Bayesian analysis of source tracking markers to estimate the effects of an urban sanitation intervention on human fecal contamination in Mozambique

AUTHORS AND AFFILIATIONS

David A. Holcomb\textsuperscript{a,b}, Jackie Knee\textsuperscript{c,d}, Drew Capone\textsuperscript{a,c}, Trent Sumner\textsuperscript{e}, Zaida Adriano\textsuperscript{e}, Rassul Nalá\textsuperscript{f}, Oliver Cumming\textsuperscript{d}, Joe Brown\textsuperscript{a}, Jill R. Stewart\textsuperscript{a,*}

\textsuperscript{a}Department of Environmental Sciences and Engineering, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States

\textsuperscript{b}Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States

\textsuperscript{c}School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia, United States

\textsuperscript{d}Department of Disease Control, London School of Hygiene & Tropical Medicine, London, United Kingdom

\textsuperscript{e}WE Consult, Maputo, Moçambique

\textsuperscript{f}Instituto Nacional de Saúde, Ministério da Saúde, Maputo, Moçambique

CORRESPONDING AUTHOR

* E-mail: jill.stewart@unc.edu; phone: (919) 966-7553
ABSTRACT

Fecal source tracking (FST) may be useful to assess pathways of fecal contamination in domestic environments and to estimate the impacts of water, sanitation, and hygiene (WASH) interventions in low-income settings. We measured two non-specific and two human-associated fecal indicators in water, soil, and surfaces before and after a shared latrine intervention from low-income households in Maputo, Mozambique participating in the Maputo Sanitation (MapSan) trial. Up to a quarter of households were impacted by human fecal contamination, but trends were unaffected by improvements to shared sanitation facilities. The intervention reduced *E. coli* gene concentrations in soil but did not impact culturable *E. coli* or the prevalence of human FST markers in a difference-in-differences analysis. Using a novel Bayesian hierarchical modeling approach, we revealed a high amount of uncertainty associated with the FST marker measurements. Diagnostic accuracy-corrected intervention effect estimates were near the null and were imprecise. The field of microbial source tracking would benefit from adding measures of diagnostic accuracy to better interpret findings, particularly when FST analyses convey insufficient information for robust inference. With improved measures, FST could help identify dominant pathways of human and animal fecal contamination in communities and guide implementation of effective interventions to safeguard health.

KEYWORDS

Diagnostic accuracy; water, sanitation, and hygiene; shared sanitation; microbial source tracking; fecal indicator; qPCR; Bayesian hierarchical model
Introduction

Water, sanitation, and hygiene (WASH) interventions aim to improve health by preventing exposure to enteric pathogens, which are introduced to the environment in the feces of infected human and animal hosts. Environmental pathways of pathogen exposure include contaminated environmental compartments like water, soil, and surfaces, as well as hands, flies, food, and fomites that have been in contact with contaminated environments.

Recent evaluations of a range of WASH interventions found inconsistent and largely negligible impacts on child diarrhea, growth, and enteric infection. Notably, combined interventions did not provide greater protection than their constituent interventions alone, suggesting that key sources of pathogens and pathways of exposure are inadequately addressed by conventional WASH strategies.

Characterizing fecal contamination in potential exposure pathways may help explain why specific interventions do or do not improve health by identifying which pathways the intervention interrupts and which remain unaffected. Fecal contamination is typically assessed by measuring fecal indicator organisms, microbes abundant in feces used to infer the presence of fecal contamination and therefore the likely presence of enteric pathogens, which are challenging to measure directly due to their diversity and low environmental concentrations. Indicator organisms can also be used for fecal source tracking (FST) by targeting microbes specific to the feces of a particular host. Animals are important sources of fecal contamination in both domestic and public environments but traditional efforts have focused on preventing exposure to human feces; differentiating between human and various animal feces could inform more appropriate intervention approaches.
Fecal indicator approaches have increasingly been applied to domestic environments in low-income settings with high burdens of enteric disease.\(^3,15\)–\(^{18,21}^{26}\) Occurrence of non-specific indicators like *Escherichia coli* is challenging to interpret in these settings due to elevated and highly variable ambient concentrations, possibly from naturalized sources, which are typically assessed in limited numbers of (cross-sectional) observations from each environmental compartment.\(^16,27\)–\(^{30}\) Other than ruminant FST markers, host-associated fecal indicators have demonstrated poor diagnostic accuracy in domestic settings.\(^17,31,32,26,16\) The use of multiple FST markers has been proposed to help address the limited accuracy of individual indicators.\(^33,34\) Several studies have calculated the conditional probability of contamination by a specific fecal source given the detection of one or more source-associated indicators in one sample.\(^31,34\)–\(^{36}\) Such analyses provide valuable intuition about the uncertainty associated with individual measurements, which can be particularly important in decision-making contexts like beach closures. To our knowledge, diagnostic performance has not been similarly accounted for when FST has been used to infer patterns and predictors of source-specific fecal contamination in domestic environments, likely overstating the confidence of such estimates.\(^4,17,18,26,37\)–\(^{39}\)

In this study, we analyze two non-specific and two human-associated fecal indicators in water, soil, and surfaces from low-income households in Maputo, Mozambique before and after a shared sanitation intervention. We explore the conditional probability of human fecal contamination in individual samples under different prevalence and indicator detection scenarios and develop a Bayesian hierarchal modeling approach that accounts for the diagnostic accuracy of multiple markers to estimate the prevalence of source-specific fecal contamination. Finally, we implement these models using both human markers to estimate intervention effects on the prevalence of human fecal contamination in multiple exposure pathways.
Materials and Methods

Study setting and intervention

We characterized fecal contamination of households with children participating in the Maputo Sanitation (MapSan) study (clinicaltrials.gov NCT02362932), a prospective, controlled before and after health impact trial of an urban, onsite sanitation intervention. The intervention was delivered to compounds (self-defined clusters of households sharing outdoor space) in low-income neighborhoods of Maputo, Mozambique, areas with high burdens of enteric disease and predominantly onsite sanitation infrastructure. Similar compounds that did not receive the intervention were recruited to serve as control sites. At baseline, both intervention and control compounds shared sanitation facilities in poor condition. The existing shared latrines in intervention compounds were replaced with pour-flush latrines that discharged aqueous effluent to infiltration pits and had sturdy, private superstructures. Intervention latrines were constructed between 2015 – 2016 by the nongovernmental organization (NGO) Water and Sanitation for the Urban Poor (WSUP), which selected intervention sites according to engineering and demand criteria (Table S1).

Data collection

The intervention impact on fecal contamination was evaluated using a controlled before-and-after (CBA) study design. Children in each intervention compound were enrolled immediately before the new latrine was opened for use, with concurrent enrollment of control compounds conducted at a similar frequency. Follow-up visits to each compound were conducted approximately 12 months following baseline enrollment. We administered compound-, household-, and child-level surveys and collected child stool during both baseline and follow-up visits, as described elsewhere. Concurrent with child stool collection during May – August
2015, we opportunistically collected environmental samples at a subset of MapSan study compounds from the shared outdoor space and from each household with children participating in the health study. During the 12-month follow-up phase in June – September 2016, we revisited the original subset of compounds and collected environmental samples from additional study compounds not sampled at baseline, as time permitted.

Detailed descriptions of environmental sample collection, processing, and analysis have been published previously. Briefly, we assessed fecal indicators in five environmental compartments: compound source water, household stored water, latrine entrance soil, household entrance soil, and household food preparation surfaces. Source water and latrine soil were sampled once per compound on each visit, while stored water, food preparation surfaces, and household soil were collected from each household with children enrolled in the health impacts study. Samples were processed by membrane filtration, preceded by manual elution for soil and swab samples, and the sample filters were analyzed for microbial indicators of fecal contamination using both culture- and molecular-based detection. We enumerated culturable *E. coli* (cEC) from filters on modified mTEC broth (Hi-Media, Mumbai, India) and immediately archived additional filters at -80°C for molecular analysis. Archived filters were analyzed by three locally validated real-time polymerase chain reaction (qPCR) assays targeting fecal microbe genes. The EC23S857 (EC23S) assay targets *E. coli* and served as an indicator of non-specific fecal contamination, while HF183/BacR287 (HF183) and Mnif both target microbes specific to human feces and served as indicators of human-source fecal contamination. DNA was isolated from soil and surface sample filters using the DNeasy PowerSoil Kit (Qiagen, Hilden, Germany) and from water sample filters with the DNA-EZ ST01 Kit (GeneRite, North Brunswick, NJ, USA), with a positive control (PC) and negative extraction.
control (NEC) included in each batch of up to 22 sample filters. PCs consisted of a clean filter spiked with $2 \times 10^8$ copies of each composite DNA standard (Table S4). Filters were treated with 3 µg salmon testes DNA (MilliporeSigma, Burlington, MA, USA) immediately before extraction as a specimen processing control (SPC) to assess PCR inhibition. We tested each extract with four qPCR assays, three targeting fecal microbes and Sketa22 targeting the salmon DNA SPC, with 10% of each sample type analyzed in duplicate for all microbial targets. Each reaction consisted of 12.5 µL TaqMan Environmental Master Mix 2.0, 2.5 µL 10x primers/probe mix, 5 µL nuclease free water (NFW), and 5 µL DNA template, for 25 µL total reaction volume. After an initial 10-minute, 95°C incubation period, cycling conditions specified by the original assay developers were followed using a CFX96 Touch thermocycler (Bio-Rad, Hercules, CA). Samples with Sketa22 cycle quantification (Cq) values > 3 above the mean Cq of extraction controls (NEC and PC) were considered inhibited and diluted 1:5 for further analysis. Each plate included three no-template controls (NTCs) and five-point, ten-fold dilution series of three extracted PCs, corresponding to triplicate reactions with $10^5 - 10^1$ or $10^6 - 10^2$ target gene copies (gc). Target concentrations were estimated from calibration curves fit to the standard dilution series using multilevel Bayesian regression with varying slopes and intercepts by extraction batch and instrument run.

