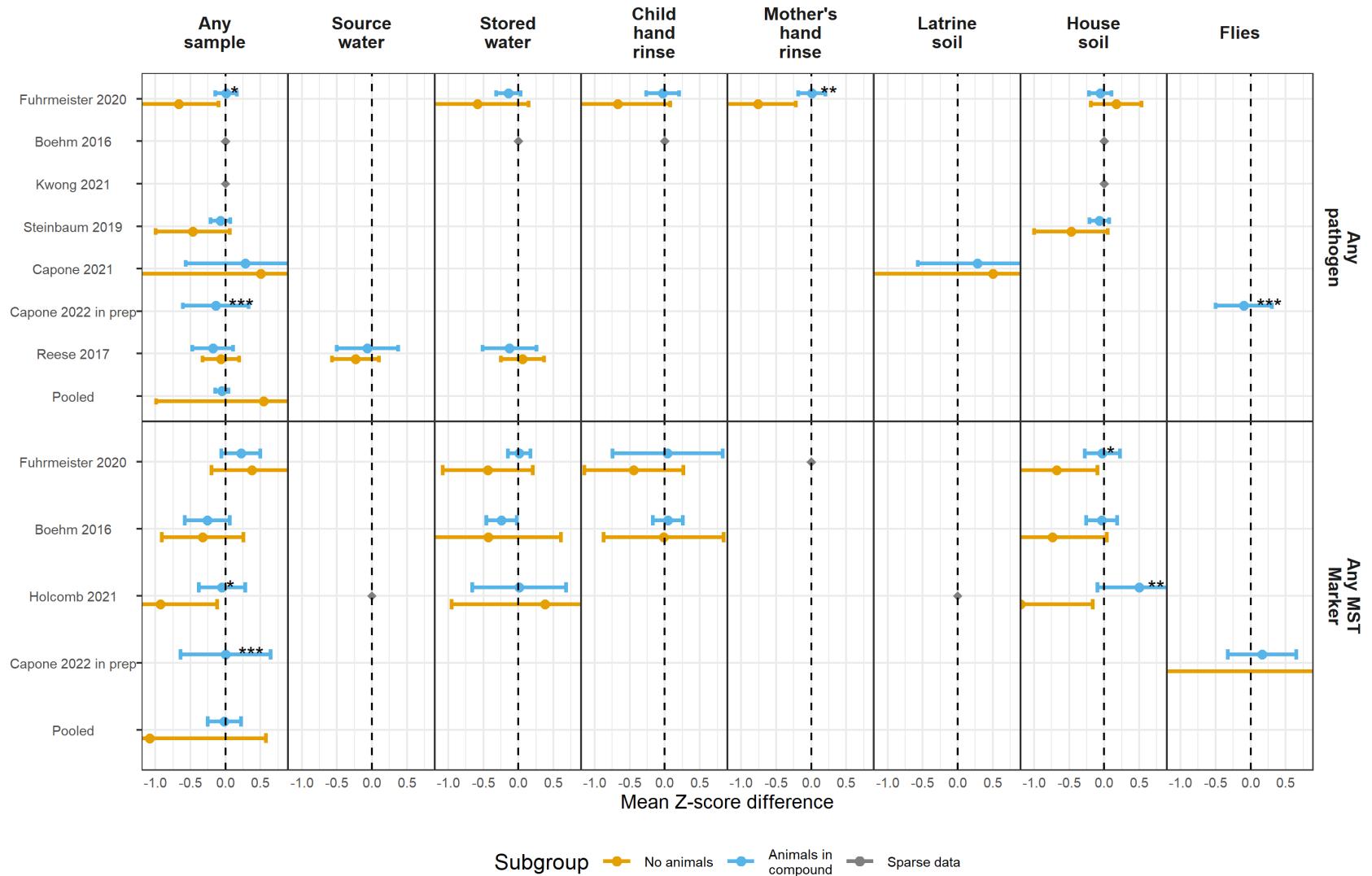


Figure S10. Forest plots of associations between any enteropathogen/any MST markers in different types of environmental samples and child height-for-age Z-scores (HAZ), stratified by whether any animals were present in the compound. Significant effect modification, as determined by the p-values on the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

