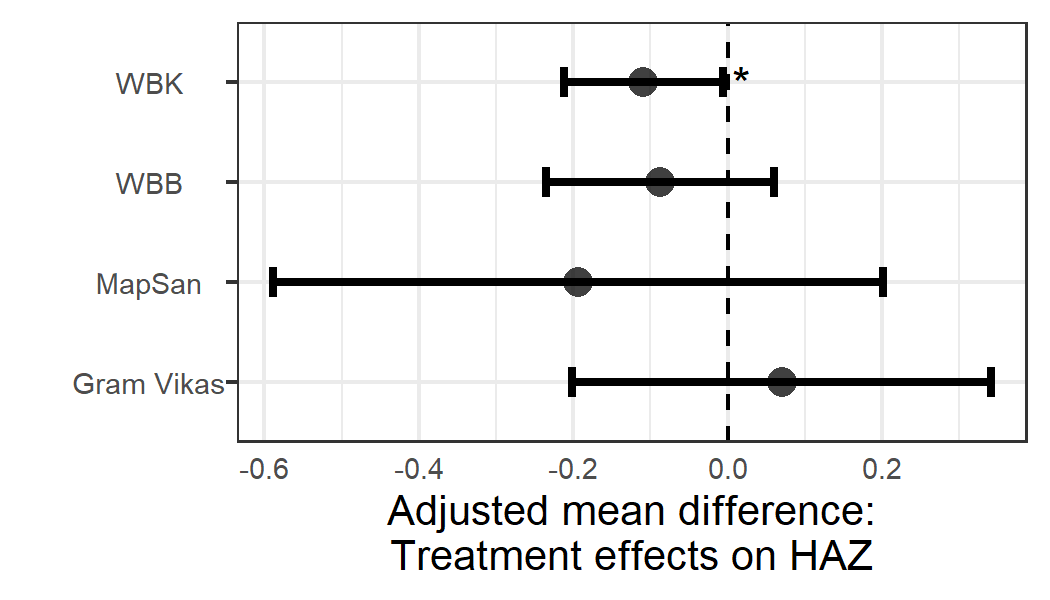
Supplementary Figures and Tables

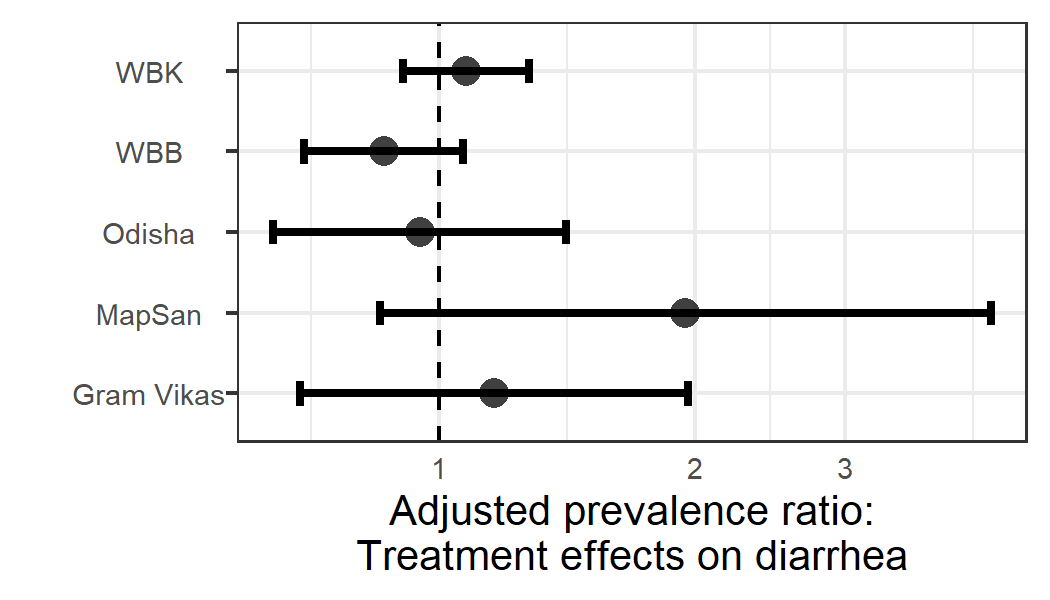
Associations between detection of enteropathogens and microbial source tracking markers in the environment and child enteric infections and growth: an individual participant data meta-analysis



