Effect of water, sanitation and hygiene interventions on pathogens in the environment: Individual participant data meta-analysis

Andrew Mertens, Jack Colford, Oliver Cumming, Joe Brown, Jill Stewart, David Holcomb, 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, Ayse Ercumen (middle order not finalized)

# Abstract

TESTTESTTEST

# Background

Water, sanitation and hygiene (WASH) improvements aim to reduce childhood enteric infections, subsequent diarrhea disease and growth faltering by reducing child exposure to fecal-orally transmitted pathogens in the environment. Until recently, trials of WASH interventions have primarily focused on documenting health outcomes such as caregiver-reported diarrhea without quantifying intermediate outcomes along the causal chain, such as detection of pathogens in environmental samples and in human biological specimens. Without measuring these causal intermediates, trials are limited to a “black box” understanding, where underlying mechanisms of interventions are unknown and investigators can only speculate about reasons for intervention success or failure. Inspecting the causal chain is especially important given the small or null effects on child diarrhea and growth achieved in recent well-conducted trials of WASH interventions.1,2,**humphreyIndependentCombinedEffects2019a?**

Additionally, studies that have collected environmental samples to date have primarily focused on drinking water (and, to a smaller extent, hands and food) while other pathways such as soil and surfaces in the domestic environment have received less attention.3 Furthermore, most studies have relied on measuring indicator organisms in the environment as a proxy for pathogens; these indicators can originate from non-fecal sources,4 cannot differentiate between human vs. animal fecal contamination,5 and correlate poorly with the presence of pathogen.6 Recent advances in DNA-based diagnostics now allow detection of a range of enteropathogens in human biological specimens and environmental samples,7 as well as distinction between human vs. animal fecal sources through molecular source tracking (MST) markers.8

(Add a paragraph on existing studies on WASH intervention effects on MST markers that weren’t included)

We aimed to assess the effect of WASH interventions on specific pathogens, pathogen types (viruses, bacteria, protozoa, helminths) and human vs. animal MST markers in the environment. We conducted a systematic review of WASH intervention studies that have measured pathogens and/or MST markers in environmental samples, and we conducted an individual participant data (IPD) meta-analysis of WASH trials that have measured pathogens and/or MST markers in the environment to explore causal relationships between WASH interventions and pathogen and MST presence and abundance in the environment environment..

# Methods

We obtained individual participant data from WASH intervention studies that measured pathogens or MST markers in environmental sampples for an IPD analysis of the effects of the interventions on pathogen/MST marker prevalence and abundance in the environment. We first conducted a systematic literature search to identify WASH intervention trials and quasi-experimental (matched cohort or controlled before-and-after) studies that have measured pathogens and/or MST markers in environmental samples as well as at least one of the health outcomes of interest (caregiver-reported diarrhea, child growth or pathogens in stool). We developed a search strategy from a two-step process. First, the known studies meeting out inclusion criteria were examined for keywords and Medical Subject Heading (MeSH) terms relating to each of the following categories of terms comprising our search string: WASH interventions; microbial source tracking and environmental contamination; enteric infection; diarrhea; and child growth and development. Next, we performed an initial search using these terms and extracted other relevant terms and synonyms from relevant articles in the search results. Search terms are listed in Supplementary Table 1.

We queried the following databases for relevant studies: PubMed, Embase, CAB Direct Global Health, Agricultural & Environmental Science Database, Web Of Science, Scopus. We only included English language publications published in 2000 or after so that only studies with more recently developed pathogen detection methods were included. We included studies meeting the following inclusion criteria: 1) Randomized controlled trial or quasi-randomized study (i.e., matched cohort, controlled before-and-after study) of water, sanitation or hygiene intervention, 2) measure pathogens and/or MST markers in environmental samples, and 3) measured at least one health outcome of interest: pathogens in stool, self-reported diarrhea, or child anthropometry. The third criteria was needed to examine the hypothesized causal pathway from WASH intervention to child health through reduced environmental contamination, and the analysis is presented in a companion paper (Mertens et al. 2021).