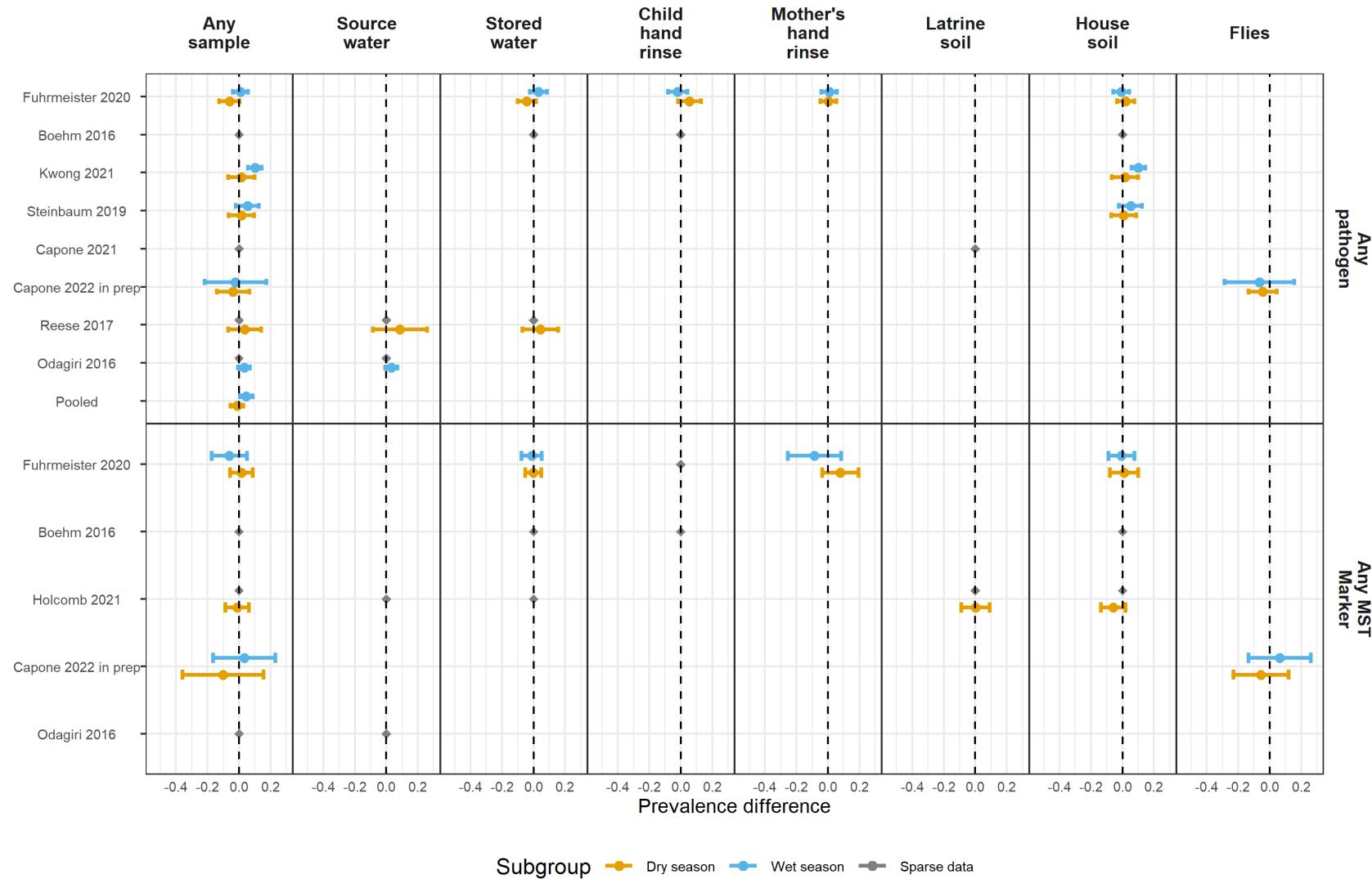
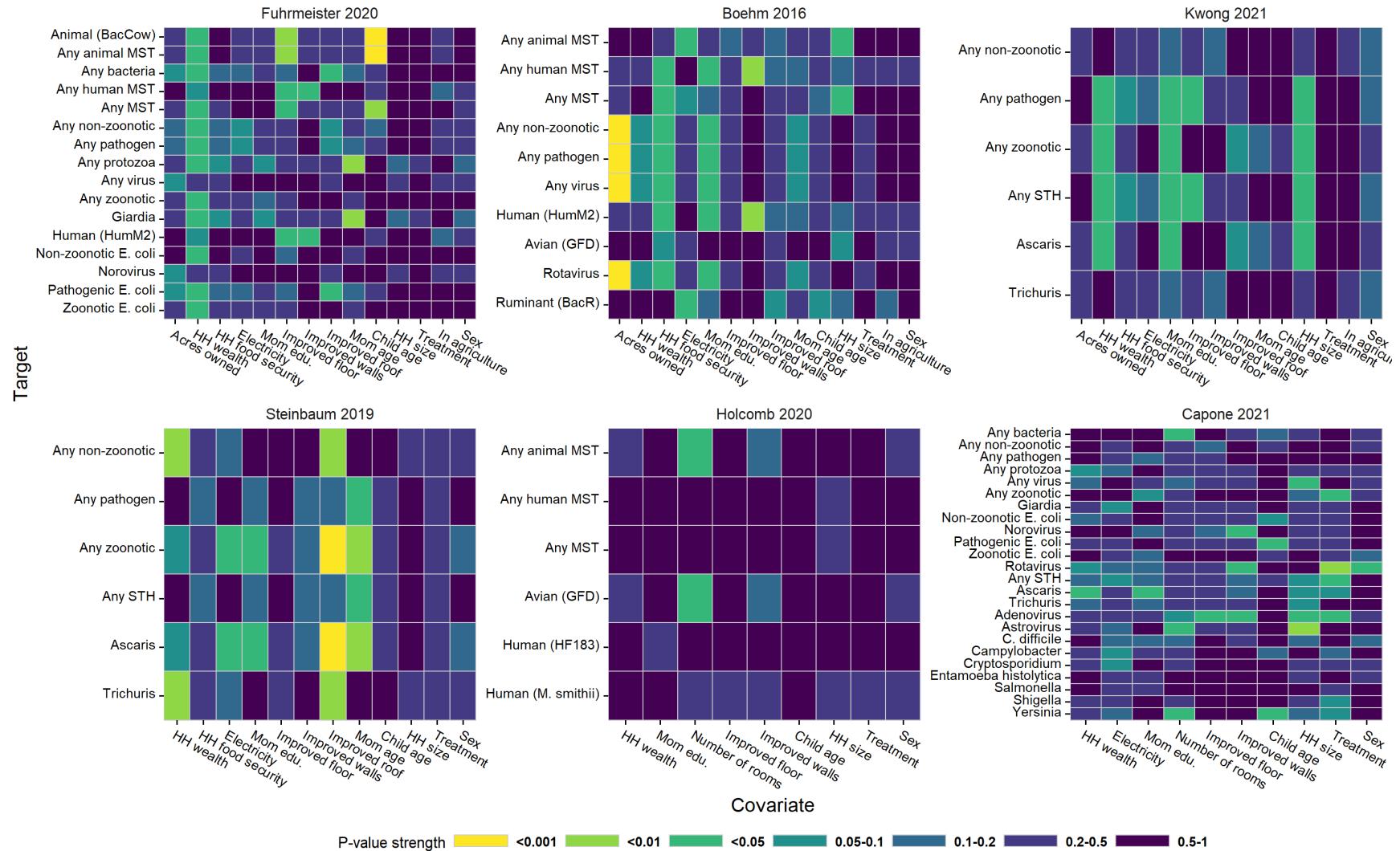