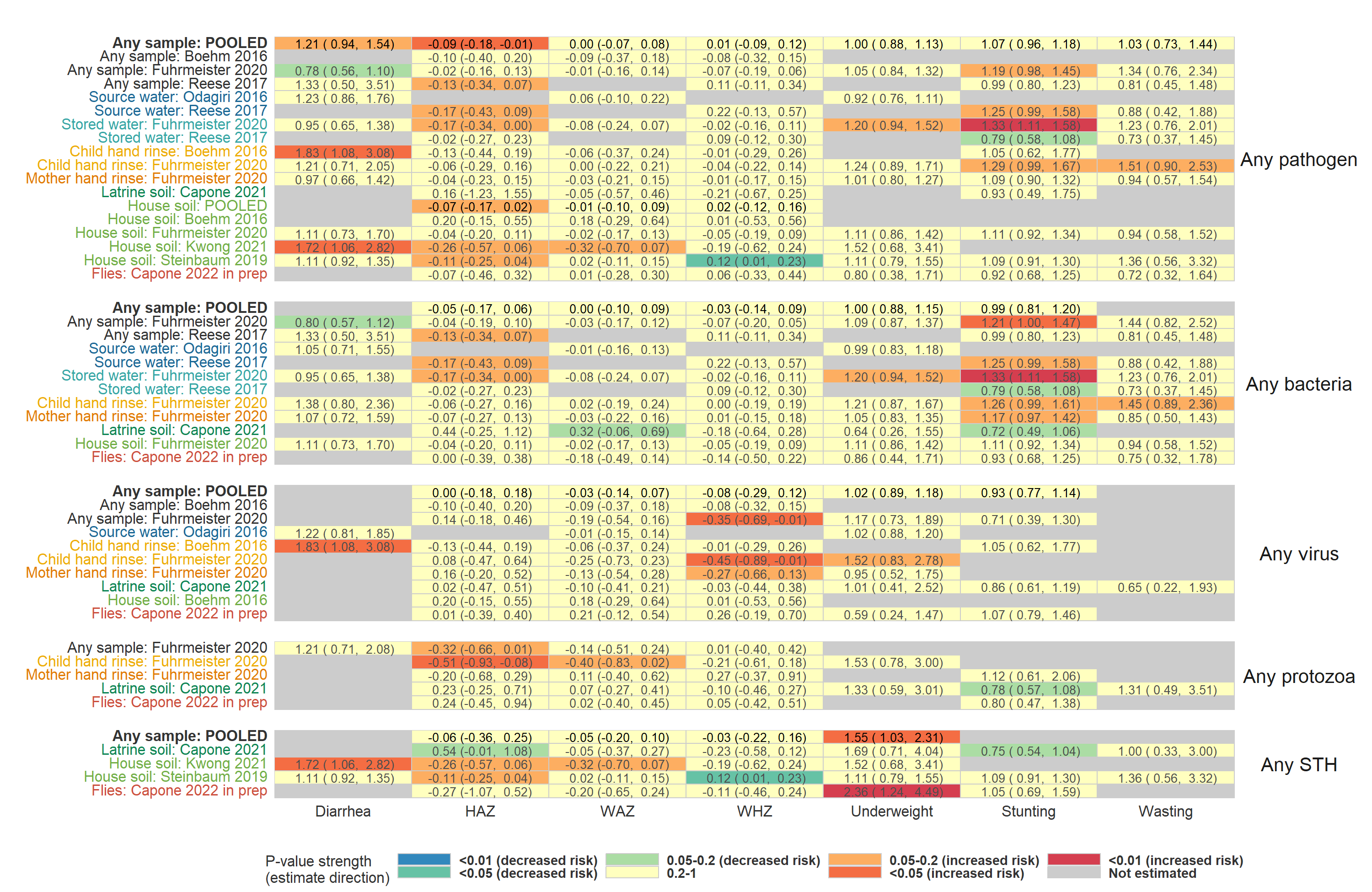
**Figure S1.** CONSORT flow diagram



**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.

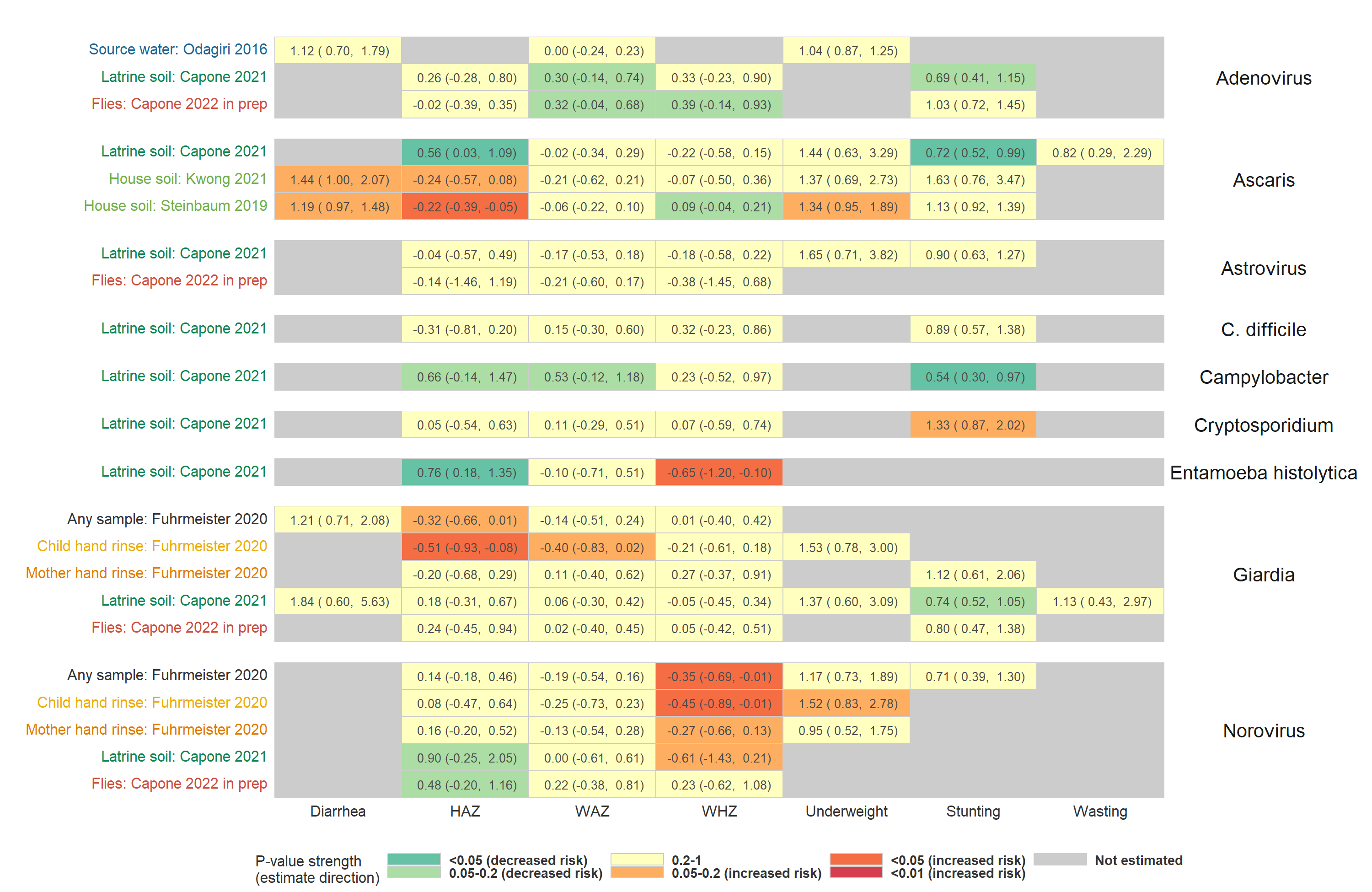


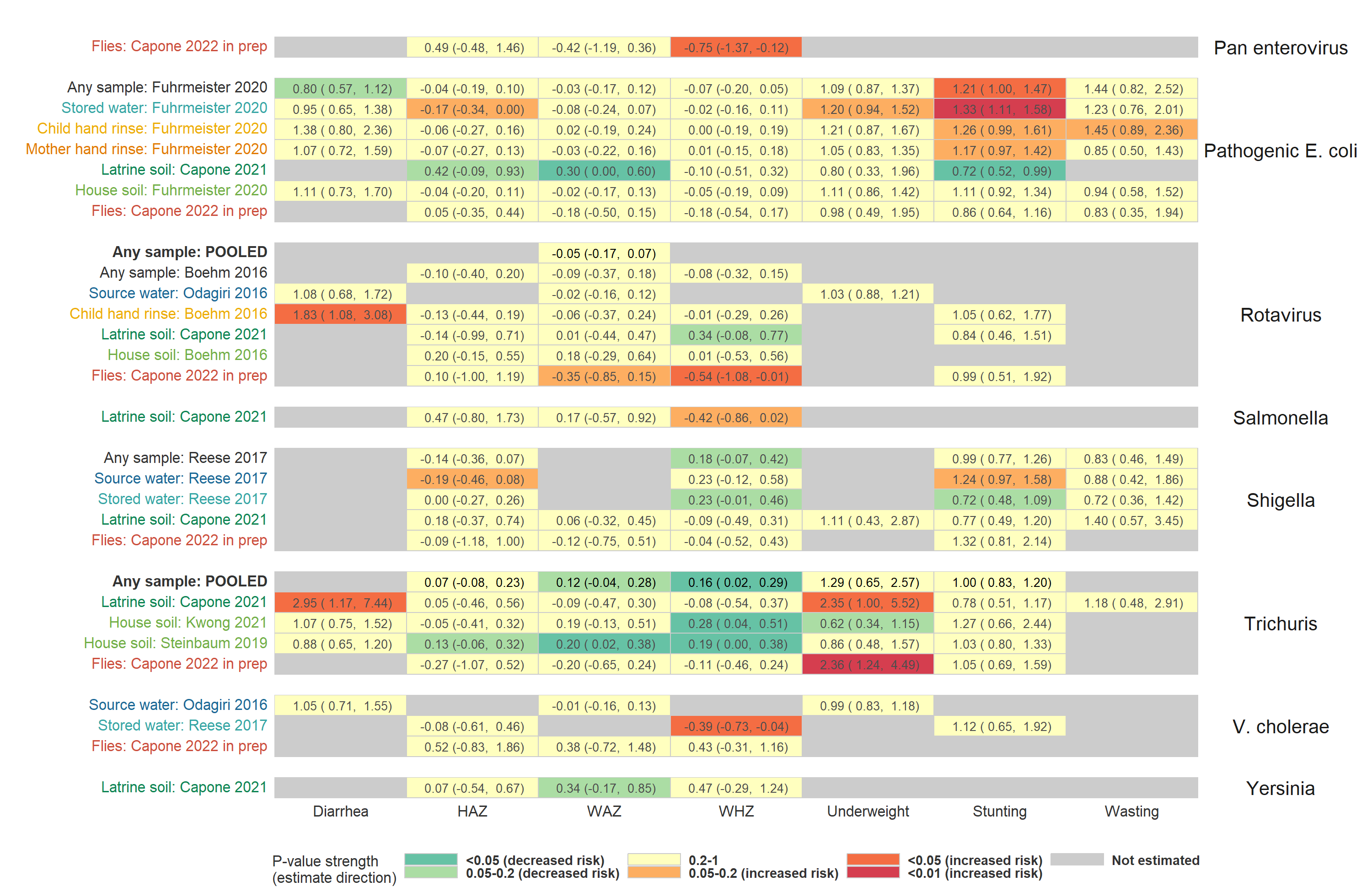
**Figure S3.** WASH intervention effects on child diarrhoeal disease within the subset of children used in the primary analysis who had time-matched diarrhoea observations and environmental samples.