From each identified study, we requested presence and abundance data from all pathogen and MST measures in environmental samples, as well as child health data, and potentially confounding baseline characteristics. To avoid sharing identifiable data, no GPS locations were shared, and the sampling dates were coarsened to a monthly resolution.

The primary outcomes analyzed were the prevalence of any enteropathogen, and of any general, human or animal MST markers in the environmental samples. Prevalences were analyzed by sample type (e.g., water, hands, soil, flies). Secondary outcomes included the prevalence of specific pathogen types (viruses, bacteria, protozoa, helminths), the prevalence and abundance of individual enteropathogens, and the prevalence and abundance of individual MST markers. For prevalence outcomes, we estimated prevalence ratios using modified poisson regressions. For abundance outcomes, we estimated differences of log-transformed abundances using linear regressions, and we estimated count differences of soil-transmitted helminth egg counts using negative binomial regressions.

We estimated prevalence ratios and abundance differences between intervention and control arms for each pathogen and microbial target by sample type, and also aggregated across samples collected from the same household during the same sampling round. We also estimated prevalence ratios and abundance differences between intervention and control arms for any pathogen presence, any MST marker, any bacteria, any virus, any protozoa, any helminth, any general MST, any human MST, or any animal MST.

All analyses were adjusting for potential confounders. While estimated effects of WASH interventions from randomized controlled trials should be unconfounded due to randomization, covariate adjustment may increase statistical efficiency, and improve exchangeability in comparing matched cohorts and non-randomized trials.9 Potential confounders were prescreened using likelihood ratio tests, and only variables associated with the outcome with a P-value < 0.2 were included in the model. We included the following variables in the prescreening set if they were measured within an included study: asset-based household wealth, number people in the household, number of rooms, household construction (walls, floor, roof), electrification, age and education of primary caregiver in the household, if anyone in the household works in agriculture, and land ownership. For rare binary outcomes, we only included one independent variable in the model (including the treatment arm) per 10 positive samples (or per 10 negative samples if >50% of samples were positive). We therefore did not estimate prevalence ratios for sample-target combinations with less than 10 positives or negatives and to avoid extreme estimated we did not estimate contrasts if there was less than two positives or negatives in the intervention or control arm. Because of repeated sampling or clustered designs in some studies, we used the Huber Sandwich Estimator to calculate robust standard errors for all estimates (Freedman 2006). We used the compound (group of households with shared sanitation) as the independent unit in non-cluster-randomized studies and the cluster in cluster-randomized trials. Additionally, we conducted subgroup analyses by animal ownership, season (dry vs. wet), study setting (rural vs. urban)

Due to differences in environmental and WASH conditions across studies, we reported study-specific estimates for all analyses. For analyses conducted using four or more studies, we tested for heterogeneity in estimates using Cochran’s Q-test.18,19 If there was not significant heterogeneity between estimates (P-value>0.2), we pooled estimates using fixed-effects models. If there is evidence for heterogeneity but there is qualitative support for pooling across studies, we pooled estimates using random-effects models.

All analyses were conducted in R 4.0, and analysis scripts are publically available (<https://github.com/amertens/wash-ipd>). The systematic review search stratigies and the analysis plan were pre-registered on Open Science Framework (<https://osf.io/8sgzn/>).

# Results

The systematic review was conducted on 1/19/2021 and returned 3,376 results after removing duplicates. 3,253 were excluded as irrelevant in the abstract screening stage, and of the 125 remaining studies, 15 met the inclusion criteria after full text screening by two authors. The 15 publications identified included multiple publications from the same intervention studies, and 6 unique intervention studies were identified by the systematic review.

1. The WASH Benefits Bangladesh trial 10
2. The WASH Benefits Kenya trial10
3. The MapSan trial in Mozambique11
4. The Gram Vikas matched cohort study in India12
5. The Total Sanitation Campaign trial in India13
6. The CHoBI7 Trial in Bangladesh (Add citation)

Data was shared by the primary investigators of the first 5 studies, but the authors of the CHoBI7 Trial declined to share the data.