Figure S11. Forest plots of child diarrheal disease prevalence differences between environmental samples with and without any enteropathogen or any MST marker detected, stratified by whether the diarrheal disease occurred during the wet versus dry season (defined by the 6 months of highest average rainfall). Significant effect modification, as determined by the p-values on the regression model interaction term, is marked above points with asterisks ($P < 0.05 = ^*$, $P < 0.01 = ^{**}$, $P < 0.001 = ^{***}$).



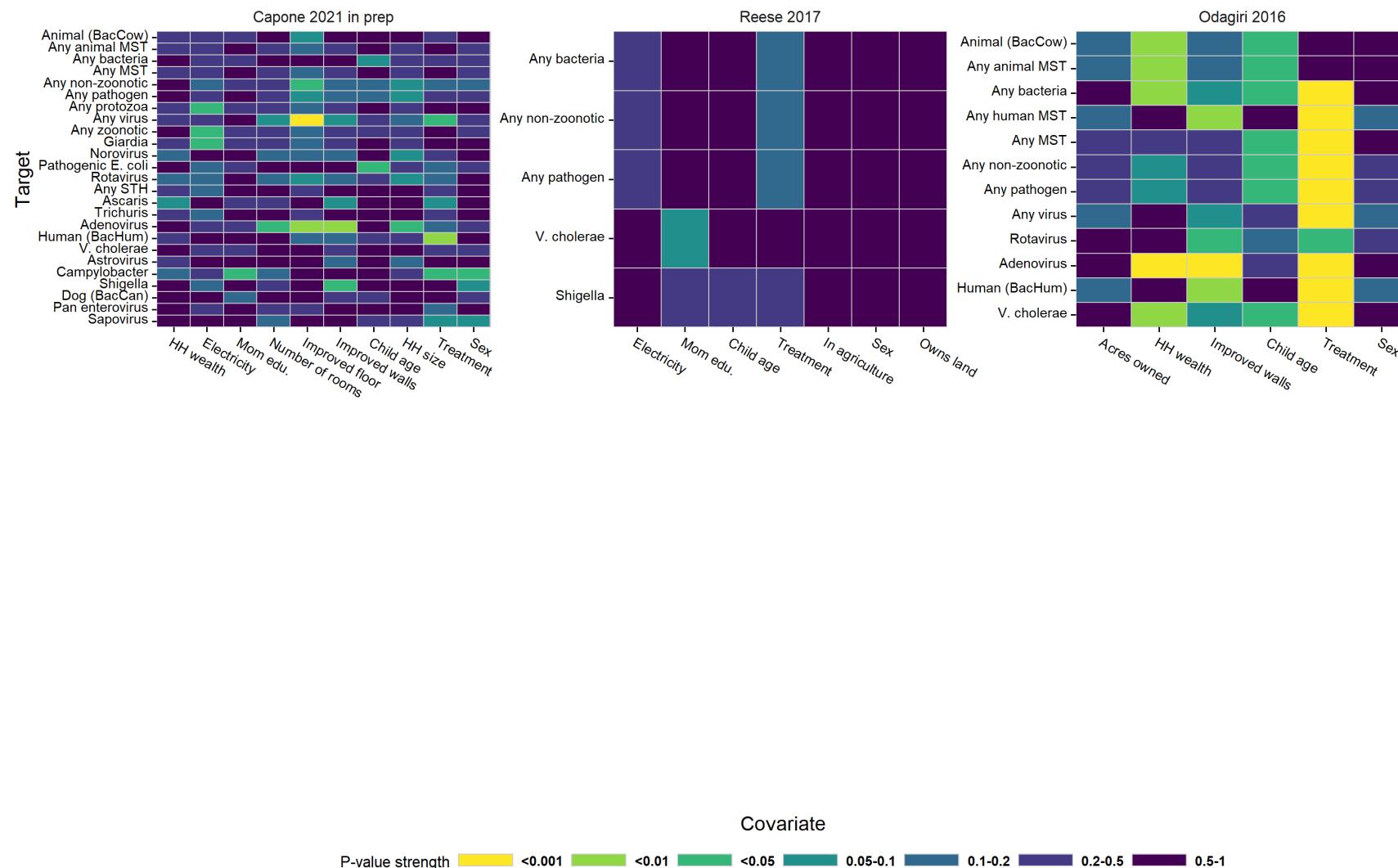


Figure S12. Study-specific associations between adjustment covariates and the presence of different enteropathogen and MST markers in aggregated environmental samples. The columns are different pre-screened confounders, and the rows are specific enteropathogens and MST markers. Cells of the heatmaps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