**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.

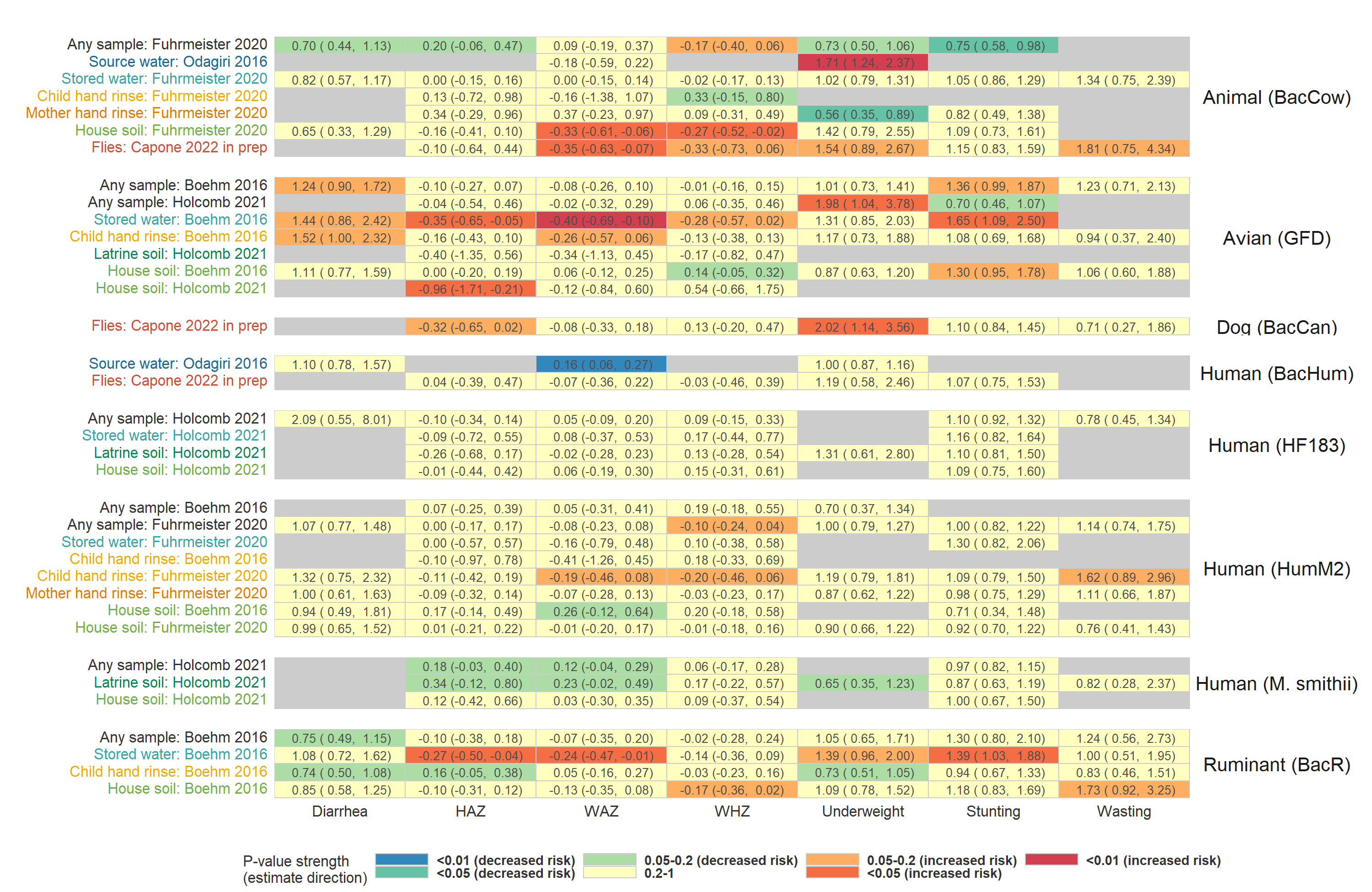




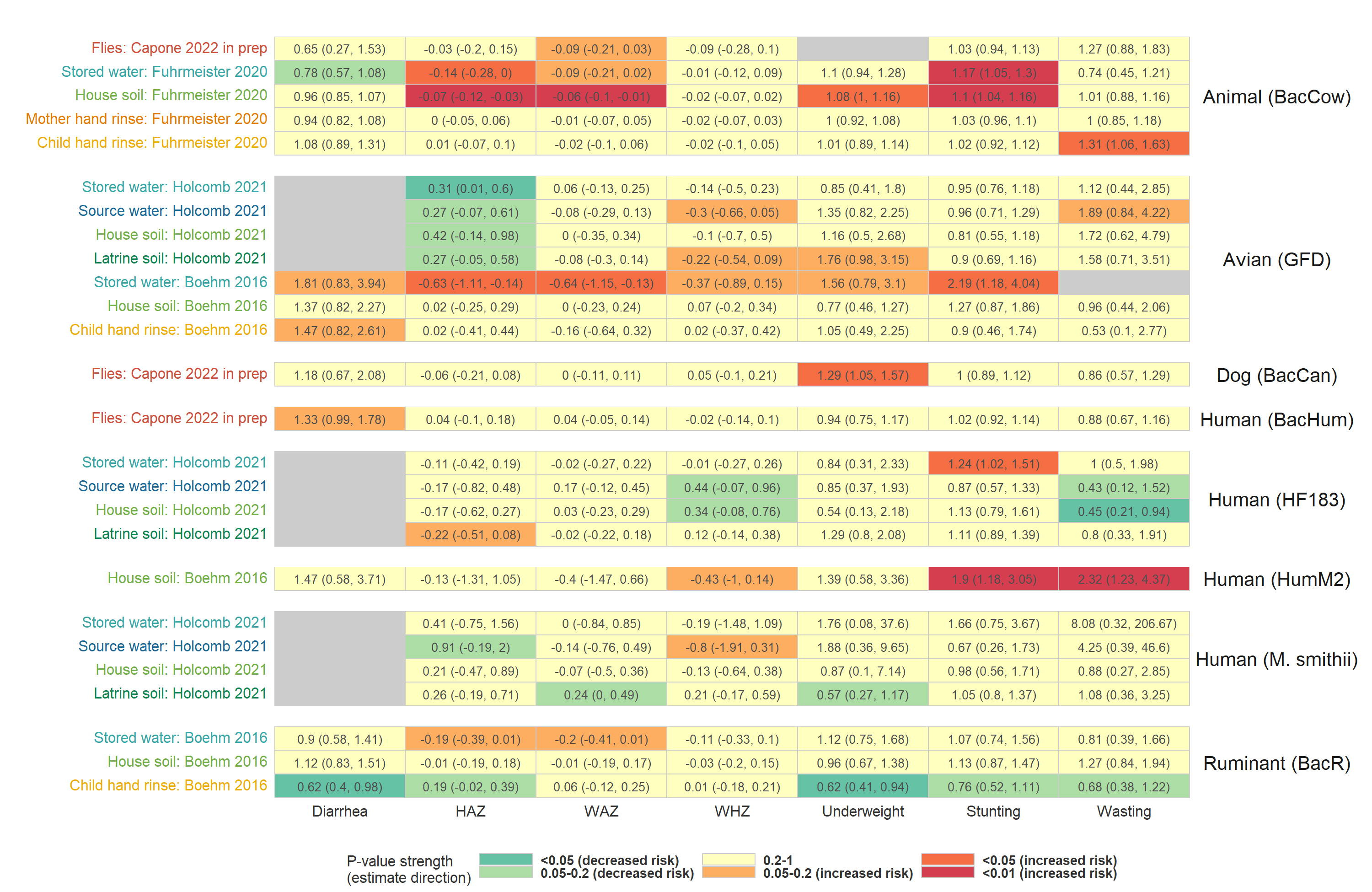
**Figure S5.** Heatmap of significance and direction of associations between specific pathogens in environmental samples (rows) and child diarrhoea and growth outcomes (columns). Cells are colored by the strength of significance and direction of association, and the point estimate and confidence intervals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



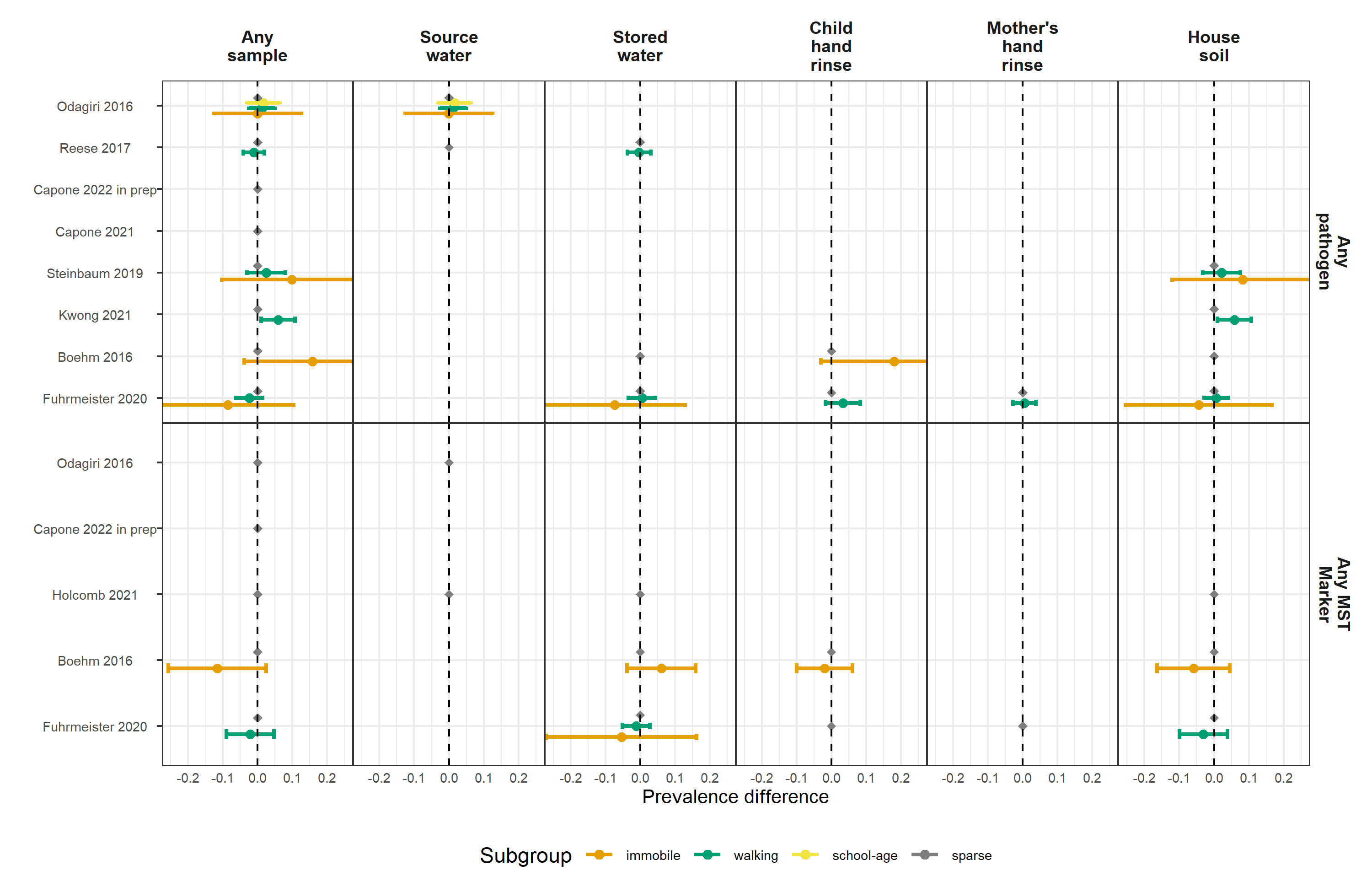
**Figure S6.** Heatmap of significance and direction of associations between the abundance of specific pathogens in environmental samples (rows) and child diarrhoea and growth outcomes (columns). Cells are colored by the strength of significance and direction of association, and the point estimate and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