There were 56018 total samples in the shared data across the 5 trials. Table 1 shows the specific pathogen prevalence by study and sample type. Table 2 shows the specific MST prevalence by study and sample type.

* (Note all interventions included sanitation, give details)
* Data from each study (Targets, methods of ascertainment, number of sample types, number of samples)
* Number of outcomes for each study, number of samples
* Prevalences for primary outcomes
* PR’s
* Abundance results (Table 3)
* Effect modification results

#### Results notes:

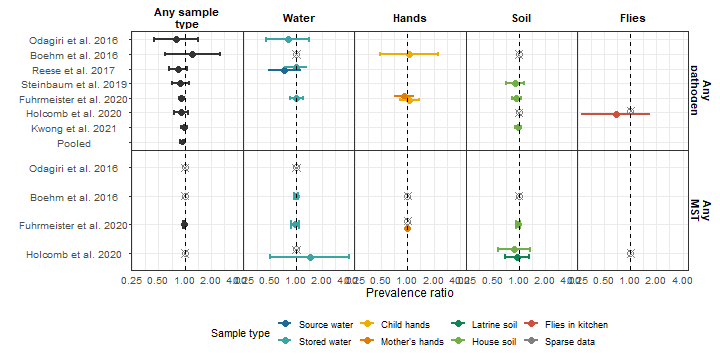
* Note food prep samples dropped from figures due to very few positives

# Discussion

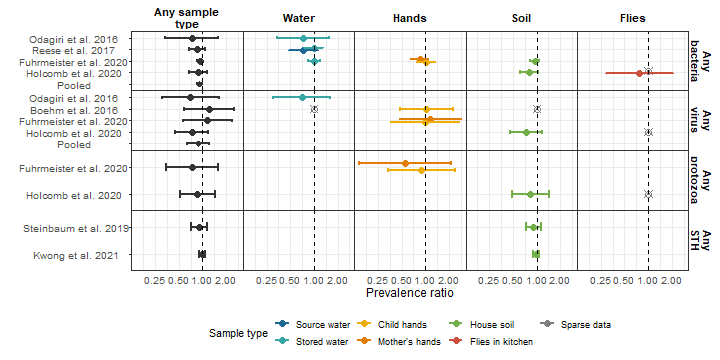
We obtained individual participant data from identified studies for an IPD analysis. Unlike conventional meta-analyses that statistically pool reported summary estimates from different studies, IPD analysis pools raw data from individual studies and then estimates the effect of interest.14 The IPD approach has several advantages over pooling independently derived effect estimates from different studies. While effect estimates are often reported differently across studies (e.g., odds ratio vs. relative risk), individual participant data allow direct derivation of the effect estimate of interest, regardless of how it was reported in the original studies. The statistical analysis approach can also be standardized across studies, including adjusting for baseline factors consistently and using consistent statistical models. Additionally, the IPD approach allows estimating effects in subgroups of participants and exploration of factors that mediate any intervention/treatment effects.15-17

* Discuss significant results and any consistencies across studies.
* Discuss differences in interventions across studies
* Differences in adjusted vs. unadjusted estimates
* Effect modification
  + In the analysis plan, we say we’ll conduct a subgroup analysis of pathogens with human only vs. human and animal hosts, but this isn’t a seperate analysis, just interpretation of the results (because we didn’t have enough to pool), so discuss here
* Limitations \* Sparse in many categories \* Abundance imputation \* Look at primary manucript discussions
* Future research needs?