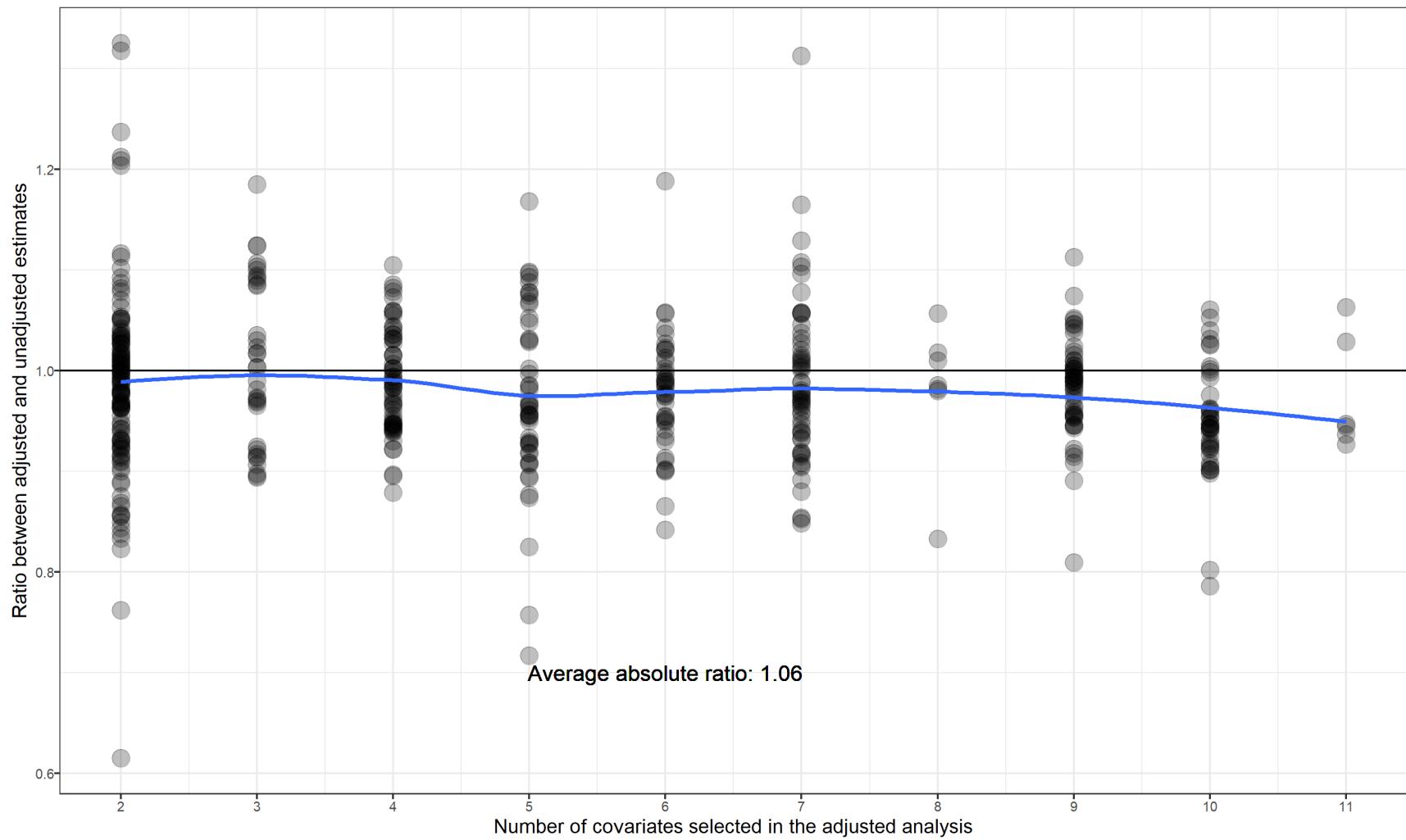


Figure 13. Comparison between associations estimated with and without including potential confounders for the binary diarrhea and growth outcomes. Points mark the ratio of relative risks estimated using adjusted and unadjusted generalized linear models. The blue line shows the average absolute ratio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