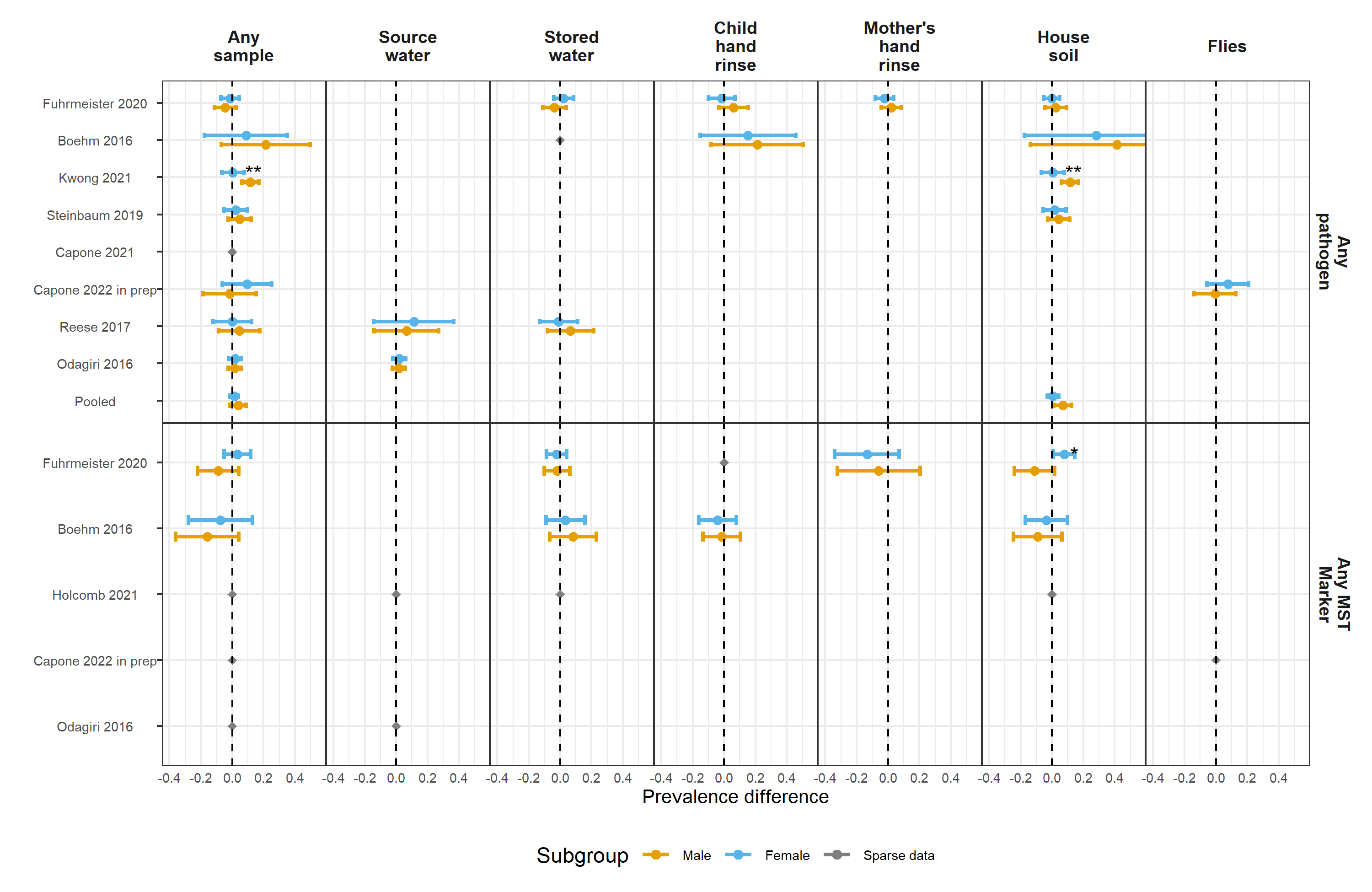


**Figure S9.** Forest plots of child diarrhoeal disease prevalence differences between environmental samples with and without any enteropathogen or any MST marker detected, stratified by child age. Grey points mark sparse age strata without estimated relative risks. Significant effect modification, as 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 = \*\*\*).

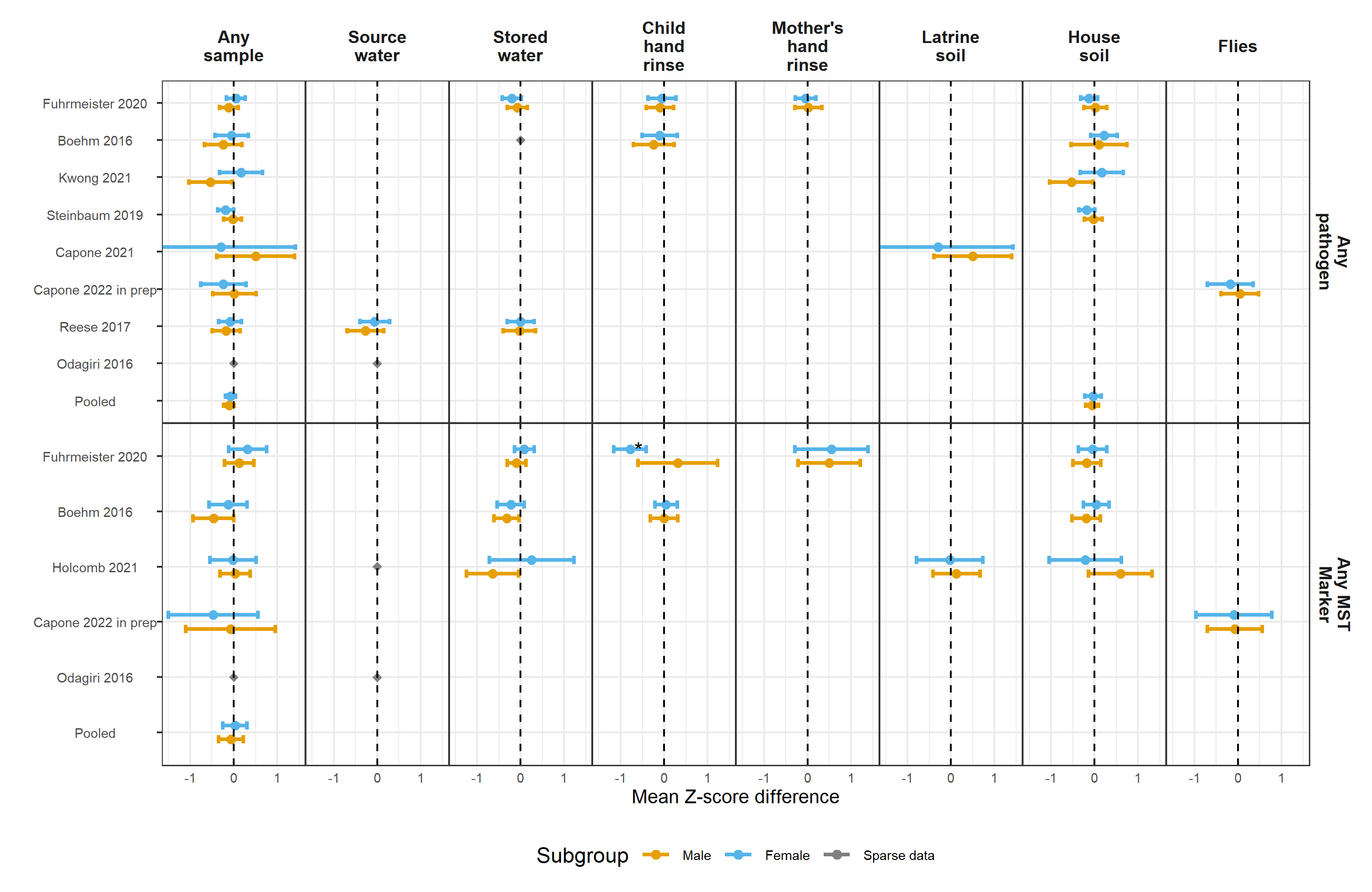
Chart, line chart, scatter chart, box and whisker chart

Description automatically generated

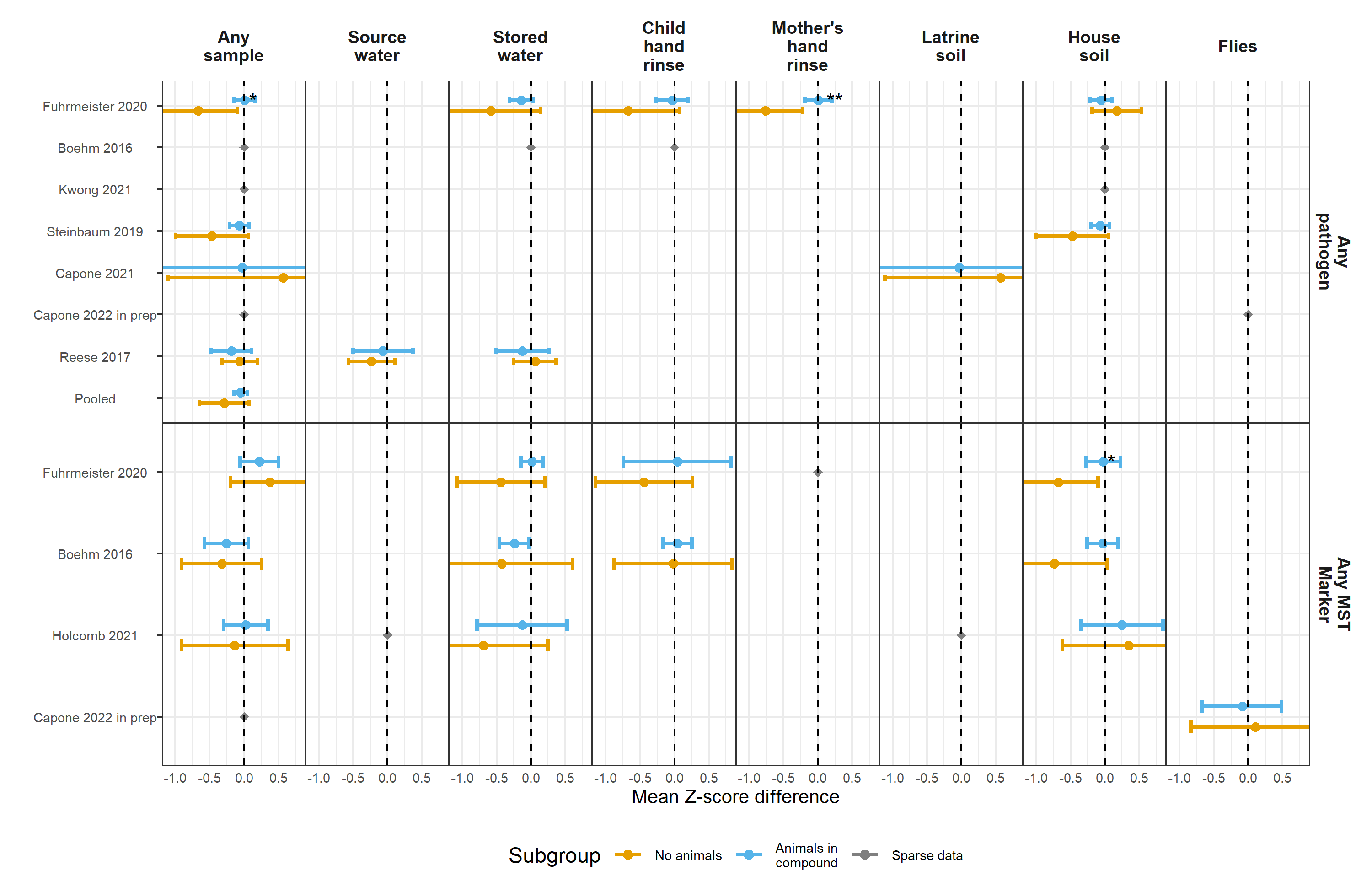
**Figure S10.** Forest plots of associations between any enteropathogen/any MST markers in different types of environmental samples and child height-for-age Z-score (HAZ), stratified by child age. Grey points mark sparse age strata without estimated mean differences. Significant effect modification, 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 = \*\*\*).