Fecal indicator concentrations were log$_{10}$ transformed and normalized according to sample matrix, expressed as log$_{10}$ colony forming units (cfu) or gc per 100 mL water, 100 cm$^2$ surface, and 1 dry gram soil. Observations below the limit of detection (LOD) or too numerous to count were imputed by sample type from a truncated normal distribution with mean and standard deviation (SD) obtained through maximum likelihood estimation assuming the log$_{10}$ concentration was normally distributed and subject to left- and right-censoring, as described.
Because HF183 and Mnif were detected infrequently in baseline samples, we analyzed human-associated indicators as binary detect/non-detect variables and considered samples with ≥ 50% of replicate reactions above the LOD as positive.²⁶

**Estimating intervention effects**

We used a difference-in-differences (DID) approach to estimate the effect of the intervention on fecal indicator occurrence. DID enables unbiased estimation of the treatment effect in the absence of randomization under the "parallel trend" assumption that all unmeasured time-varying covariates related to the outcome are constant across treatment groups and that unmeasured covariates varying between treatment groups are constant through time.⁴³,⁵⁴,⁵⁵ Direct DID estimates for the mean concentration of non-specific indicators (including imputed values) and the prevalence of human-associated indicators were obtained using a bootstrap approach with 2000 samples. We calculated the mean concentration or prevalence in each of the four design strata (pre-treatment intervention compounds, post-treatment intervention compounds, pre-treatment control compounds, and post-treatment controls) by sample type, from which the DID was calculated directly [see Supporting Information (SI)]. Bootstrap 95% compatibility intervals (CI) were obtained as the 2.5 and 97.5 percentile values of the bootstrap samples.⁵⁶

We also conducted regression analyses incorporating potential confounding variables to obtain conditional DID estimates. We used the product-term representation of the DID estimator, in which binary indicators of treatment group, study phase, and their product (interaction) were included as linear predictors. The coefficient on the product term provides the conditional DID estimate.⁵⁴,⁵⁷ Separate models were fit for each combination of fecal indicator and sample type using Bayesian multilevel models with compound-varying intercepts.
Censored linear regression was used to estimate the intervention impact on the log$_{10}$ concentration of non-specific indicators using regularizing normal priors with SD = 5 on the population-level intercept and SD = 2 on predictor coefficients, including the DID terms.$^{58-60}$ We estimated the effect of the intervention on human-associated indicator prevalence using logistic regression and the prevalence odds ratio (POR) as the measure of effect. The population-level intercept and predictor coefficients were given regularizing normal priors with SD = 1.5 and SD = 0.5, respectively, which on the probability scale corresponded to a ~95% chance the population-level prevalence was between 0.05 and 0.95 and an effect of up to $\pm 0.23$ for each predictor at a population-level prevalence of 0.5. All models were fit with the package brms in R version 4.0.2 using 1500 warmup and 1000 sampling iterations on four chains. Estimates of the intervention effect were summarized by the mean and central 95% CI of the resulting 4000 posterior draws.$^{58,61}$

Adjusted models included variables for selected compound, household, meteorological, and sample characteristics. Compound population density, presence of domestic animals, and asset-based household wealth scores were derived from household and compound surveys administered during each study phase.$^{42,62}$ Previous day mean temperature and seven-day antecedent rainfall were drawn from daily summary records for a local weather station. For stored water samples, we considered whether the storage container was covered and if the mouth was wide enough to admit hands. The surface material was considered for food surface swabs, and for soil samples we accounted for sun exposure and visibly wet soil surfaces. Covariate data sources and processing have been described previously.$^{26,42}$
Conditional probability analysis

Both HF183 and Mnif were previously found to frequently misdiagnose human feces in our study area.\textsuperscript{26} An indicator's diagnostic accuracy is described by its sensitivity (Se), the probability of detecting the indicator when contamination is present, and specificity (Sp), the probability of not detecting the indicator when contamination is absent. The probability that a positive sample is contaminated depends on the marker sensitivity and specificity and the prevalence of human fecal contamination. This marginal probability of contamination can be approximated as the frequency of indicator detection among all samples to explore indicator reliability in a specific study.\textsuperscript{31} We assessed the probability that human feces were present in an environmental sample in which HF183 or Mnif was detected using Bayes' Theorem and the local sensitivity and specificity of the two markers (see SI).\textsuperscript{34–36} We calculated the conditional probability of contamination for HF183 and Mnif separately and for each combination of the two indicators by sample type. The marginal probability of contamination was approximated as the detection frequency of HF183 among all samples of a given type.

Accounting for diagnostic accuracy

Fecal indicator measurements are used as proxies for unobserved fecal contamination to estimate its prevalence and associations of interest, such as the effects of mitigation practices. This approach is vulnerable to measurement error, illustrated by the limited diagnostic accuracy of many host-associated fecal indicators.\textsuperscript{16} Bias due to inaccurate diagnostic tests can be mitigated by incorporating external information on the sensitivity and specificity of the test.\textsuperscript{53}

The expected detection frequency, $p$, of a test with sensitivity $Se$ and specificity $Sp$ is given by

$$p = Se \times \pi + (1 - Sp)(1 - \pi)$$

(1)
for an underlying condition with prevalence $\pi$.\textsuperscript{63,64} We adapted the approach of Gelman and Carpenter to estimate the intervention effect on human fecal contamination prevalence from observations of human-associated fecal indicators by incorporating external information on indicator performance within a Bayesian hierarchical framework.\textsuperscript{64} We included the product-term representation of the DID estimator and other covariates as linear predictors of the prevalence log-odds. Assuming indicator detection in the $i$th sample, $y_i$, was Bernoulli-distributed with probability $p_i$, where $p_i$ was related to the prevalence as shown in Equation (1), the accuracy-adjusted prevalence model was

$$y_i \sim \text{Bernoulli}(p_i)$$

$$p_i = Se \times \pi_i + (1 - Sp)(1 - \pi_i)$$

$$\logit(\pi_i) = \beta^0 + \beta^P P_i + \beta^T T_i + \beta^{DID} P_i \times T_i + \mathbf{X}_i \mathbf{y}$$

where $\beta^0$ is the intercept; $\beta^P$, $\beta^T$, and $\beta^{DID}$ are the parameters corresponding to indicators for study phase ($P$), treatment group ($T$), and their product; and $\mathbf{y}$ is the vector of parameters for the additional predictor variables in matrix $\mathbf{X}$.

We fit three models that differed by definition of $Se$ and $Sp$. In the simplest case (Model 1), we assumed a perfectly accurate test with $Se = Sp = 1$, thus $p = \pi$. We used the same priors as in the DID analysis. The second model (Model 2) incorporated observations from the local validation analysis by assuming

$$y^{Se} \sim \text{binomial}(n^{Se}, Se)$$

$$y^{Sp} \sim \text{binomial}(n^{Sp}, Sp)$$

for $y^{Se}$ positive results in $n^{Se}$ human fecal samples and $y^{Sp}$ negative results in $n^{Sp}$ non-human fecal samples. Informative $\text{beta}(3,2)$ priors were used for $Se$ and $Sp$, corresponding to a 95% chance they fall between 0.19 and 0.93 on the probability scale, with mean 0.6.
Because our validation sample set was small and performance estimates vary widely between studies, we fit a third model (Model 3) featuring a meta-analysis of indicator sensitivity and specificity (see SI). We assumed the log-odds of the sensitivity in the \( k \)th study, \( S_{e[k]} \), were normally distributed with mean \( \mu^{Se} \) and SD \( \sigma^{Se} \), such that

\[
y_{e[k]}^{Se} \sim \text{binomial}(n_{e[k]}, S_{e[k]})
\]

\[
\text{logit}(S_{e[k]}) \sim \text{normal}(\mu^{Se}, \sigma^{Se})
\]

with an equivalent structure for the specificity. We assigned \( k = 1 \) to our local validation study, using \( S_{e[1]} \) and \( S_{p[1]} \) as the values of \( Se \) and \( Sp \) in Equation (2).\textsuperscript{26,64} This emphasized the local performance data while allowing information from other settings to influence the estimates through partial pooling, with the extent of pooling learned from the data (expressed through \( \sigma^{Se} \) and \( \sigma^{Sp} \)).\textsuperscript{59} \( \mu^{Se} \) and \( \mu^{Sp} \) were assigned \textit{normal}(0.5, 1) priors, providing approximately equivalent coverage on the probability scale as the previous beta priors, with \textit{half-normal}(0, 0.5) priors on \( \sigma^{Se} \) and \( \sigma^{Sp} \).\textsuperscript{64}