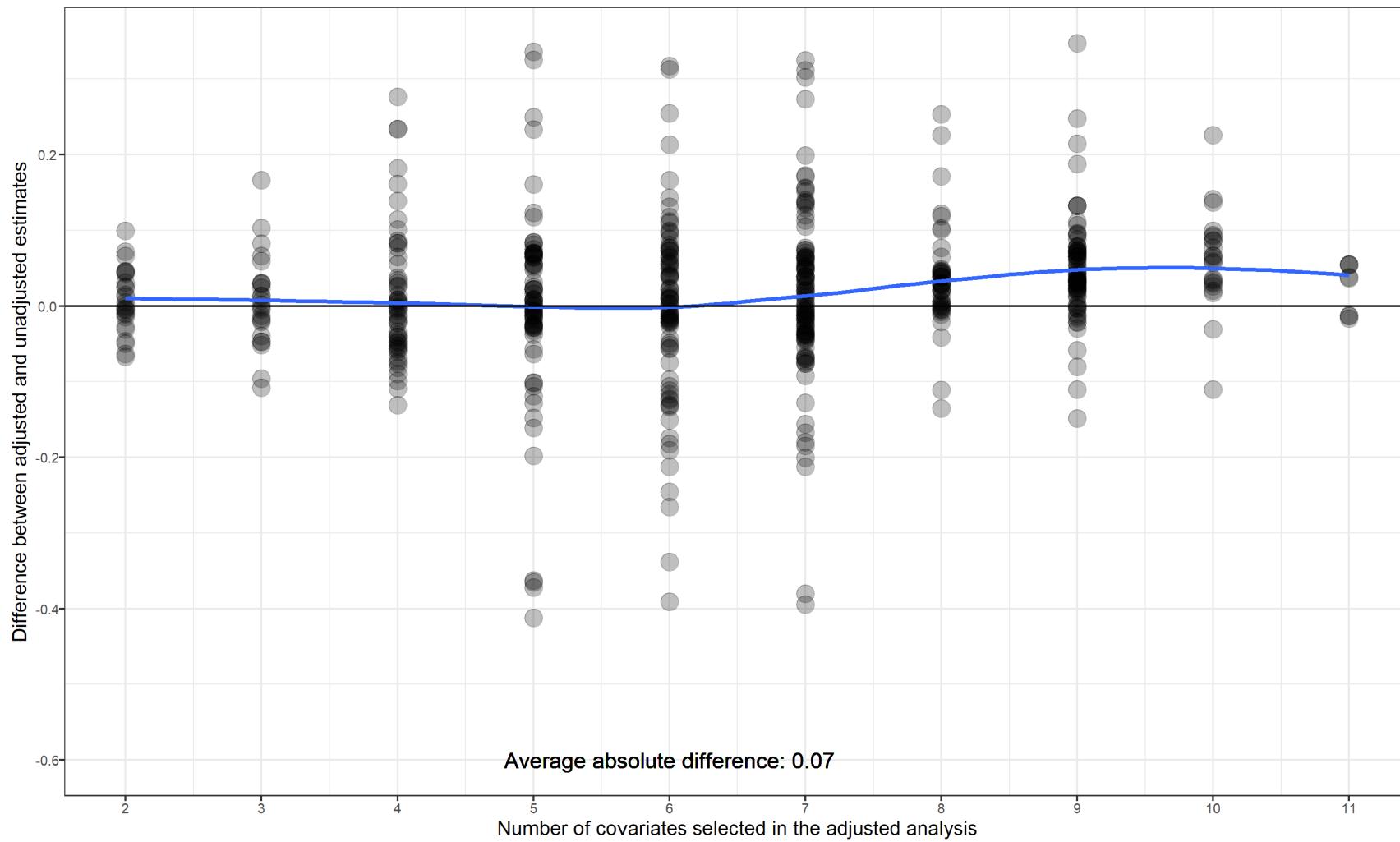


Figure 14. Comparison between associations estimated with and without including potential confounders for the continuous growth outcomes. Points mark the differences between mean differences estimated using adjusted and unadjusted generalized linear models. The blue line shows the average difference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

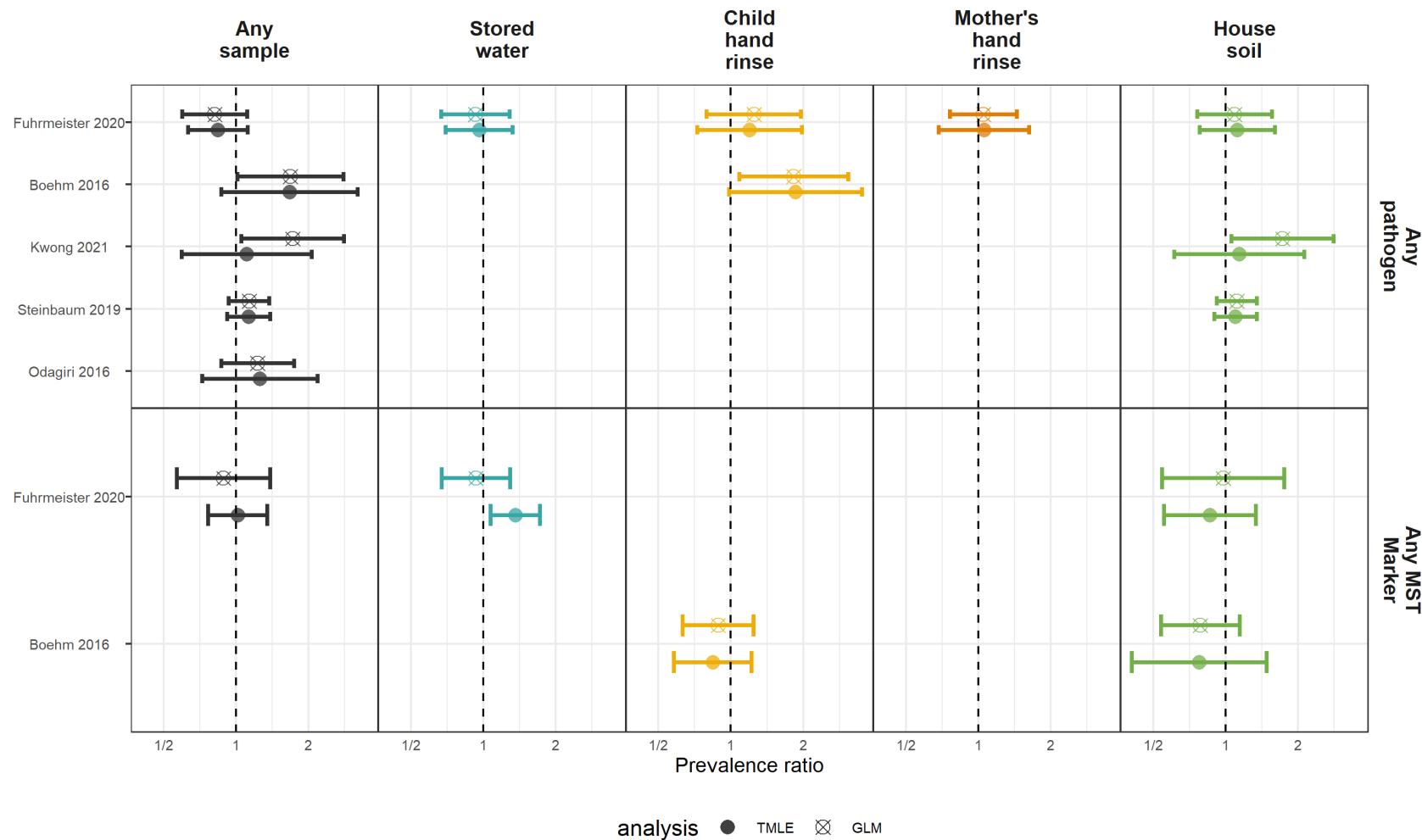


Figure S15. Comparison between associations estimated with generalized linear models (GLM) and machine-learning based targeted likelihood estimation models (TMLE) for the diarrhea outcome.