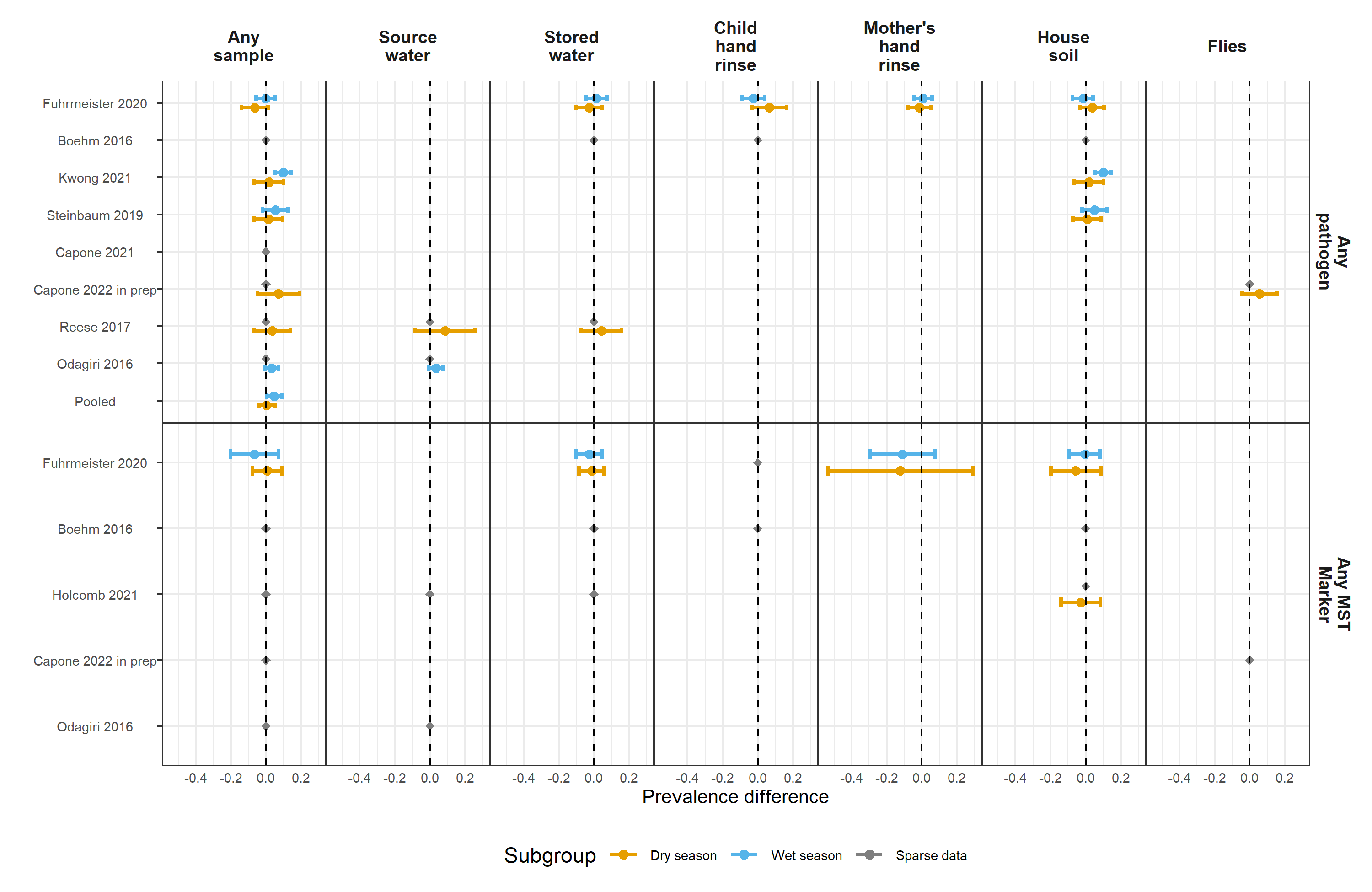
**Figure S11.** Forest plots of child diarrhoeal disease prevalence differences between environmental samples with and without any enteropathogen or any MST marker detected, stratified by child sex. Significant effect modification, as determined by the p-values on the regression model interaction term,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



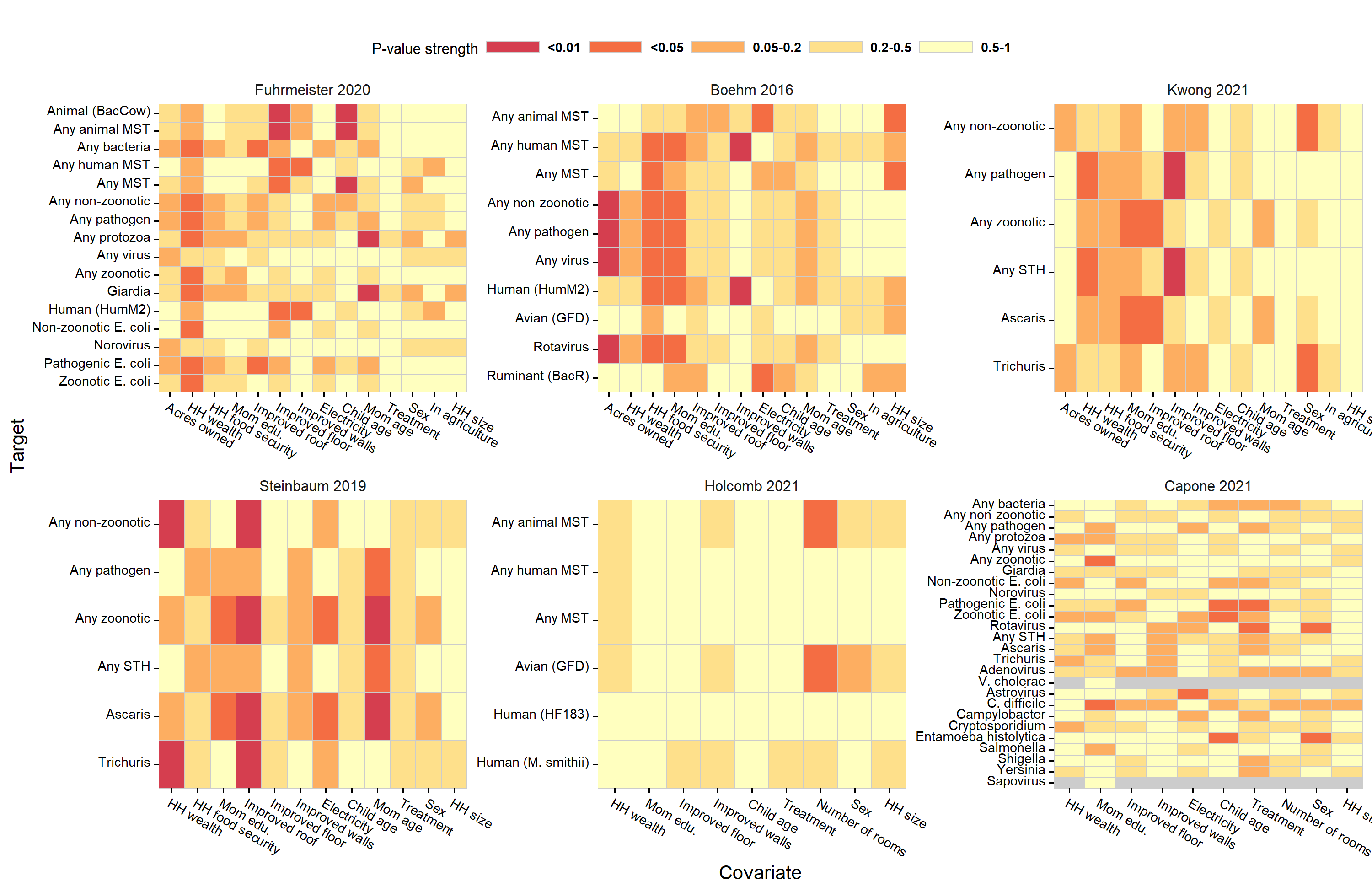
**Figure S12.** Forest plots of associations between any enteropathogen/any MST markers in different types of environmental samples and child height-for-age Z-scores (HAZ), stratified by child sex. Significant effect modification, as determined by the p-values on the regression model interaction term,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