We coded each model in the probabilistic programming language \textbf{Stan} (see SI for Stan code).\textsuperscript{65} Models were fit separately for latrine entrance soil, household entrance soil, and stored water for both HF183 and Mnif using the \textbf{RStan} interface with four chains of 1000 warmup and 1000 sampling iterations each, for a total of 4000 posterior samples.\textsuperscript{66} In addition to the DID POR given by the product-term parameter, we used the posterior predictive distribution to estimate the prevalence of human fecal contamination in each stratum and to directly calculate DID on the probability scale.\textsuperscript{59,67} Models were adjusted for the same covariates as the DID regression models.
Modeling latent human fecal contamination

Fecal contamination can be understood as a latent environmental condition for which fecal indicators serve as imperfect diagnostic tests. Information from multiple fecal indicators may be utilized by modeling each as arising from the same underlying contamination to potentially improve inference. We extended the meta-analytic model (Model 3) to include observations of both HF183 and Mnif in the same samples (Model 4), with separate detection probabilities, $p_i^{hf}$ and $p_i^{mn}$, obtained from indicator-specific sensitivity and specificity estimates applied to the same underlying prevalence, $\pi_i$. As in previous models, the DID estimator and other predictor variables were included in a linear model on the log-odds of $\pi_i$, assuming that intervention effects and other covariates acted directly on the latent prevalence.

As environmental compartments from the same compound share sources of fecal exposure, observations in different sample types can be understood as separate realizations of the same latent fecal contamination of the compound environment. Because the latrine intervention was delivered to compounds for the purpose of containing a key human fecal source, we wished to evaluate the intervention effect on compound-level human fecal contamination. We extended the previous model to simultaneously consider observations of latrine soil, household soil, and stored water in each compound (Model 5). Sample type-specific prevalence variables, $\pi_i^{[\text{type}]}$, were modeled as linear deviations from the compound-level prevalence $\pi_j$ on the log-odds scale:

$$\text{logit}(\pi_i^{[\text{type}]}) = \alpha^{[\text{type}]} + X_i^{[\text{type}]} \gamma^{[\text{type}]} + \text{logit}(\pi_i^{\text{comp}})$$

$$\text{logit}(\pi_j^{\text{comp}}) = \alpha_j^{\text{comp}} + \beta^P P_j + \beta^T T_j + \beta^\text{DID} P_j \times T_j + X_j^{\text{comp}} \gamma^{\text{comp}}$$

for sample $i$ of a given type (latrine soil, household soil, or stored water) in compound $j$, where $\alpha_j^{\text{comp}}$ is a compound-varying intercept and $\alpha^{[\text{type}]}$ is a varying intercept by sample type.
Compound-level predictors, including the DID estimator terms, were placed on the compound-prevalence log-odds. Parameters for sample-level and meteorological predictors in $X^{\text{type}}_i$ were estimated separately for each sample type. As before, all $\beta$ and $\gamma$ parameters were given normal priors with SD = 0.5. We assumed $\alpha^{\text{comp}}(j) \sim \text{normal}(\mu^{\text{comp}}, \sigma^{\text{comp}})$ and $\alpha^{\text{type}} \sim \text{normal}(0, \sigma^{\text{type}})$, with half-normal, SD = 0.5 priors on $\sigma^{\text{comp}}$ and $\sigma^{\text{type}}$ and the same $\text{normal}(0, 1.5)$ prior on $\mu^{\text{comp}}$ used for the fixed intercept $\beta^0$ in previous models. Models 4 and 5 were coded in Stan and fit using Rstan with the same conditions described for Models 1–3.

**Ethical approval**

This study was approved by the Institutional Review Board of the University of North Carolina at Chapel Hill (IRB # 15-0963) and the associated health study was approved by the Comité Nacional de Bioética para a Saúde (CNBS), Ministério da Saúde, Republic of Mozambique (333/CNBS/14), the Ethics Committee of the London School of Hygiene and Tropical Medicine (reference # 8345), and the Institutional Review Board of the Georgia Institute of Technology (protocol # H15160). Environmental samples were only collected from households with enrolled children for whom written, informed parental or guardian consent had been given.

**Results**

**Sample characteristics**

We collected a total of 770 environmental samples from 507 unique locations at 139 households in 71 compounds. Samples were collected both pre- and post-intervention at 263 locations (52%), for a total of 526 paired samples and 244 unpaired samples (Table S2). Characteristics expected to confound the relationship between sanitation and fecal contamination...
were largely similar between treatment arms during each study phase (Table 1). Cumulative precipitation was higher on average in intervention compounds at baseline and in control compounds at follow-up. Water storage containers were also more frequently covered in intervention (75%) than control households (57%) at baseline, though the majority of containers were covered in all strata. Soil surfaces were more often visibly wet in control households (51%) than intervention (33%) at follow-up, both of which were lower than at baseline (57% and 48%, respectively). Most food preparation surfaces were plastic, though more often so in control households during both study phases. A higher percentage of compounds from both treatment arms reported owning domestic animals at follow-up (80–88%) than baseline (47–68%), which may be related to differences in the questionnaire between survey phases. Median household wealth was 40–45 on a 100-point index, with higher variance among controls at follow-up. Median compound population density ranged from 5.5–8.1 residents/100 m².
<table>
<thead>
<tr>
<th>characteristic</th>
<th>level</th>
<th>metric</th>
<th>before</th>
<th></th>
<th></th>
<th>after</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>control</td>
<td>intervention</td>
<td>summary</td>
<td>control</td>
<td>intervention</td>
<td>summary</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N</td>
<td>summary</td>
<td>N</td>
<td>summary</td>
<td>N</td>
<td>summary</td>
<td>N</td>
</tr>
<tr>
<td>animals present</td>
<td>compound</td>
<td>n (%)</td>
<td>32</td>
<td>15 (47)</td>
<td>25</td>
<td>17 (68)</td>
<td>30</td>
<td>24 (80)</td>
</tr>
<tr>
<td>population density (persons/100 m²)</td>
<td>compound</td>
<td>median (IQR)</td>
<td>29</td>
<td>5.5 (3.5)</td>
<td>23</td>
<td>8.1 (5.9)</td>
<td>28</td>
<td>5.9 (4.8)</td>
</tr>
<tr>
<td>wealth index (0 - 100)</td>
<td>household</td>
<td>median (IQR)</td>
<td>51</td>
<td>43 (12)</td>
<td>40</td>
<td>43 (12)</td>
<td>55</td>
<td>45 (19)</td>
</tr>
<tr>
<td>previous day mean temperature (°C)</td>
<td>date</td>
<td>median (IQR)</td>
<td>19</td>
<td>21 (2)</td>
<td>16</td>
<td>20 (2)</td>
<td>17</td>
<td>20 (1)</td>
</tr>
<tr>
<td>seven-day cumulative precipitation (mm)</td>
<td>date</td>
<td>median (IQR)</td>
<td>19</td>
<td>9 (3)</td>
<td>16</td>
<td>14 (3)</td>
<td>17</td>
<td>13 (39)</td>
</tr>
<tr>
<td>water container covered</td>
<td>sample</td>
<td>n (%)</td>
<td>44</td>
<td>25 (57)</td>
<td>28</td>
<td>21 (75)</td>
<td>38</td>
<td>21 (55)</td>
</tr>
<tr>
<td>narrow-mouth water container</td>
<td>sample</td>
<td>n (%)</td>
<td>44</td>
<td>13 (30)</td>
<td>28</td>
<td>10 (36)</td>
<td>38</td>
<td>13 (34)</td>
</tr>
<tr>
<td>plastic food surface material</td>
<td>sample</td>
<td>n (%)</td>
<td>34</td>
<td>30 (88)</td>
<td>23</td>
<td>18 (78)</td>
<td>29</td>
<td>26 (90)</td>
</tr>
<tr>
<td>shaded latrine soil</td>
<td>sample</td>
<td>n (%)</td>
<td>32</td>
<td>24 (75)</td>
<td>17</td>
<td>12 (71)</td>
<td>30</td>
<td>25 (83)</td>
</tr>
<tr>
<td>shaded household soil</td>
<td>sample</td>
<td>n (%)</td>
<td>42</td>
<td>31 (74)</td>
<td>28</td>
<td>24 (86)</td>
<td>35</td>
<td>32 (91)</td>
</tr>
<tr>
<td>wet latrine soil surface</td>
<td>sample</td>
<td>n (%)</td>
<td>32</td>
<td>20 (62)</td>
<td>17</td>
<td>13 (76)</td>
<td>30</td>
<td>18 (60)</td>
</tr>
<tr>
<td>wet household soil surface</td>
<td>sample</td>
<td>n (%)</td>
<td>42</td>
<td>24 (57)</td>
<td>27</td>
<td>13 (48)</td>
<td>35</td>
<td>18 (51)</td>
</tr>
<tr>
<td>latrine soil moisture (%)</td>
<td>sample</td>
<td>median (IQR)</td>
<td>33</td>
<td>9.8 (9.8)</td>
<td>23</td>
<td>8.4 (7.2)</td>
<td>30</td>
<td>10.0 (7.9)</td>
</tr>
<tr>
<td>household soil moisture (%)</td>
<td>sample</td>
<td>median (IQR)</td>
<td>49</td>
<td>9.9 (8.6)</td>
<td>35</td>
<td>6.9 (6.1)</td>
<td>47</td>
<td>7.8 (5.4)</td>
</tr>
</tbody>
</table>