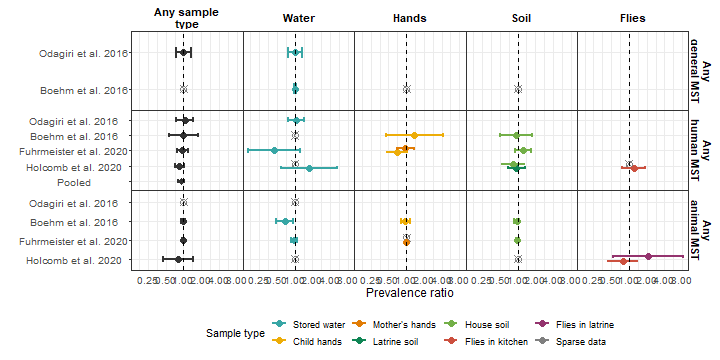
# Figures



**Figure 1.** Forest plots of intervention effects on any enteropathogen, and 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 and target combination. Empty, grey, crossed points denote contrasts where data existed but with either too low or too high a prevalence to estimate a prevalence ratio. All estimated are adjusted for potential confounders.



**Figure 2.** Forest plots of intervention effects on any virus, any bacteria, any protozoa and any STH prevalences in different types of environmental samples.



**Figure 3.** Forest plots of intervention effects on any general, human and animal MST markers in different samples of environmental samples.

# Tables

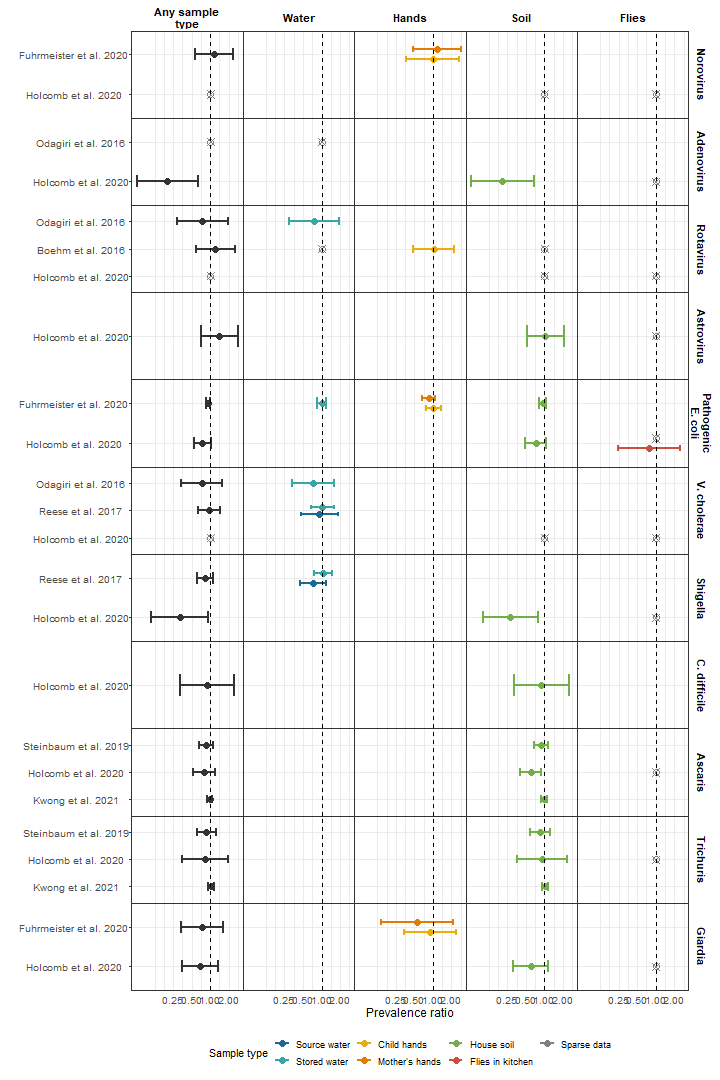
### Table 1. Prevalence of pathogens by sample type tested in each study