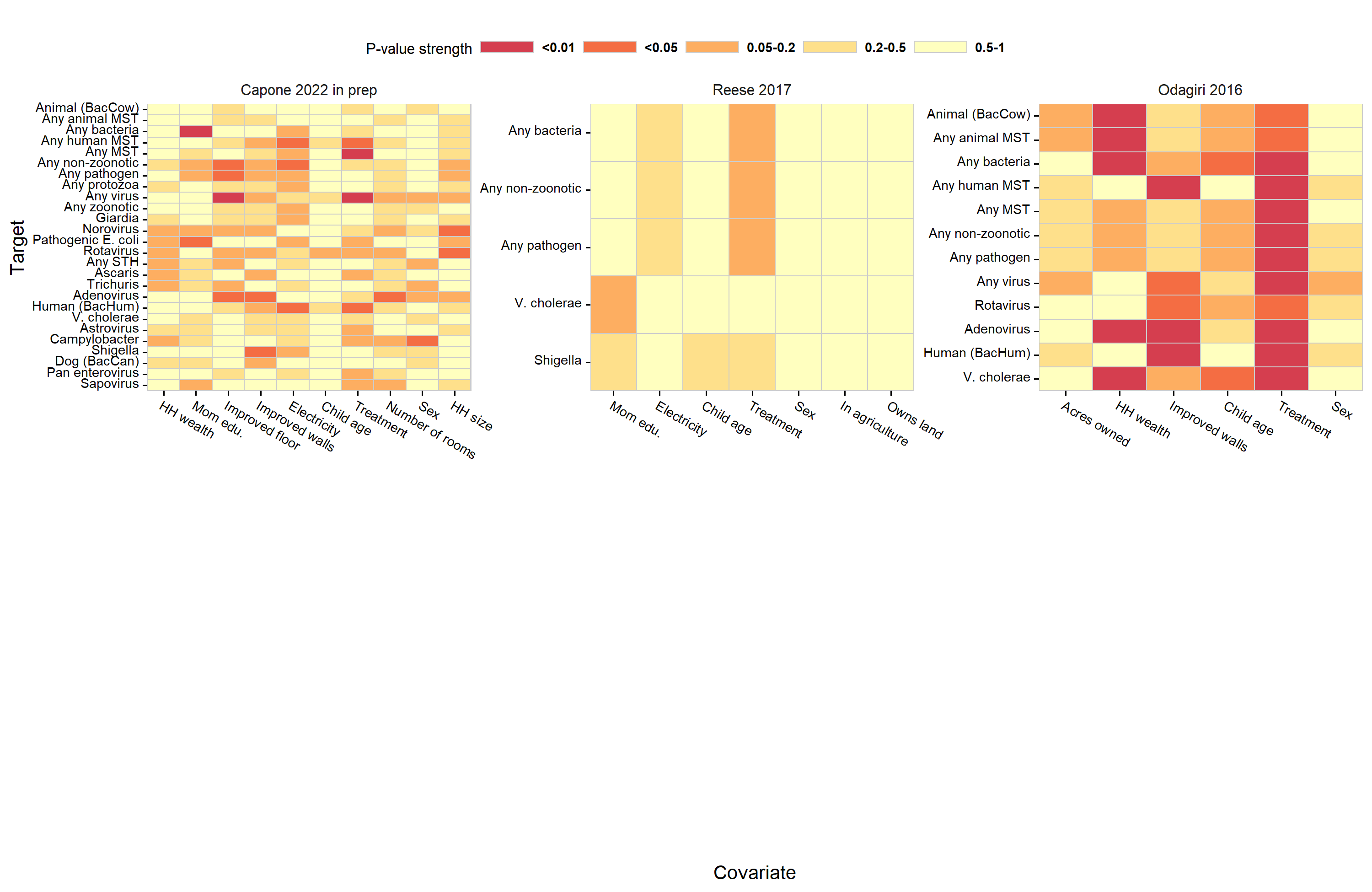


**Figure S13.** Forest plots of associations between any enteropathogen/any MST markers in different types of environmental samples and child height-for-age Z-scores (HAZ), stratified by whether any animals were present in the compound. Significant effect modification, as determined by the p-values on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

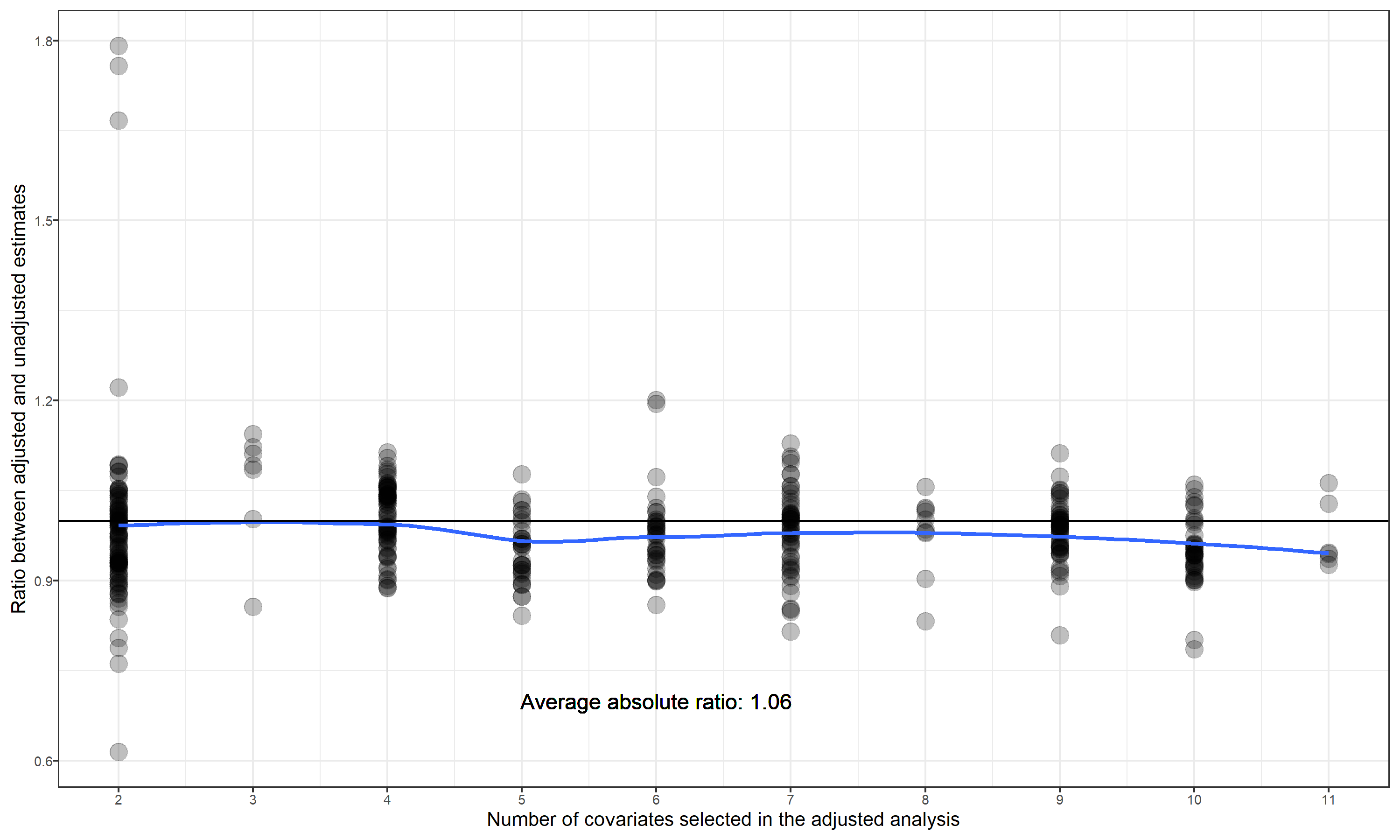


**Figure S14.** Forest plots of child diarrhoeal disease prevalence differences between environmental samples with and without any enteropathogen or any MST marker detected, stratified by whether the diarrhoeal disease occurred during the wet versus dry season (defined by the 6 months of highest average 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 = \*\*\*).