\(^a\) interquartile range
**Fecal indicator occurrence**

At least one fecal indicator was detected in 94% of samples (720/770) and *E. coli* was detected in 718 samples: by culture in 81% (611/755) and by qPCR in 86% (655/763). Mean cEC concentrations were lower at follow-up for all sample types in both treatment arms, a pattern not observed for EC23S concentrations (Figure 1). Of the 763 samples tested for human-associated indicators, 28% (217) were positive for at least one human marker. Human-associated indicators were common in soils (23–65% prevalence) but only HF183 was regularly detected in stored water (10–22%) and both indicators were rare on food surfaces (0–9%).

Bootstrap DID estimates suggest the intervention reduced EC23S concentrations on food preparation surfaces and HF183 prevalence in household soil but minimally impacted fecal indicator occurrence in other sample types (Table S5). Notably, HF183 prevalence in household soil was similar among intervention households in both study phases but increased among control compounds at follow-up. By contrast, model-based DID estimates, adjusted for potential confounding, were consistent with no intervention effect on food preparation surface EC23S concentration or household soil HF183 prevalence (Table S6). Adjusted models instead indicate the intervention reduced latrine soil concentrations of EC23S [mean difference: -1.2 (95% CI: -2.1, -0.30) log10 gc/dry g]. Due to low detection frequency, models were not fit for either human marker on food surfaces or for Mnif in stored water; source water samples were excluded from all analyses.26
Figure 1. Bootstrap estimates of fecal indicator occurrence by study phase and treatment arm. Points indicate mean log10 concentration for *E. coli* indicators and prevalence of human-associated indicators, with bars presenting bootstrap 95% CIs.

**Conditional probability of human fecal contamination**

The probability that a sample is contaminated with human feces given the detection of a human indicator is a function of the indicator's sensitivity and specificity and the prevalence of human contamination in the study environment. While the prevalence of human contamination is unknown, the relatively poor sensitivity (HF183: 64%; Mnif: 71%) and specificity (HF183: 67%; Mnif: 70%) of both markers present ample opportunity for both false positive and false negative test results. At 15% prevalence (approximately the detection frequency of HF183 in stored water), the probability of human contamination given a positive test was 26% for HF183 and
30% for Mnif. Only with prevalence above 30–35% was detecting either indicator more likely than not to correctly diagnose the presence of human fecal contamination. Combining test results from both indicators improved identification of human contamination, increasing the probability of contamination to 45% when both markers were positive and the prevalence was 15% (Figure 2). However, the two human markers frequently disagreed when assessed in the same sample, conflicting in 44% of household soil, 43% of latrine soil, and 15% of stored water samples. Furthermore, at 44% prevalence (the highest detection frequency for HF183, observed in latrine soils), there remained a >20% chance that a sample positive for both indicators was not contaminated. Among lower-prevalence sample types the conditional probability never reached 50%. Unless the background prevalence in the study area was about 45% or greater, it is unlikely that the use of HF183 and Mnif reliably identified human contamination in individual samples, particularly given the frequent disagreement between the two markers.
Diagnostic accuracy of human fecal indicators

We considered HF183 validation data from 14 studies of diagnostic sensitivity and 15 studies of specificity (Table S7). For Mnif, we incorporated data from four validation studies. The number of samples ranged from 5–62 for human feces and 15–359 for non-human. Reported crude sensitivity ranged from 29-100% for HF183 and 60-82% for Mnif. Crude specificities were reported from 40–100% for HF183 and 58-93% for Mnif.
By incorporating indicator validation data into models of human fecal contamination, we obtained estimates of indicator sensitivity and specificity that were partially informed by observations in environmental samples with unknown fecal contamination status. For this reason, accuracy estimates from the same model differed by sample type (Table 2). We first considered only validation data from our study area (Model 2), which produced slightly lower point estimates for HF183 sensitivity (60%) than obtained from crude calculations (64%). The sensitivity estimates were similar for all three sample types, with the greatest uncertainty in stored water (95% CI: 38-81%). The HF183 specificity estimate was similar to the crude value (67%) for latrine soil [66% (53-80%)] but the estimates were higher for household soil [72% (61-83%)] and stored water [85% (77-91%)], in which HF183 was detected less frequently.

Sensitivity and specificity patterns for Mnif followed similar patterns.

Meta-analytic sensitivity and specificity estimates were slightly higher than the single-study estimates when using a single indicator (Model 3) and when combining both indicators (Model 4). However, the 95% CIs remained similar for all three models, suggesting the inclusion of additional information minimally impacted estimates of diagnostic accuracy. Model 5, which simultaneously considered both indicators in all sample types, provided a single set of sensitivity and specificity estimates for each indicator. The estimates were comparable to values observed for individual sample types but featured narrower 95% CIs, indicating reduced uncertainty in indicator performance with the inclusion of data from multiple sample types. While the estimated sensitivity for Mnif [71% (59-83%)] remained similar to the crude value (71%), the estimated specificity of both indicators and HF183 sensitivity were somewhat higher than expected from crude calculations.
<table>
<thead>
<tr>
<th>marker</th>
<th>sensitivity (95% CI)</th>
<th>specificity (95% CI)</th>
<th>N</th>
<th>prevalence estimate (95% CI)</th>
<th>prevalence DID¹ (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>latrine soil</td>
<td></td>
<td></td>
<td></td>
<td>before</td>
<td>after</td>
</tr>
<tr>
<td>bootstrap</td>
<td>HF183</td>
<td>0.60 (0.42, 0.79)</td>
<td>0.66 (0.53, 0.80)</td>
<td>98</td>
<td>0.38 (0.05, 0.88)</td>
</tr>
<tr>
<td></td>
<td>Mniff</td>
<td>0.64 (0.47, 0.82)</td>
<td>0.66 (0.51, 0.81)</td>
<td>98</td>
<td>0.48 (0.09, 0.90)</td>
</tr>
<tr>
<td>model 2b</td>
<td>HF183</td>
<td>0.65 (0.45, 0.85)</td>
<td>0.68 (0.55, 0.82)</td>
<td>98</td>
<td>0.34 (0.05, 0.83)</td>
</tr>
<tr>
<td></td>
<td>Mniff</td>
<td>0.70 (0.56, 0.83)</td>
<td>0.72 (0.58, 0.85)</td>
<td>98</td>
<td>0.49 (0.14, 0.84)</td>
</tr>
<tr>
<td>model 4d</td>
<td>HF183</td>
<td>0.64 (0.47, 0.82)</td>
<td>0.71 (0.57, 0.84)</td>
<td>98</td>
<td>0.39 (0.11, 0.73)</td>
</tr>
<tr>
<td></td>
<td>Mniff</td>
<td>0.71 (0.58, 0.84)</td>
<td>0.71 (0.57, 0.84)</td>
<td>98</td>
<td>0.34 (0.12, 0.65)</td>
</tr>
<tr>
<td>household soil</td>
<td></td>
<td></td>
<td></td>
<td>before</td>
<td>after</td>
</tr>
<tr>
<td>bootstrap</td>
<td>HF183</td>
<td>0.60 (0.38, 0.80)</td>
<td>0.72 (0.61, 0.83)</td>
<td>147</td>
<td>0.26 (0.15, 0.41)</td>
</tr>
<tr>
<td></td>
<td>Mniff</td>
<td>0.57 (0.34, 0.80)</td>
<td>0.73 (0.63, 0.84)</td>
<td>146</td>
<td>0.37 (0.23, 0.52)</td>
</tr>
<tr>
<td>model 3c</td>
<td>HF183</td>
<td>0.66 (0.43, 0.85)</td>
<td>0.74 (0.63, 0.85)</td>
<td>147</td>
<td>0.25 (0.04, 0.63)</td>
</tr>
<tr>
<td></td>
<td>Mniff</td>
<td>0.68 (0.50, 0.82)</td>
<td>0.76 (0.67, 0.86)</td>
<td>146</td>
<td>0.26 (0.03, 0.60)</td>
</tr>
<tr>
<td>model 5c</td>
<td>HF183</td>
<td>0.72 (0.57, 0.87)</td>
<td>0.85 (0.78, 0.91)</td>
<td>146</td>
<td>0.26 (0.09, 0.49)</td>
</tr>
<tr>
<td></td>
<td>Mniff</td>
<td>0.71 (0.59, 0.83)</td>
<td>0.78 (0.68, 0.86)</td>
<td>146</td>
<td>0.30 (0.08, 0.64)</td>
</tr>
<tr>
<td>stored water</td>
<td></td>
<td></td>
<td></td>
<td>before</td>
<td>after</td>
</tr>
<tr>
<td>bootstrap</td>
<td>HF183</td>
<td>0.60 (0.38, 0.81)</td>
<td>0.85 (0.78, 0.91)</td>
<td>170</td>
<td>0.15 (0.02, 0.40)</td>
</tr>
<tr>
<td>model 2d</td>
<td>HF183</td>
<td>0.67 (0.43, 0.85)</td>
<td>0.86 (0.79, 0.92)</td>
<td>170</td>
<td>0.15 (0.02, 0.38)</td>
</tr>
<tr>
<td>model 5c</td>
<td>HF183</td>
<td>0.72 (0.57, 0.87)</td>
<td>0.85 (0.78, 0.91)</td>
<td>169</td>
<td>0.19 (0.04, 0.43)</td>
</tr>
<tr>
<td>latent compound</td>
<td></td>
<td></td>
<td></td>
<td>before</td>
<td>after</td>
</tr>
<tr>
<td>model 5c</td>
<td>HF183</td>
<td>0.72 (0.57, 0.87)</td>
<td>0.85 (0.78, 0.91)</td>
<td>109</td>
<td>0.27 (0.09, 0.52)</td>
</tr>
</tbody>
</table>
a model 1: single sample type, single marker assuming perfect sensitivity and specificity