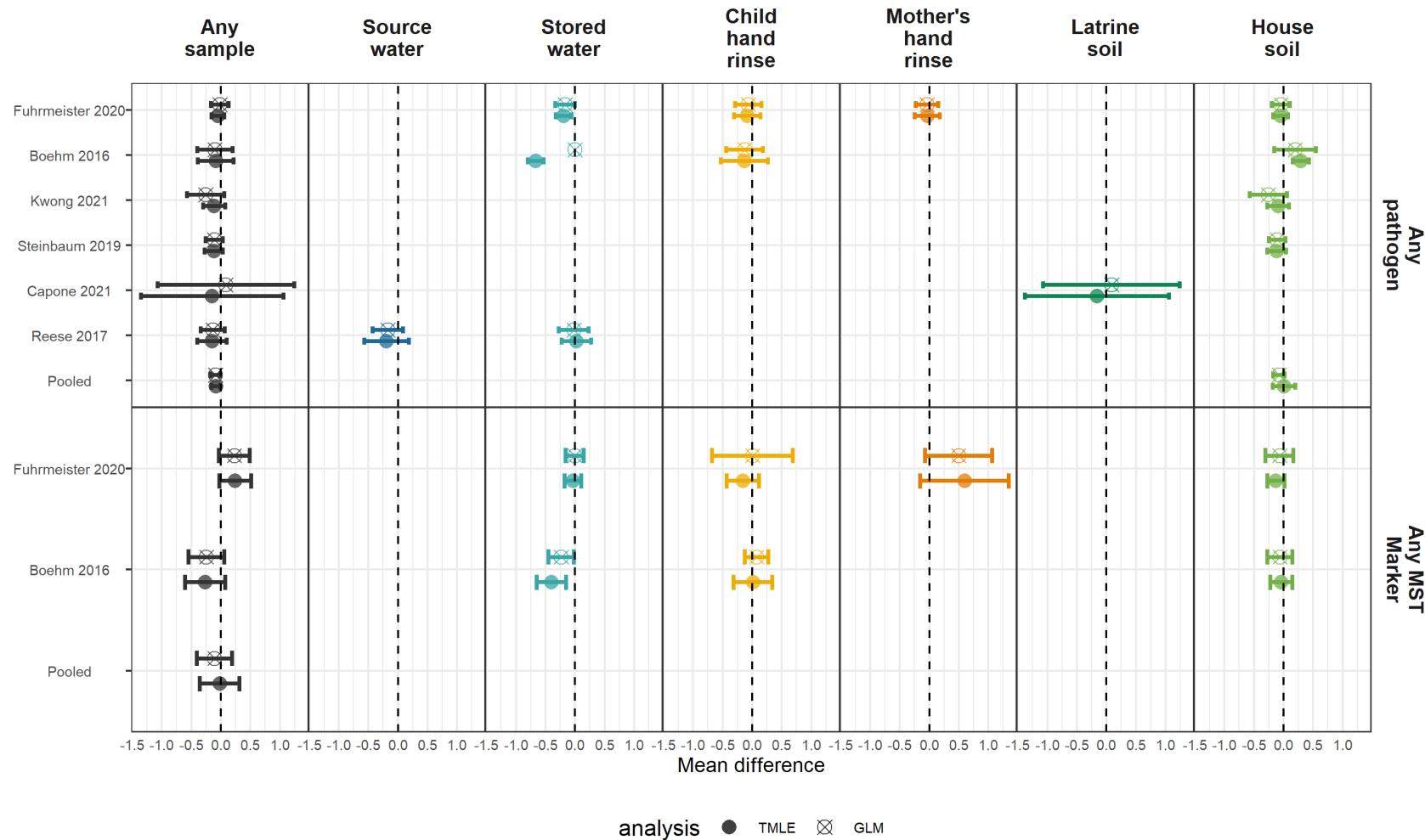


Figure S16. Comparison between associations estimated with generalized linear models (GLM) and machine-learning based targeted likelihood estimation models (TMLE) for the height-for-age (HAZ) Z-score outcome.