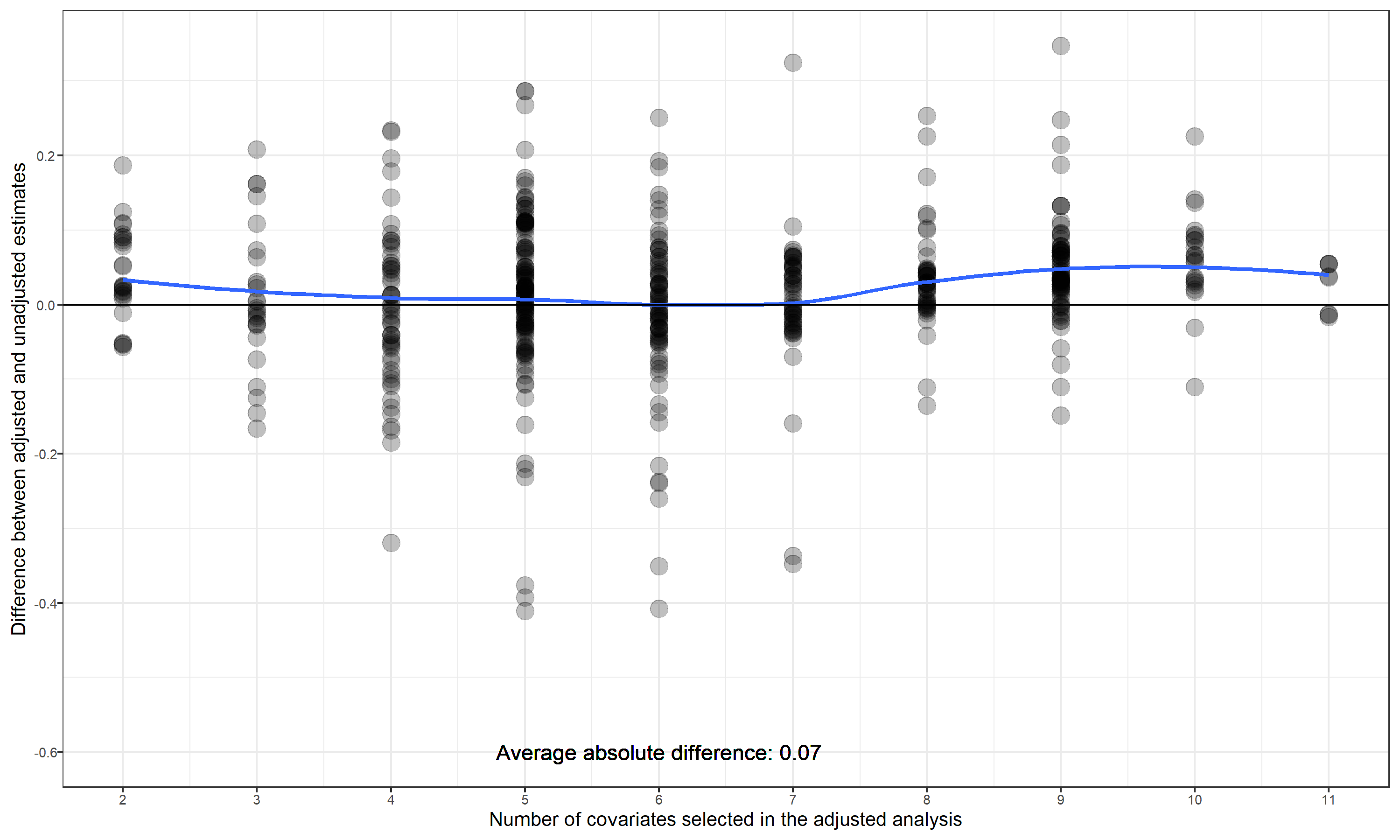




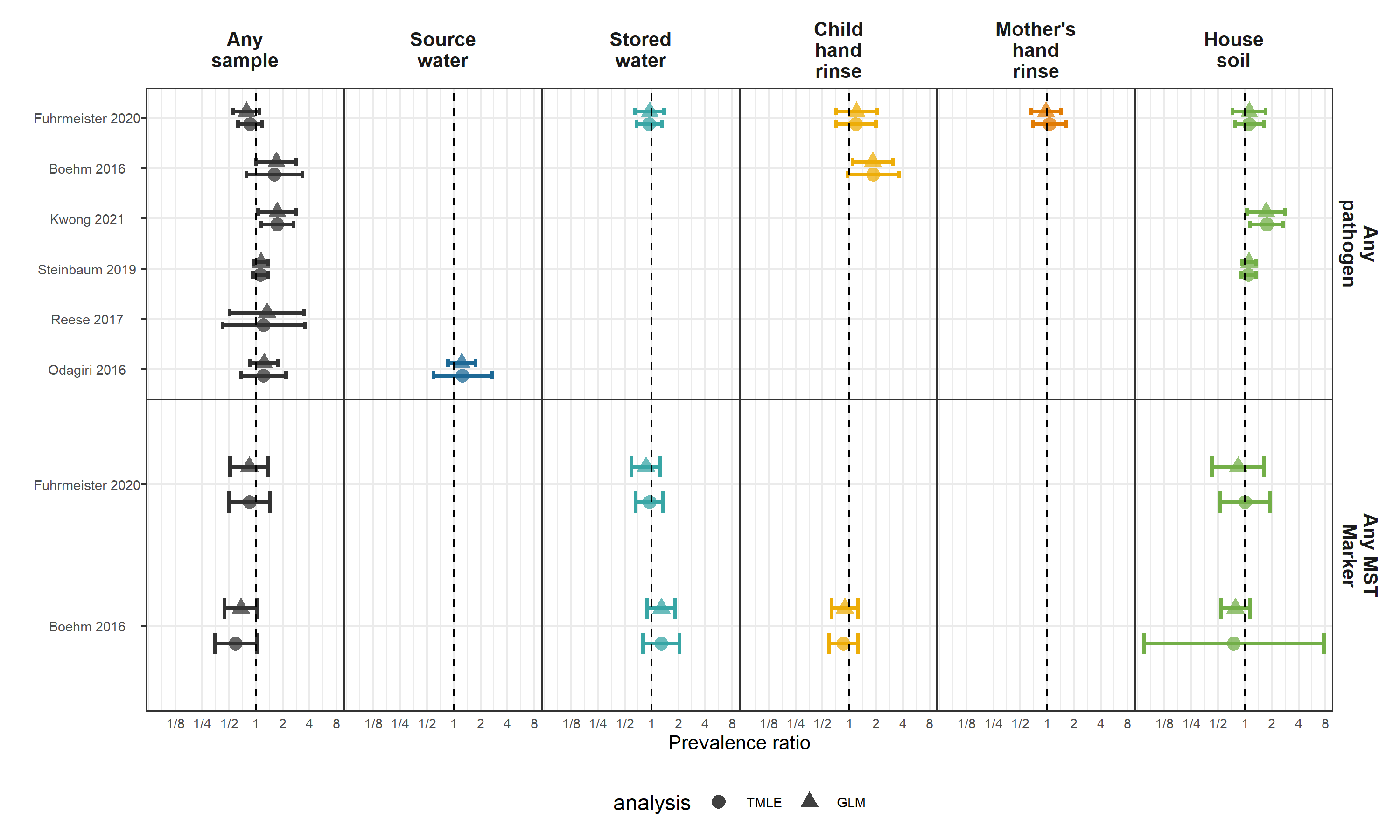
**Figure S15.** Study-specific associations between adjustment covariates and the presence of different enteropathogen and MST markers in aggregated environmental samples. The columns are different pre-screened confounders, and the rows are specific enteropathogens and MST markers. Cells of the heatmaps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



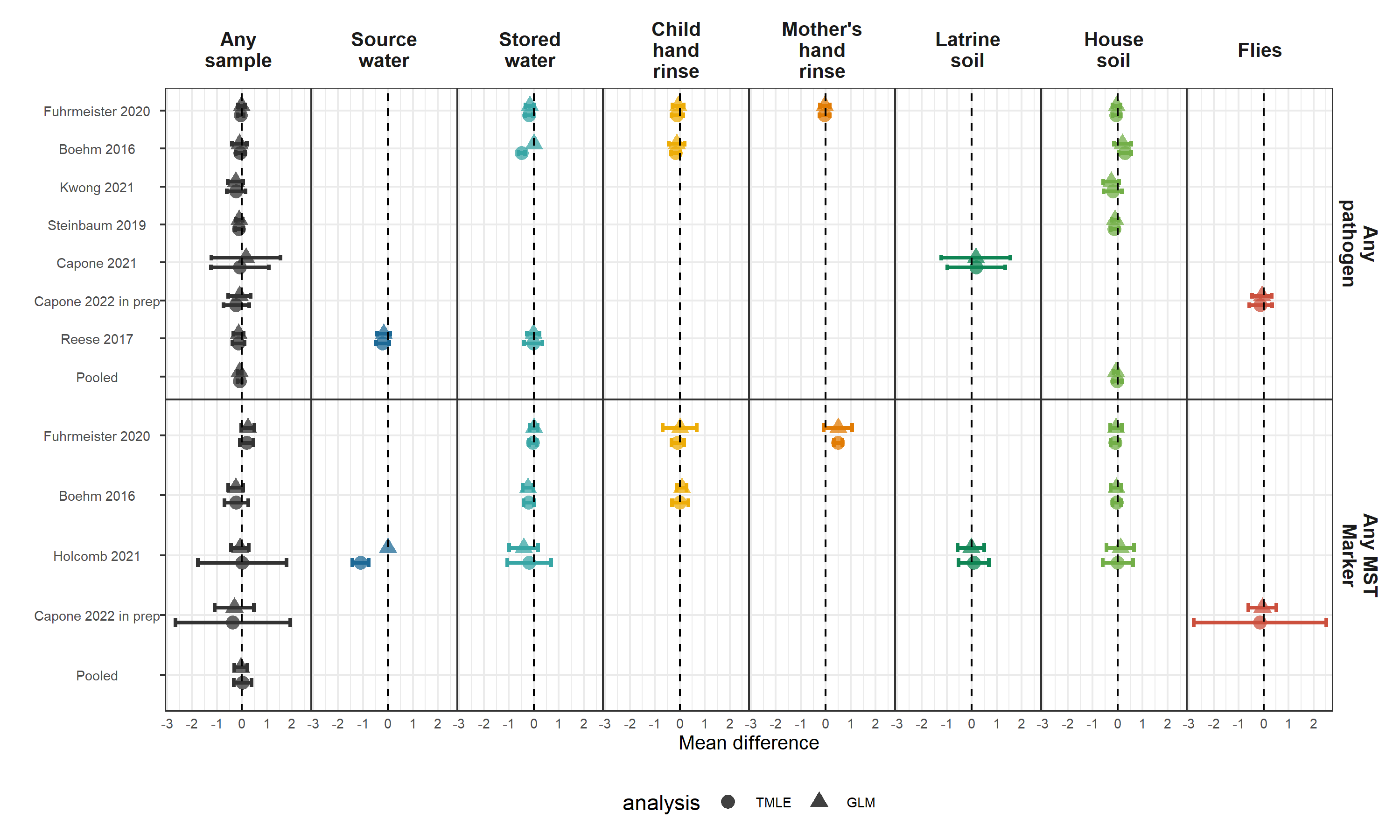
**Figure S16.** Comparison between associations estimated with and without including potential confounders for the binary diarrhoea and growth outcomes. Points mark the ratio of relative risks estimated using adjusted and unadjusted generalized linear models. The blue line shows the average ratio between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