b model 2: single sample type, single marker with sensitivity and specificity from local validation study

c model 3: single sample type, single marker with meta-analytic sensitivity and specificity

d model 4: single sample type, two markers with meta-analytic sensitivity and specificity

e model 5: three sample types, two markers with meta-analytic sensitivity and specificity

f difference-in-differences
Prevalence of human fecal contamination

Posterior predictions from each of the five models were used to estimate stratum-specific prevalence of human fecal contamination. To compare by treatment assignment and study phase, we predicted prevalence for compounds with no animals or antecedent precipitation and mean population density (7 persons/100 m²), wealth score (46), and previous-day temperature (20.4 °C), in which soil surfaces were dry and shaded and water storage containers possessed wide, uncovered mouths. The prevalence estimates were notably imprecise: the 95% CI of the HF183 prevalence in post-treatment latrine soil ranged from 3% to 92% for Model 2 (Table 2). The 95% CI widths were similar for Model 1 and the bootstrap estimates but were substantially wider for the other four models. The intervals narrowed somewhat when both indicators were considered (Model 4) and narrowed further when all sample types were incorporated (Model 5) but were still wider than the estimates that ignored diagnostic accuracy.

Although we did not formally assess the paired differences between prevalence estimates, the wide and largely overlapping posterior predictive CIs indicate a limited ability to distinguish between prevalence point estimates in different strata or between models. That the DID estimates on the probability scale were strongly consistent with no effect for all model specifications further suggests that the available data were insufficiently informative to assess prevalence differences between strata. The corresponding prevalence odds ratio estimates obtained directly from the DID product term were likewise imprecise (Figure S1). Nonetheless, the model-based point estimates were consistently more similar between strata than the corresponding bootstrap estimates, indicating the models induced shrinkage in the stratum-specific prevalence estimates. This trend was notable for Model 5, which estimated sample-type specific prevalence as linear deviations on the logit scale from a shared compound prevalence. The effects of study phase and
treatment assignment were estimated for the latent compound prevalence and thus identical for each sample type. The latent compound prevalence estimates were quite similar, particularly between study phases for the same treatment group: 27% (95% CI: 9-52%) at baseline and 28% (9-53%) at follow-up for control compounds and 22% (6-50%) at baseline and 22% (6-47%) at follow-up for intervention compounds. The corresponding estimates for household soil were nearly identical to the compound-level estimates, with somewhat higher estimates for latrine soil and lower for stored water. Although the physical interpretation of this latent compound-level construct is uncertain, these estimates suggest that about a quarter of compounds were meaningfully impacted by human fecal contamination, which was unaffected by improvements to shared sanitation facilities.

**Discussion**

The provision of shared latrines reduced average soil concentrations of the molecular *E. coli* marker EC23S at latrine entrances by more than 1-log$_{10}$ but did not have a comparable effect on culturable *E. coli*. EC23S latrine soil concentrations rose more in control compounds than they fell in intervention compounds, which under the parallel trends assumption is interpreted as a secular trend upwards that the intervention mitigated, for a much smaller absolute reduction than suggested by the DID estimate (Figure 1). However, an opposite, downward trend was observed for all cEC concentrations. This discrepancy between two tests for the same organism complicates the interpretation of the relatively strong intervention effect estimated for EC23S.

A cluster-randomized trial in rural Bangladesh likewise found no reduction in culturable *E. coli* concentrations from sanitation improvements. Latrine provision also did not reduce the prevalence of pathogenic *E. coli* genes in soil, meaning neither culture- nor molecular-based measurements of soil *E. coli* were affected. Other recent trials have not assessed intervention
impacts on fecal contamination of soil, but several have evaluated contamination of drinking water, with some also testing child hands, food, or fomites.\textsuperscript{15} As with the present study, all found no effect of sanitation-only interventions on any environmental compartment; combined water, sanitation, and hygiene interventions improved drinking water quality in two studies.\textsuperscript{13,14} Two cluster-randomized trials, in India and Bangladesh, also found no effect of rural sanitation interventions on the prevalence of human-associated indicators in stored drinking water.\textsuperscript{37,39} Both studies also assessed human markers in mother and child hand rinse samples, which were not collected in this study. No effect was observed for either hand type in India or on mother hands in Bangladesh, although the human marker prevalence may have been reduced on child hands.\textsuperscript{39} Many have speculated that sanitation's apparent lack of effect may be due in part to animal fecal contamination.\textsuperscript{12,22} Animal feces often contain pathogens capable of infecting humans and animal fecal biomass in domestic environments is estimated to far exceed that from humans.\textsuperscript{22,72–74} Inadequate management of child feces and fecal sludge, contamination of food and water outside the home, and inadequate community-level drainage, solid waste, and sanitation services all present potential pathways of continued contamination despite household sanitation improvements.\textsuperscript{24,75–80} Recognizing calls for "transformative" WASH to address these multifarious hazards, sustained progress may require high standards of housing and public services, necessitating multi-sectoral coordination and financing, that render WASH considerations somewhat incidental.\textsuperscript{12,81–83} Nevertheless, quality sanitation infrastructure can provide important benefits irrespective of preventing pathogen exposure, particularly in crowded urban settlements.\textsuperscript{84,85} For example, previous research found users of MapSan intervention latrines and similar facilities in the same neighborhoods reported reduced disgust and embarrassment about unhygienic conditions and improved perceptions of security and privacy.\textsuperscript{86}
Accounting for the diagnostic accuracy of FST markers revealed far greater uncertainty about host-specific fecal contamination, both of individual samples and population averages, than indicated by the raw indicator measurements. The relatively poor sensitivity and specificity of both human markers in this setting severely limited their ability to identify specific samples contaminated with human feces. Correcting for indicator sensitivity and specificity to human-source contamination, coupled with the limited observations of each sample type, yielded imprecise prevalence estimates that were consistent with both near absence and almost omnipresence of contamination. This imprecision inhibited detecting intervention effects. While the point estimates for the intervention effect were relatively close to the null, the full posterior distributions encompassed both large reductions and substantial increases in prevalence attributable to the intervention. This analysis does not rule out the possibility that sanitation improvements reduced the prevalence of human fecal contamination. Rather, it strongly suggests that the tools used were inadequate, conveying too little information to address the research question with an acceptable degree of confidence.

These limitations highlight the importance of conducting local validation studies for any new FST application to characterize and correct for diagnostic performance in order to mitigate overconfidence and overinterpretation. Given the intermingling in low-income settings of humans and animals, and their gut microbiomes, alternative FST targets such as mitochondrial DNA could prove more accurate. Recent technological advances also present opportunities for new approaches that might bypass the limitations of the current FST paradigm, including portable, long-read sequencing platforms for metagenomic-based source tracking and parallel PCR platforms that render simultaneous analysis of multiple FST markers and comprehensive direct pathogen detection increasing feasible. However, technology in the near term is
unlikely to satisfactorily address the substantial variability, limited analytical sensitivity, and matrix interference characteristic of environmental microbial assessments. The low signal typical of environmental measurements suggests that study designs—preferably longitudinal—that maximize observations on select pathways of greatest interest should be prioritized to support more robust inference, regardless of analytical approach.