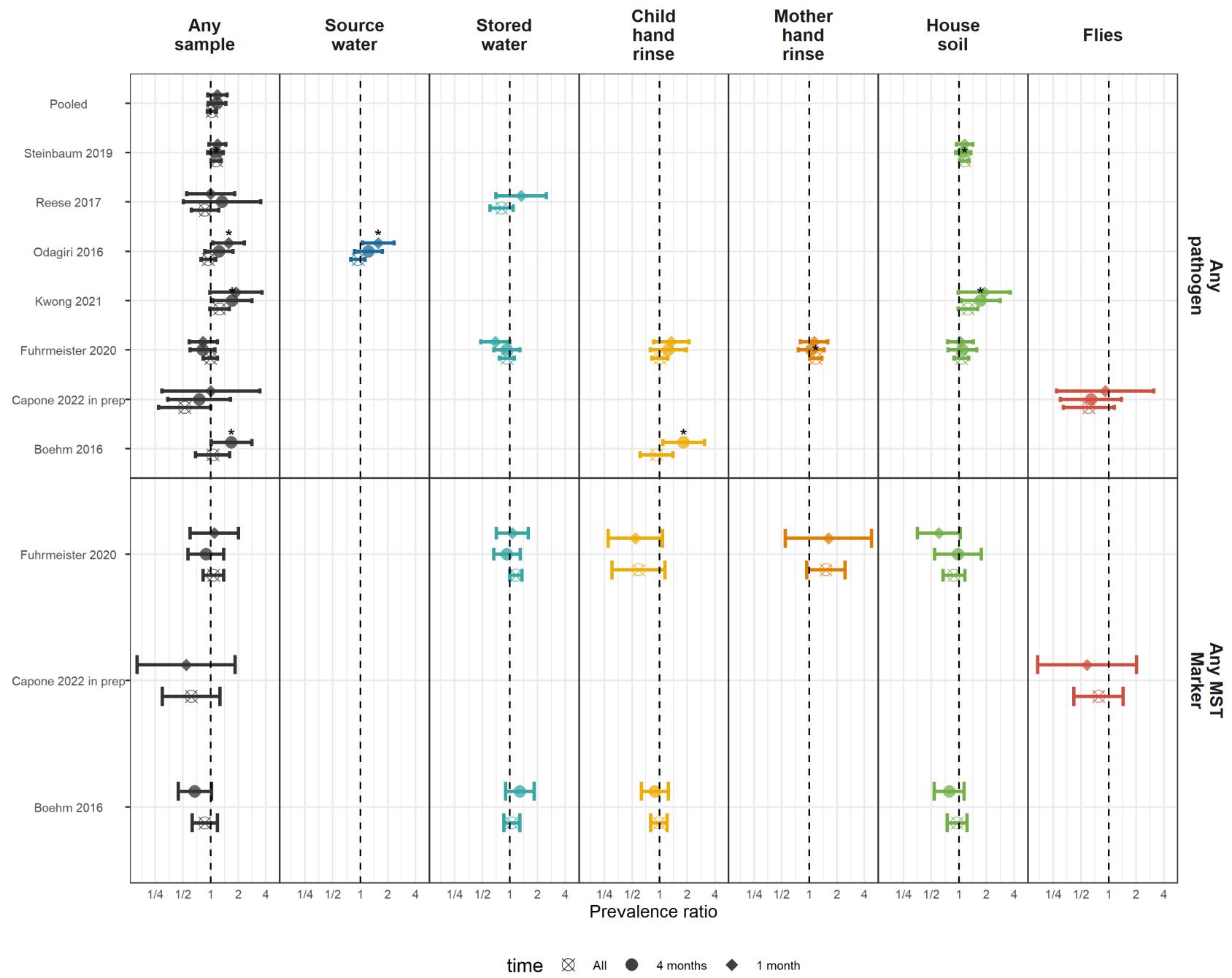


Figure S17. Comparison between associations estimated in the primary diarrhea analysis (diarrheal disease occurring after environmental sampling, but no more than 4 months later with associations estimated only using diarrheal disease cases within 1 month, or occurring at any time). For the analysis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

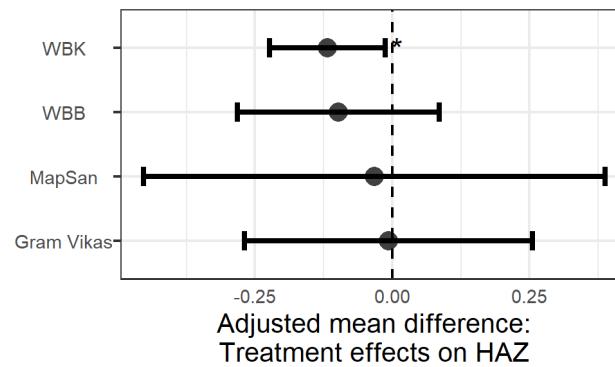


Figure S18. WASH intervention effects on child height-for-age Z-scores within the subset of children used in the primary analysis who had time-matched growth measurements and environmental samples.

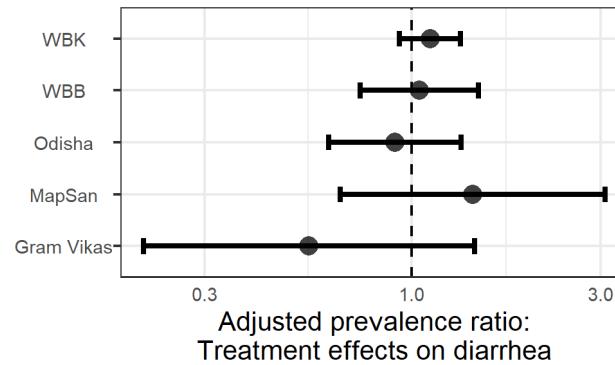


Figure S19. WASH intervention effects on child diarrheal disease within the subset of children used in the primary analysis who had time-matched diarrhea observations and environmental samples.

TABLES

Table 1. Descriptive statistics of child health outcomes by study. Pathogen-specific infection prevalence is the prevalence of at least one pathogen detected in child stool, and the number of pathogen infections is the total number of detected infections, where individual children can have infections from multiple pathogens.

Study	Trial	Distinct pathogens measured	# children with pathogens measured	# pathogen infections	Pathogen prev.	# diarrhe a obs.	# diarrhe a cases	Diarrh ea prev.	# HAZ obs.	Mean HAZ	Stunting prev.	# WAZ obs.	Mean WAZ	Underweight prev.	# WHZ obs.	Mean WHZ	Wasting prev.
Reese 2017	Gram Vikas					210	17	8.1	578	-1.78	42.2				576	-0.87	13.4
Holcomb 2021	Map San					227	20	8.8	232	-1.74	49.1	231	-0.64	10.8	228	0.21	7.0
Capone 2021	Map San	15	246	1,009	87.1	289	33	11.4	317	-1.55	40.7	321	-0.66	12.1	309	0.08	9.4
Capone 2022 in prep	Map San	10	255	803	82.2	244	27	11.1	291	-1.67	42.3	293	-0.69	14.3	280	0.14	7.1
Odagiri 2016	Odis ha					2,036	188	9.2				4,152	-1.38	29.1			
Fuhrmeister 2020	WBB	2	89	34	19.1	1,598	189	11.8	858	-1.81	40.9	872	-1.54	30.5	860	-0.85	10.0
Boehm 2016	WBB					412	99	24.0	411	-1.35	26.3	412	-1.35	24.3	412	-0.74	9.5
Kwong 2021	WBB	2	1,243	615	33.1	1,080	141	13.1	103	-1.58	30.1	103	-1.55	29.1	103	-0.97	8.7
Steinbaum 2019	WBK	2	1,609	338	20.6	1,912	496	25.9	1,800	-1.54	31.6	1,852	-0.73	9.7	1,797	0.10	1.5

HAZ: Height-for-age Z-score; WAZ: Weight-for-age Z-score; WHZ: Weight-for-height Z-score.