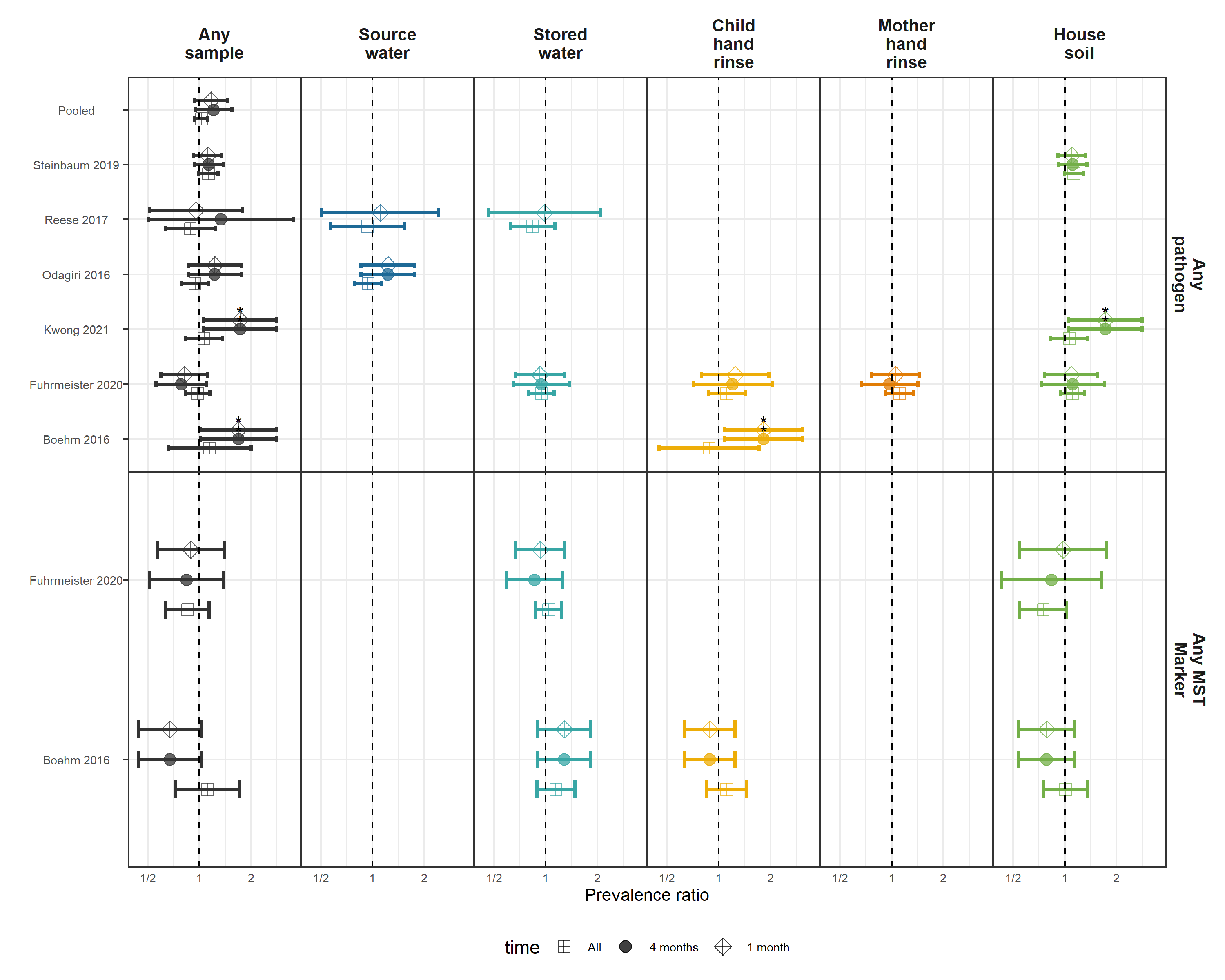
**Figure S17.** Comparison between associations estimated with and without including potential confounders for the continuous growth outcomes. Points mark the differences between mean differences estimated using adjusted and unadjusted generalized linear models. The blue line shows the average difference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

# Supplementary Tables

## Table S1. PRISMA Checklist

| **Topic** | **No.** | **Item** | **Location where item is reported** |
| --- | --- | --- | --- |
| **TITLE** |  |  |  |
| **Title** | 1 | Identify the report as a systematic review. | Page 1 |
| **ABSTRACT** |  |  |  |
| **Abstract** | 2 | See the PRISMA for Abstracts checklist below |  |
| **INTRODUCTION** |  |  |  |
| **Rationale** | 3 | Describe the rationale for the review in the context of existing knowledge. | Introduction, paragraph 1 |
| **Objectives** | 4 | Provide an explicit statement of the objective(s) or question(s) the review addresses. | Introduction, paragraph 1 |
| **METHODS** |  |  |  |
| **Eligibility criteria** | 5 | Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses. | Methods, paragraph 1 |
| **Information sources** | 6 | Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted. | Methods, paragraph 1 |
| **Search strategy** | 7 | Present the full search strategies for all databases, registers and websites, including any filters and limits used. | Methods, paragraph 1, citing related article. |
| **Selection process** | 8 | Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process. | Methods, paragraph 1, citing related article. |
| **Data collection process** | 9 | Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process. | Methods, paragraph 1, citing related article. |
| **Data items** | 10a | List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect. | Methods, paragraph 2 |
|  | 10b | List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information. | Methods, paragraph 1, citing related article. |
| **Study risk of bias assessment** | 11 | Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process. | Methods, paragraph 1, Table S2 |
| **Effect measures** | 12 | Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results. | Methods, paragraph 3 |
| **Synthesis methods** | 13a | Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item 5)). | Methods, paragraph 3 |
|  | 13b | Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions. | Methods, paragraph 3 |
| 13c | Describe any methods used to tabulate or visually display results of individual studies and syntheses. | Figure captions |
| 13d | Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used. | Methods, paragraph 3 |
| 13e | Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression). | Methods, paragraphs 4,5 |
| 13f | Describe any sensitivity analyses conducted to assess robustness of the synthesized results. | Methods, paragraph 5 |
| **Reporting bias assessment** | 14 | Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases). | Not applicable |
| **Certainty assessment** | 15 | Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome. | Not applicable |
| **RESULTS** |  |  |  |
| **Study selection** | 16a | Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram. | Figure S1 |
|  | 16b | Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded. | Results, paragraph 1 |
| **Study characteristics** | 17 | Cite each included study and present its characteristics. | Results, paragraph 1 |
| **Risk of bias in studies** | 18 | Present assessments of risk of bias for each included study. | Table S2 |
| **Results of individual studies** | 19 | For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots. | Figures 1 ,2 3, S2-S20 |
| **Results of syntheses** | 20a | For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies. | Not applicable |
|  | 20b | Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect. | Figures1 ,2 3, S2-S20 |
| 20c | Present results of all investigations of possible causes of heterogeneity among study results. | Results, Subgroup analyses section, and Figures S9-S14 |
| 20d | Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results. | Results, Sensitivity analyses section, and Figures S15-S20 |
| **Reporting biases** | 21 | Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed. | Not applicable |
| **Certainty of evidence** | 22 | Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed. | Figures 1 ,2 3, S2-S20 |
| **DISCUSSION** |  |  |  |
| **Discussion** | 23a | Provide a general interpretation of the results in the context of other evidence. | Discussion, paragraph 1 |
|  | 23b | Discuss any limitations of the evidence included in the review. | Discussion, paragraphs 2, 7 |
| 23c | Discuss any limitations of the review processes used. | Discussion, paragraph 2, cited related article |
| 23d | Discuss implications of the results for practice, policy, and future research. | Discussion, paragraph 5, 6, 8 |
| **OTHER INFORMATION** |  |  |  |
| **Registration and protocol** | 24a | Provide registration information for the review, including register name and registration number, or state that the review was not registered. | https://osf.io/8sgzn/ |
|  | 24b | Indicate where the review protocol can be accessed, or state that a protocol was not prepared. | https://osf.io/8sgzn/ |
| 24c | Describe and explain any amendments to information provided at registration or in the protocol. | Not applicable |
| **Support** | 25 | Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review. | Abstract, Funding |
| **Competing interests** | 26 | Declare any competing interests of review authors. | Not applicable |
| **Availability of data, code and other materials** | 27 | Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review. | https://github.com/amertens/wash-ipd |

# PRISMA for abstracts

| **Topic** | **No.** | **Item** | **Reported?** |
| --- | --- | --- | --- |
| **TITLE** |  |  |  |
| **Title** | 1 | Identify the report as a systematic review. | Yes |
| **BACKGROUND** |  |  |  |
| **Objectives** | 2 | Provide an explicit statement of the main objective(s) or question(s) the review addresses. | Yes |
| **METHODS** |  |  |  |
| **Eligibility criteria** | 3 | Specify the inclusion and exclusion criteria for the review. | Yes |
| **Information sources** | 4 | Specify the information sources (e.g. databases, registers) used to identify studies and the date when each was last searched. | No |
| **Risk of bias** | 5 | Specify the methods used to assess risk of bias in the included studies. | Yes |
| **Synthesis of results** | 6 | Specify the methods used to present and synthesize results. | Yes |
| **RESULTS** |  |  |  |
| **Included studies** | 7 | Give the total number of included studies and participants and summarise relevant characteristics of studies. | Yes |
| **Synthesis of results** | 8 | Present results for main outcomes, preferably indicating the number of included studies and participants for each. If meta-analysis was done, report the summary estimate and confidence/credible interval. If comparing groups, indicate the direction of the effect (i.e. which group is favoured). | Yes |
| **DISCUSSION** |  |  |  |
| **Limitations of evidence** | 9 | Provide a brief summary of the limitations of the evidence included in the review (e.g. study risk of bias, inconsistency and imprecision). | Yes |
| **Interpretation** | 10 | Provide a general interpretation of the results and important implications. | Yes |
| **OTHER** |  |  |  |
| **Funding** | 11 | Specify the primary source of funding for the review. | Yes |
| **Registration** | 12 | Provide the register name and registration number. | Yes |

## 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 diarrhoea, 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 | \* | \* included negative control outcome, participants not blinded | \* 94% complete FU | \* household-level interventions | caregiver recall for diarrhoea, direct measurement for growth, and laboratory detection for pathogen-specific infections | possible, data collectors not blinded (statistical 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 | \* | \* included negative control outcome, participants not blinded | \* <1% loss to FU | \* household-level interventions | caregiver recall for diarrhoea, direct measurement for growth, and laboratory detection for pathogen-specific infections | possible, data collectors not blinded (statistical analysis blinded) | \*\* | 5 |
| 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 diarrhoea, 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 diarrhoea 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 diarrhoea, 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.