Acknowledgment

This study was funded by the United States Agency for International Development under Translating Research into Action, Cooperative Agreement No. GHS-A-00-09-00015-00, the Bill and Melinda Gates Foundation grant OPP1137224, and the National Institute of Environmental Health Sciences (T32ES007018). We gratefully acknowledge our implementing partner, Water and Sanitation for the Urban Poor; the field research team at WE Consult; the laboratory staff at the Mozambique National Institute of Health; and Olimpio Zavale, Olivia Ginn, and Anna Stamatogiannakis for sampling support.

Supporting Information Available

Site selection criteria; samples collected; laboratory quality control; conditional probability calculations; difference-in-differences estimates; human fecal indicator validation studies; accuracy-adjusted intervention effect estimates; unadjusted human fecal contamination prevalence estimates; model Stan code.
References


P.; Lin, A.; Luby, S. P.; Mertens, A.; Njenga, S. M.; Nyambane, G.; Ram, P. K.; Colford, J.
M. Effects of Water Quality, Sanitation, Handwashing, and Nutritional Interventions on
Lancet Glob. Health 2018, 6 (3), e316–e329. https://doi.org/10.1016/S2214-
109X(18)30005-6.

(7) Luby, S. P.; Rahman, M.; Arnold, B. F.; Unicomb, L.; Ashraf, S.; Winch, P. J.; Stewart, C.
P.; Begum, F.; Hussain, F.; Benjamin-Chung, J.; Leontsini, E.; Naser, A. M.; Parvez, S. M.;
Effects of Water Quality, Sanitation, Handwashing, and Nutritional Interventions on
Diarrhoea and Child Growth in Rural Bangladesh: A Cluster Randomised Controlled Trial.
Lancet Glob. Health 2018, 6 (3), e302–e315. https://doi.org/10.1016/S2214-
109X(17)30490-4.

(8) Grembi, J. A.; Lin, A.; Karim, M. A.; Islam, M. O.; Miah, R.; Arnold, B. F.; McQuade, E.
A.; Holmes, S. P.; Stewart, C. P.; Benjamin-Chung, J.; Colford, J. M.; Houpt, E. R.; Luby,
S. P. Effect of Water, Sanitation, Handwashing and Nutrition Interventions on
Enteropathogens in Children 14 Months Old: A Cluster-Randomized Controlled Trial in

(9) Humphrey, J. H.; Mbuya, M. N. N.; Ntozini, R.; Moulton, L. H.; Stoltzfus, R. J.; Tavengwa,
N. V.; Mutasa, K.; Majo, F.; Mutasa, B.; Mangwadu, G.; Chasokela, C. M.; Chigumira, A.;


https://doi.org/10.1021/acs.est.8b05153.


https://doi.org/10.1101/731992.


(34) Johnston, C.; Byappanahalli, M. N.; Gibson, J. M.; Ufnar, J. a.; Whitman, R. L.; Stewart, J. R. Probabilistic Analysis Showing That a Combination of Bacteroides and Methanobrevibacter Source Tracking Markers Is Effective for Identifying Waters


Bauza, V.; Madadi, V. O.; Ocharo, R. M.; Nguyen, T. H.; Guest, J. S. Microbial Source Tracking Using 16S rRNA Amplicon Sequencing Identifies Evidence of Widespread Contamination from Young Children’s Feces in an Urban Slum of Nairobi, Kenya. 


Acharya, K.; Khanal, S.; Pantha, K.; Amatya, N.; Davenport, R. J.; Werner, D. A


Li, X.; Harwood, V. J.; Nayak, B.; Staley, C.; Sadowsky, M. J.; Weidhaas, J. A


Ercumen, A.; Arnold, B. F.; Naser, A. Mohd.; Unicomb, L.; Colford, J. M.; Luby, S. P.

Bayesian analysis of source tracking markers to estimate the effects of an urban sanitation intervention on human fecal contamination in Mozambique: Supporting Information

David A. Holcomb, Jackie Knee, Drew Capone, Trent Sumner, Zaida Adriano, Rassul Nalá, Oliver Cumming, Joe Brown, Jill R. Stewart

This supporting information contains 34 pages, 1 figure, and 8 tables.
Table of Contents

S1. Site selection criteria .................................................................................................................. 3
S2. Samples collected ......................................................................................................................... 3
S3. Laboratory quality control ......................................................................................................... 4
S4. Conditional probability calculations ......................................................................................... 7
S5. Difference-in-differences estimates ........................................................................................... 7
S6. Human fecal indicator validation studies ................................................................................... 10
S7. Accuracy-adjusted intervention effect estimates ....................................................................... 11
S8. Human fecal contamination prevalence estimates ................................................................... 13
S9. Stan code .................................................................................................................................. 15
  S9.1. Model 1 ............................................................................................................................... 15
  S9.2. Model 2 ............................................................................................................................... 17
  S9.3. Model 3 ............................................................................................................................... 19
  S9.4. Model 4 ............................................................................................................................... 21
  S9.5. Model 5 ............................................................................................................................... 24
S10. Supplementary References ...................................................................................................... 29
S1. Site selection criteria

Intervention sites were selected by the nongovernmental organization (NGO) Water and Sanitation for the Urban Poor (WSUP) according to feasibility and demand criteria (Table S1). The MapSan research were not involved in the design of the intervention or selection of the intervention sites, but did recruit similar compounds to serve as control sites according to a reduced set of the same selection criteria applied to intervention sites.

Table S1. Baseline site selection criteria for intervention and control compounds

<table>
<thead>
<tr>
<th>criterion</th>
<th>required for intervention compounds</th>
<th>control compounds</th>
</tr>
</thead>
<tbody>
<tr>
<td>located in the 11 pre-defined implementation neighborhoods</td>
<td>yes</td>
<td>no²</td>
</tr>
<tr>
<td>residents share sanitation in poor condition</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>at least 12 residents</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>residents willing to contribute to latrine construction costs</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>sufficient space available for construction of the new facility</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>accessible for transportation of construction materials and tank-emptying activities</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>legal access to piped water supply</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>groundwater level deep enough to accommodate a septic tank</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>at least one child younger than 48 months old in residence</td>
<td>no</td>
<td>yes</td>
</tr>
</tbody>
</table>

² also recruited from 6 similar, adjacent neighborhoods; see Knee et al. (2020)²

S2. Samples collected

We visited additional households and compounds in the follow-up study phase, collecting more samples than at baseline (Table S2).² Fewer source water samples were collected at follow-up because the municipal water supply was available less often. One control compound (with two enrolled households) independently upgraded the latrine after baseline and was excluded from the follow-up sample.
Table S2. Number of observations of compounds, households, and each sample type

<table>
<thead>
<tr>
<th>observation unit</th>
<th>before control</th>
<th>before intervention</th>
<th>after control</th>
<th>after intervention</th>
</tr>
</thead>
<tbody>
<tr>
<td>compounds</td>
<td>33</td>
<td>25</td>
<td>30</td>
<td>34</td>
</tr>
<tr>
<td>households</td>
<td>86</td>
<td>66</td>
<td>86</td>
<td>90</td>
</tr>
<tr>
<td>latrine entrance soil</td>
<td>33</td>
<td>23</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>household entrance soil</td>
<td>49</td>
<td>36</td>
<td>47</td>
<td>45</td>
</tr>
<tr>
<td>compound source water</td>
<td>23</td>
<td>21</td>
<td>19</td>
<td>22</td>
</tr>
<tr>
<td>household stored water</td>
<td>50</td>
<td>41</td>
<td>52</td>
<td>55</td>
</tr>
<tr>
<td>food preparation surfaces</td>
<td>50</td>
<td>40</td>
<td>53</td>
<td>51</td>
</tr>
</tbody>
</table>

S3. Laboratory quality control

Sterile PBS was filtered as a laboratory blank for approximately every 10 samples filtered, all of which were culture-negative for cEC (n = 151). At least three no template control (NTC) reactions were included on every qPCR run using 5 µL nuclease-free water in place of sample template. Each qPCR run typically included samples from three extraction batches, and a negative extraction control (NEC) was included from each extraction batch represented on the plate. HF183 was absent in all NTC (n = 46) and NEC (n = 46) reactions. Mnif was likewise not detected in any NTC (n = 42) or NEC (n = 46) reaction. However, EC23S was detected in 2% of NTC (1/45) and 11% of NEC (5/46) reactions. EC23S concentrations were low in the contaminated negative controls, with a mean Cq of 38.1 and a minimum Cq of 37.1—only slightly above the detection limit of 39 cycles. Such low levels of contamination have been reported previously and are thought to be due to residual *E. coli* DNA in the Environmental Master Mix from the production process.\(^2,3\) One latrine soil sample was considered inhibited based on a greater than 3 Cq deviation from the mean Sketa22 Cq of all NECs and positive controls (PC) on the same plate. This sample was diluted 1:5 in all further analyses. For each sample type, 10% of samples were randomly selected to be analyzed by each qPCR assay in technical duplicate reactions. The detection status of EC23S agreed in 96% (74/77) of replicate
pairs. HF183 results matched for 86% (66/77) of paired reactions and Mnif agreed in 95% (71/75).