### Table 2. Prevalence of microbial source tracking markers by sample type tested in each study

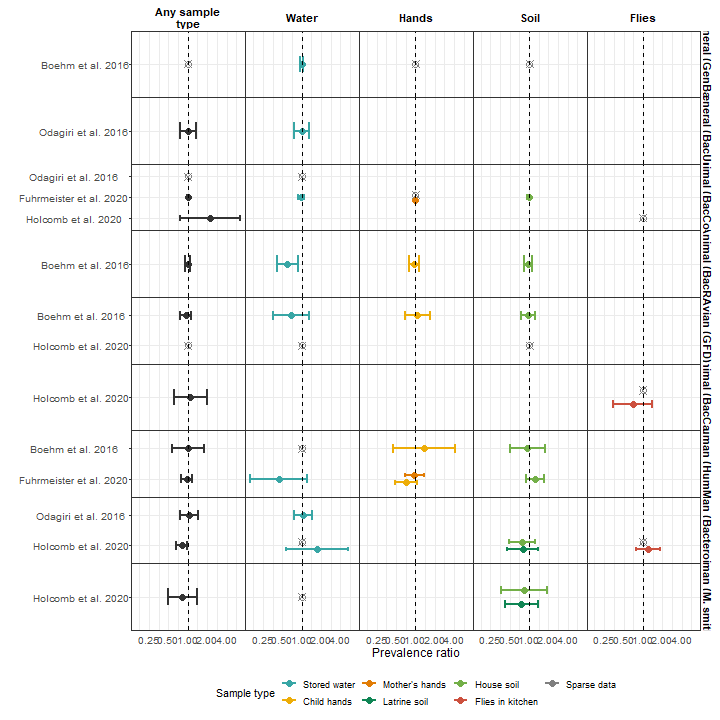
### Table 3. Mean (SD) abundances by intervention arm and adjusted abundance differences

\* marks estimates from negative binomial models.

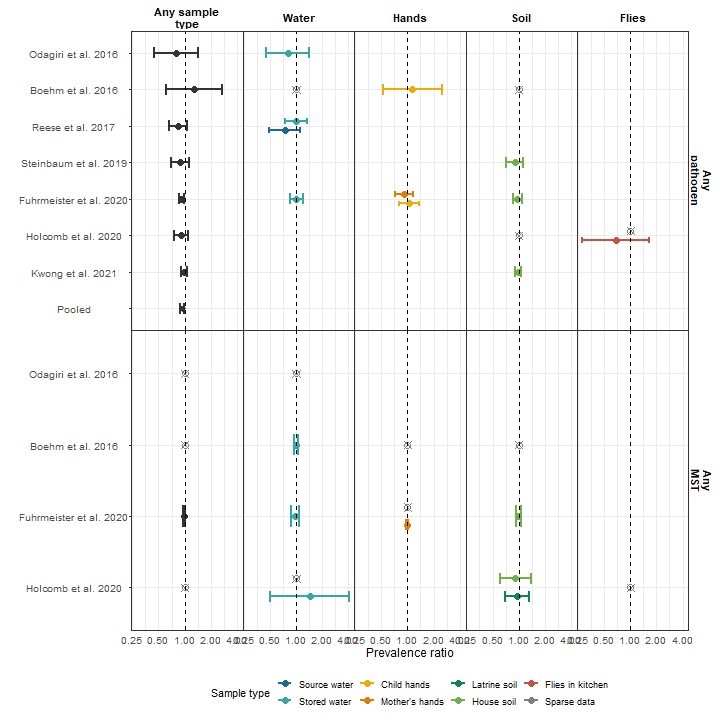
# Supplimentary Figures



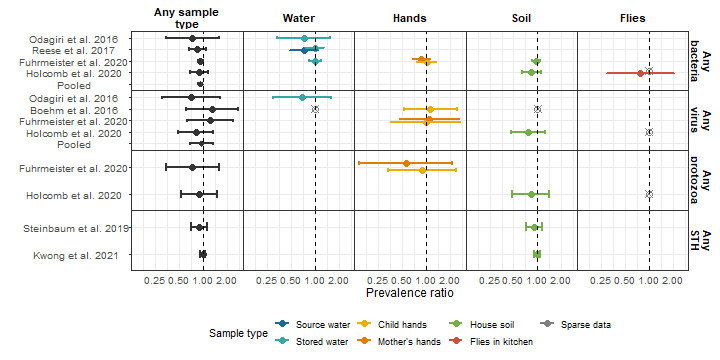
**Supplimentary Figure 1.** Prevalence of specific pathogens