Table S1. PRISMA Checklist

(See separate attachment)

Table S2. Risk of bias based on modified Newcastle-Ottawa scale

Stars are given for low risk of bias in each category, up to a total of nine stars. Scoring details are in the footnotes.

Reference	Selection bias	Response bias	Follow-up bias	Misclassification bias	Outcome assessment	Outcome measurement	Bias in analysis	Total
	Is there evidence of selection bias, which refers to systematic differences between baseline characteristics of the groups that are compared? ^a	Is there evidence of response bias? ^b	Is there evidence of bias due to missing follow-up data? ^c	Is there risk of households not receiving the intervention being misclassified as having received it, or vice versa? ^d	Is there evidence of bias arising from how the outcome was assessed? ^e	Is there evidence of ascertainment bias? ^f	Is there evidence that analysis was not appropriately adjusted for clustering and/or confounding, if appropriate? ^g	Total number of stars (x/9 possible stars).
Clasen T, et al. Effectiveness of a rural sanitation programme on diarrhoea, soil-transmitted helminth infection, and child malnutrition in Odisha, India: a cluster-randomised trial. Lancet Glob Health. 2014.	*	possible (no blinding)	possible (86% of possible weeks are reported weeks)	* household-level interventions	caregiver recall for diarrhea, direct measurement for growth, and laboratory detection for pathogen-specific infections	possible (no blinding of assessor or person under study)	** adjusted for clustering	4
Luby, S.P. et al.. Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhoea and child growth in rural Bangladesh: a cluster randomised controlled trial. The Lancet Global Health 2018	*	include d negative control outcome, participants not blinded	* 94% complete FU	* household-level interventions	caregiver recall for diarrhea, direct measurement for growth, and laboratory detection for pathogen-specific infections	possible, data collector s not blinded (statistic al analysis blinded)	**	6
Null, C. et al., Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhoea and child growth in rural Kenya: a cluster-randomised controlled trial. The Lancet Global Health 2018	*	include d negative control outcome, participants not blinded	* <1% loss to FU	* household-level interventions	caregiver recall for diarrhea, direct measurement for growth, and laboratory detection for pathogen-specific infections	possible, data collector s not blinded (statistic al analysis blinded)	**	5

Reference	Selection bias	Response bias	Follow-up bias	Misclassification bias	Outcome assessment	Outcome measurement	Bias in analysis	Total
Reese, H. et al. Assessing longer-term effectiveness of a combined household-level piped water and sanitation intervention on child diarrhoea, acute respiratory infection, soil-transmitted helminth infection and nutritional status: a matched cohort study in rural Odisha, India. International journal of epidemiology 2019	selection bias is possible, as the study is not randomized and there are some baseline differences between intervention and control group	* no, assessed through negative control outcome	substantial loss to FU	* household-level interventions	caregiver recall for diarrhea, direct measurement for growth, and laboratory detection for pathogen-specific infections	possible (no blinding of assessor or person under study)	**	4
Knee, J. et al. Effects of an urban sanitation intervention on childhood enteric infection and diarrhea in Maputo, Mozambique: A controlled before-and-after trial. eLife 2011	selection bias is possible, as the study is not randomized, but intervention and control groups were mostly balanced at baseline. Control households were more likely to have covered floors and higher quality walls and intervention groups had more people per household.	possible (no blinding)	substantial loss to FU	* household-level interventions	caregiver recall for diarrhea, direct measurement for growth, and laboratory detection for pathogen-specific infections	possible (no blinding of assessor or person under study)	**	3

^a RCTs receive 1 star, unless evidence of selection bias (e.g. randomisation procedures not followed). Meaningful differences between groups at baseline in RCTs receive 0 stars. Rates of declining to participate >10% receive 0 stars. Non- or quasi-randomised studies receive 0 stars.

^b If intervention recipient was not blinded to intervention status, 0 stars.

^c <10% receives 1 star, greater than or equal to 10% receives 0 stars.

^d Interventions delivered at the household/individual level receive 1 star. Interventions delivered at the community level that missed a substantial, i.e. greater than or equal to 10%, proportion of the target population receive 0 stars, including when there is insufficient information to verify whether this is the case. Interventions with substantial risk of contamination (control households receiving intervention) receive 0 stars.

^e Parent / person recall (=0 stars). Fieldworker assessed (=1 star). Physician/microbiologically assessed (=2 stars)

^f If outcome measurement staff were not blinded to intervention status, 0 stars.

^g Scoring is based on losing stars (max. 2). Individual RCTs with baseline balance on covariates are unlikely to require adjustment (=2 stars). Cluster-RCTs and non-randomised trials may require adjustment for clustering (-1 star if not done). RCTs or cRCTs may require adjustment for covariates, with justification (-1 star if not done). Non-randomised studies require adjustment for covariates (-1 star if not done), but also adequate justification for covariate selection (-1 star if not included), and there can be too few or too many covariates.