Calibration curves for quantifying molecular fecal indicator gene copies were fit to tenfold dilution series of positive controls that had been spiked with reference material for all three assays and extracted alongside each batch of samples. Three gBlock linear DNA fragments (Integrated DNA Technologies, Skokie, IL, USA) containing composite reference sequences for the three fecal source tracking assays considered in this paper as well as additional assays considered in the associated validation study were used as standard reference material for the positive controls (Table S4). Extracting the reference material accounted for loss of target DNA during extraction but induced additional variability between dilution series constructed from each PC. We therefore allowed both the slopes and intercepts of the calibration curves to vary by qPCR instrument run and extraction batch to account for the additional variation. The resulting curves were relatively linear ($R^2 > 95\%$) when averaged across all instrument runs and extraction batches but were somewhat inefficient, particularly for HF183 (Table S3).

<table>
<thead>
<tr>
<th>assay</th>
<th>intercept (95% CI)</th>
<th>slope (95% CI)</th>
<th>efficiency (95% CI)</th>
<th>$R^2$ (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC23S</td>
<td>47.91 (47.25, 48.64)</td>
<td>-3.50 (-3.64, -3.37)</td>
<td>93.11 (88.12, 98.14)</td>
<td>0.98 (0.97, 0.98)</td>
</tr>
<tr>
<td>HF183</td>
<td>47.45 (46.41, 48.58)</td>
<td>-3.85 (-4.07, -3.67)</td>
<td>81.83 (76.07, 87.36)</td>
<td>0.98 (0.97, 0.98)</td>
</tr>
<tr>
<td>Mnif</td>
<td>48.77 (47.66, 50.25)</td>
<td>-3.47 (-3.79, -3.23)</td>
<td>94.51 (83.57, 104.0)</td>
<td>0.95 (0.93, 0.95)</td>
</tr>
</tbody>
</table>
Table S4. Synthetic DNA reference material spiked into positive controls

<table>
<thead>
<tr>
<th>assays covered</th>
<th>sequence [5’-3’]</th>
<th>GenBank (base positions)</th>
<th>length [bases]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BacHum-UCD</strong></td>
<td>TGGTCTTAAATACCGGATACGACCACCTCTGCACGATGG</td>
<td>AB242142 (170-730)</td>
<td>560</td>
</tr>
<tr>
<td><strong>BacUni-UCD</strong></td>
<td>CCGGTGTTGGAAGTTTTCTCATTGGGAGATGGCTC</td>
<td>FJ462358 (156-746)</td>
<td>732</td>
</tr>
<tr>
<td><strong>HF183/BacR28</strong></td>
<td>GACGCCACTTCGGGGACTGAGACCGCCACAGTCCAT</td>
<td>JN084061 (29-171)</td>
<td>732</td>
</tr>
<tr>
<td><strong>GFD</strong></td>
<td>CCGGCTATATCCCGGAAGATCTTCCTGACGGATCCTGGTACG</td>
<td>AE015928 (4515891-4515973)</td>
<td>475</td>
</tr>
<tr>
<td><strong>LA35</strong></td>
<td>TGGTCTTAAATACCGGATACGACCACCTCTGCACGATGG</td>
<td>AB019138 (192-363)</td>
<td>475</td>
</tr>
<tr>
<td><strong>EC23S857</strong></td>
<td>TAACTATGCTCATCGTTCGTCAGCAGAATACAGTTAATT</td>
<td>AC_000008 (18885-19000)</td>
<td>475</td>
</tr>
<tr>
<td><strong>HAdV</strong></td>
<td>GCTACACCTGCTGAAACCACCCTACTTCTGG</td>
<td>DQ682619 (847-954)</td>
<td>475</td>
</tr>
<tr>
<td><strong>Mnif</strong></td>
<td>CAACTCTTCTGACGGATCGGTATCGTTCGTCAGCAGAATACAGTTAATT</td>
<td>GACGCCACTTCGGGGACTGAGACCGCCACAGTCCAT</td>
<td>AE015928 (4515891-4515973)</td>
</tr>
</tbody>
</table>

**GenBank**
- GenBank: Database for nucleotide and protein sequences.
S4. Conditional probability calculations

The conditional probability of human contamination, \( C \), given the detection of two human-associated markers, \( M_1 \) and \( M_2 \) is given by

\[
P(C^+|M_1^+ \cap M_2^+) = \frac{P(M_1^+|C^+) \times P(M_2^+|C^+) \times P(C^+)}{P(M_1^+|C^+) \times P(M_2^+|C^+) \times P(C) + P(M_1^+|C^-) \times P(M_2^+|C^-) \times P(C^-)}
\]

\[
= \frac{Se_1 \times Se_2 \times P(C^+)}{Se_1 \times Se_2 \times P(C^+) + (1 - Sp_1) \times (1 - Sp_2) \times (1 - P(C^+))}
\]

where \( Se_1 = P(M_1^+|C^+) \), \( Se_2 = P(M_2^+|C^+) \), \( Sp_1 = P(M_1^-|C^-) \), \( Sp_2 = P(M_2^-|C^-) \), and \( A^+ \) indicates the presence, and \( A^- \) the absence, of variable \( A \). We calculated the conditional probability of contamination for HF183 and Mnif separately and for each combination of the two indicators (\( M_1^+, M_2^+; M_1^-, M_2^-; M_1^+, M_2^-; M_1^-, M_2^+ \)) by sample type.

S5. Difference-in-differences estimates

The effect of the intervention was estimated using difference-in-differences approaches. Crude DID estimates on the probability scale were calculated using 2000 bootstrap samples (Table S5). The DID estimate was calculated as

\[
DID = \left( E[Y_{T=1,P=1}] - E[Y_{T=1,P=0}] \right) - \left( E[Y_{T=1,P=1}] - E[Y_{T=1,P=0}] \right)
\]

where \( Y_{T,P} \) are the observed indicator values for treatment group \( T \) in study phase \( P \). The value of \( T \) is 0 for control compound observations and 1 for intervention compounds; likewise, \( P \) takes the value 0 for pre-treatment (baseline) observations and 1 post-treatment (follow-up). Model-based DID estimates were obtained using the product-term representation of the DID estimator in regression models, which permits the inclusion of additional covariates. We produced both crude model-based estimates, which only included terms for the DID estimator, as well as adjusted estimates that included additional terms for meteorological, compound, household, sample characteristics (Table S6).
| indicator | sample type | before | | | after | | | | | | | | control | N | estimate | control | N | estimate | control | N | estimate | control | N | estimate | control | N | estimate | control | N | estimate | control | N | estimate | control | N | estimate | control | N | estimate | control | N | estimate | control | N | estimate | control | N | estimate | control | N | estimate | control | N | estimate | control | N | estimate | control | N | estimate | control | N | estimate | control | N | estimate | control | N | estimate | control | N | estimate | control | N | estimate | control | N | estimate | control | N | estimate | control | N | estimate |
|-----------|-------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
Table S6. Model-based difference-in-differences estimates

<table>
<thead>
<tr>
<th>target</th>
<th>latrine entrance soil</th>
<th></th>
<th>household entrance soil</th>
<th></th>
<th>household stored water</th>
<th></th>
<th>food preparation surfaces</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>crude</td>
<td>adjusted</td>
<td>crude</td>
<td>adjusted</td>
<td>crude</td>
<td>adjusted</td>
<td>crude</td>
<td>adjusted</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>DID^c</td>
<td>N</td>
<td>DID</td>
<td>N</td>
<td>DID</td>
<td>N</td>
<td>DID</td>
</tr>
<tr>
<td>cEC</td>
<td>111</td>
<td>-0.37</td>
<td>95</td>
<td>-0.42</td>
<td>175</td>
<td>-0.34</td>
<td>146</td>
<td>0.05</td>
</tr>
<tr>
<td>EC23S</td>
<td>116</td>
<td>-0.84</td>
<td>98</td>
<td>-1.22</td>
<td>176</td>
<td>0.06</td>
<td>147</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>(-1.64, -0.02)</td>
<td>(-2.11, -0.30)</td>
<td>(-0.46, 0.57)</td>
<td>(-0.26, 1.01)</td>
<td>(-0.89, -0.04)</td>
<td>(-0.92, 0.12)</td>
<td>(-1.46, -0.19)</td>
<td>(-1.32, 0.19)</td>
</tr>
<tr>
<td>HF183</td>
<td>116</td>
<td>0.94</td>
<td>98</td>
<td>0.92</td>
<td>176</td>
<td>0.83</td>
<td>147</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>(0.40, 1.87)</td>
<td>(0.38, 1.87)</td>
<td>(0.39, 1.54)</td>
<td>(0.38, 1.80)</td>
<td>(0.48, 2.27)</td>
<td>(0.44, 2.15)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mnif</td>
<td>116</td>
<td>0.73</td>
<td>98</td>
<td>0.71</td>
<td>175</td>
<td>1.24</td>
<td>146</td>
<td>1.00</td>
</tr>
</tbody>
</table>