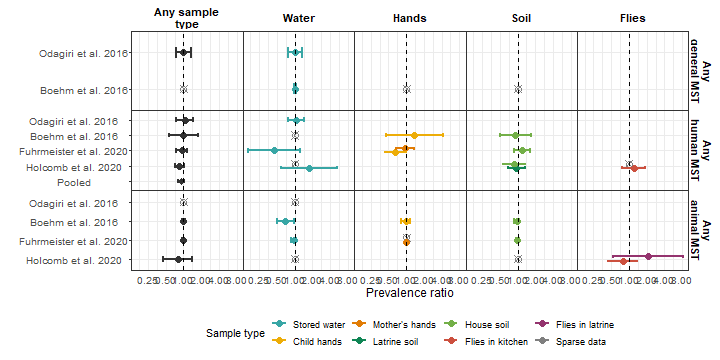
**Supplimentary Figure 2.** Prevalence of specific MST markers



**Supplimentary Figure 3.** Forest plots of unadjusted intervention effects on any enteropathogen, and 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 and target combination. Empty, grey, crossed points denote contrasts where data existed but with either too low or too high a prevalence to estimate a prevalence ratio. All estimated are adjusted for potential confounders.



**Supplimentary Figure 4.** Forest plots of unadjusted intervention effects on any virus, any bacteria, any protozoa and any STH prevalences in different types of environmental samples.



**Supplimentary Figure 5.** Forest plots of unadjusted intervention effects on any general, human and animal MST markers in different samples of environmental samples.

# Supplimentary Tables

# References

1. 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* **6**, e302–e315 (2018).

2. 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* **6**, e316–e329 (2018).

3. Sclar, G. D. *et al.* Assessing the impact of sanitation on indicators of fecal exposure along principal transmission pathways: A systematic review. *International Journal of Hygiene and Environmental Health* **219**, 709–723 (2016).

4. Hardina, C. M. & Fujioka, R. S. Soil: The environmental source of Escherichia coli and enterococci in Hawaii’s streams. *Environmental Toxicology and Water Quality* **6**, 185–195 (1991).

5. Gruber, J. S., Ercumen, A. & Colford, J. M. Coliform bacteria as indicators of diarrheal risk in household drinking water: Systematic review and meta-analysis. *PloS One* **9**, e107429 (2014).

6. Wu, J., Long, S. C., Das, D. & Dorner, S. M. Are microbial indicators and pathogens correlated? A statistical analysis of 40 years of research. *Journal of Water and Health* **9**, 265–278 (2011).

7. Liu, J. *et al.* A laboratory-developed TaqMan Array Card for simultaneous detection of 19 enteropathogens. *Journal of Clinical Microbiology* **51**, 472–480 (2013).

8. Stoeckel, D. M. & Harwood, V. J. Performance, Design, and Analysis in Microbial Source Tracking Studies. *Applied and Environmental Microbiology* **73**, 2405–2415 (2007).

9. Cochran, W. G. The Combination of Estimates from Different Experiments. *Biometrics* **10**, 101–129 (1954).

10. Arnold, B. F. *et al.* Cluster-randomised controlled trials of individual and combined water, sanitation, hygiene and nutritional interventions in rural Bangladesh and Kenya: The WASH Benefits study design and rationale. *BMJ Open* **3**, e003476 (2013).

11. Brown, J. *et al.* A controlled, before-and-after trial of an urban sanitation intervention to reduce enteric infections in children: Research protocol for the Maputo Sanitation (MapSan) study, Mozambique. *BMJ Open* **5**, e008215 (2015).

12. Reese, H. *et al.* Design and rationale of a matched cohort study to assess the effectiveness of a combined household-level piped water and sanitation intervention in rural Odisha, India. *BMJ open* **7**, e012719 (2017).

13. 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. *The Lancet. Global Health* **2**, e645–653 (2014).