E. coli log_{10} concentration change (95% CI)

<table>
<thead>
<tr>
<th>target</th>
<th>latrine entrance soil</th>
<th></th>
<th>household entrance soil</th>
<th></th>
<th>household stored water</th>
<th></th>
<th>food preparation surfaces</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>crude</td>
<td>adjusted</td>
<td>crude</td>
<td>adjusted</td>
<td>crude</td>
<td>adjusted</td>
<td>crude</td>
<td>adjusted</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>DID^c</td>
<td>N</td>
<td>DID</td>
<td>N</td>
<td>DID</td>
<td>N</td>
<td>DID</td>
</tr>
<tr>
<td>cEC</td>
<td>111</td>
<td>-0.37</td>
<td>95</td>
<td>-0.42</td>
<td>175</td>
<td>-0.34</td>
<td>146</td>
<td>0.05</td>
</tr>
<tr>
<td>EC23S</td>
<td>116</td>
<td>-0.84</td>
<td>98</td>
<td>-1.22</td>
<td>176</td>
<td>0.06</td>
<td>147</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>(-1.64, -0.02)</td>
<td>(-2.11, -0.30)</td>
<td>(-0.46, 0.57)</td>
<td>(-0.26, 1.01)</td>
<td>(-0.89, -0.04)</td>
<td>(-0.92, 0.12)</td>
<td>(-1.46, -0.19)</td>
<td>(-1.32, 0.19)</td>
</tr>
<tr>
<td>HF183</td>
<td>116</td>
<td>0.94</td>
<td>98</td>
<td>0.92</td>
<td>176</td>
<td>0.83</td>
<td>147</td>
<td>0.90</td>
</tr>
<tr>
<td></td>
<td>(0.40, 1.87)</td>
<td>(0.38, 1.87)</td>
<td>(0.39, 1.54)</td>
<td>(0.38, 1.80)</td>
<td>(0.48, 2.27)</td>
<td>(0.44, 2.15)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mnif</td>
<td>116</td>
<td>0.73</td>
<td>98</td>
<td>0.71</td>
<td>175</td>
<td>1.24</td>
<td>146</td>
<td>1.00</td>
</tr>
</tbody>
</table>

a not adjusted for covariates; include terms for study phase, treatment arm, and their product, and compound-varying intercepts
b adjusted for animal presence, population density, household wealth, temperature, precipitation, and sample-specific variables: sun exposure and surface wetness for soils, presence of lid and width of container mouth for stored water, and food surface material.
c difference-in-differences estimated as the regression coefficient on the product of study phase and treatment arm indicators
S6. Human fecal indicator validation studies

Validation studies included in the performance meta-analysis were identified from Google Scholar records of articles citing the original assay publications and from the references of each published validation study identified.\textsuperscript{7,12,14} We only included HF183 studies that assessed the HF183/BFDRev assay or its modification used in this study, HF183/BacR287.\textsuperscript{7} We identified HF183 validation studies conducted on five continents, including two in Africa, seven in Asia, two in Australia, four in North America, and one in South America (Table S7). All identified Mnif studies used samples from the USA except for our study in Mozambique. The rurality of the settings from which samples were collected could not be determined for all studies and were more often specified for studies conducted in low- and middle-income countries. We used estimates that considered samples detected below the limit of quantification (DNQ) as negatives when separate estimates were reported by DNQ definition.\textsuperscript{15}
Table S7. Published validation studies of human fecal indicators HF183 and Mnif

<table>
<thead>
<tr>
<th>study</th>
<th>location</th>
<th>setting</th>
<th>human samples</th>
<th>non-human samples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>N positive</td>
<td>sensitivity</td>
</tr>
<tr>
<td>HF183</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Mozambique</td>
<td>urban</td>
<td>14</td>
<td>9</td>
</tr>
<tr>
<td>2</td>
<td>Bangladesh</td>
<td>urban</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>Bangladesh</td>
<td>rural</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>India</td>
<td>rural</td>
<td>35</td>
<td>10(^a)</td>
</tr>
<tr>
<td>5</td>
<td>Kenya</td>
<td>rural</td>
<td>17</td>
<td>11</td>
</tr>
<tr>
<td>6</td>
<td>Thailand</td>
<td></td>
<td>28</td>
<td>25</td>
</tr>
<tr>
<td>7</td>
<td>Costa Rica</td>
<td></td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>8</td>
<td>Singapore</td>
<td>urban</td>
<td>56</td>
<td>42</td>
</tr>
<tr>
<td>9</td>
<td>Nepal</td>
<td></td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>10</td>
<td>USA</td>
<td>rural</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>11</td>
<td>Japan</td>
<td></td>
<td>20</td>
<td>20</td>
</tr>
<tr>
<td>12</td>
<td>Australia</td>
<td></td>
<td>32</td>
<td>24</td>
</tr>
<tr>
<td>13</td>
<td>USA</td>
<td></td>
<td></td>
<td>184</td>
</tr>
<tr>
<td>14</td>
<td>Australia</td>
<td></td>
<td></td>
<td>60(^a)</td>
</tr>
<tr>
<td>15</td>
<td>Peru</td>
<td>peri-urban</td>
<td>30</td>
<td>23</td>
</tr>
<tr>
<td>Mnif</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Mozambique</td>
<td>urban</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>2</td>
<td>USA</td>
<td></td>
<td>60</td>
<td>36(^a)</td>
</tr>
<tr>
<td>3</td>
<td>Indiana</td>
<td></td>
<td>59</td>
<td>40</td>
</tr>
<tr>
<td>4</td>
<td>Mississippi</td>
<td></td>
<td>62</td>
<td>51</td>
</tr>
</tbody>
</table>

\(^a\) samples below the limit of quantification considered negative

S7. Accuracy-adjusted intervention effect estimates

We constructed a series of models to estimate sanitation intervention effects on the prevalence of human fecal contamination. In the first model, human-associated fecal indicators were used as direct proxies for human fecal contamination, equivalent to assuming 100% sensitivity and specificity. The estimated intervention effect, given by the DID prevalence odds ratio, was less than one for both indicators in latrine soil and household soil and above one for HF183 in stored water (Figure S1). However, the estimates were imprecise and encompassed both large increases and large reductions within the 95% CIs, providing little evidence for an effect of the intervention on either indicator in any sample type. Incorporating local validation data to account for indicator accuracy (Model 2) further widened the 95% CIs, indicating...
additional uncertainty about the intervention effect on the prevalence of human fecal contamination. Effect estimates incorporating multiple validation datasets (Model 3) were largely similar to those using local validation data alone. When using HF183 and Mnif to jointly estimate human fecal contamination (Model 4), the DID estimates were similar to estimates for individual indicators. Similarly, combining observations from all three sample types produced an estimated intervention effect close to the null with the 95% CI encompassing both reductions and increases in the odds of human fecal contamination at the compound level [POR: 0.93 (95% CI: 0.42-2.1)]. Across all model formulations, there was little evidence for an effect of the sanitation intervention on the prevalence of human fecal contamination.

Figure S1. Difference-in-difference prevalence odds ratio (POR) estimates of the sanitation intervention effect on human fecal contamination under five different models. Model 1 made no correction for indicator accuracy, Model 2 used local validation data to account for indicator sensitivity and specificity, and Models 3, 4 and 5 used meta-analytic estimates of local indicator accuracy. Model 3 was fit separately by indicator and sample type, Model 4 was fit to both indicators by sample type, and Model 5 used both indicators in all sample types to estimate the latent compound prevalence of human fecal contamination. All models were adjusted for population density, presence of animals, wealth score, temperature, antecedent precipitation, and sun exposure and surface wetness for soil samples and storage container mouth width and cover status for water samples.
S8. Human fecal contamination prevalence estimates

Posterior predictions from each of the five models were used to estimate stratum-specific prevalence. We fit unadjusted models that included only intercepts and the DID terms (Table S8) as well as adjusted models with meteorological, compound, household, and sample characteristics included as covariates (Table 2, main text).
Table S8. Estimated sensitivity, specificity, and prevalence of human fecal contamination from unadjusted models

<table>
<thead>
<tr>
<th>Target</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>N</th>
<th>Prevalence Estimate (95% CI)</th>
<th>Prevalence DID (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latrine soil</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bootstrap</td>
<td>HF183</td>
<td>0.59 (0.41, 0.80)</td>
<td>65</td>
<td>0.41 (0.05, 0.90)</td>
<td>-0.09 (-0.27, 0.09)</td>
</tr>
<tr>
<td></td>
<td>Mnif</td>
<td>0.68 (0.50, 0.86)</td>
<td>65</td>
<td>0.53 (0.11, 0.92)</td>
<td>-0.05 (-0.25, 0.14)</td>
</tr>
<tr>
<td>model 1</td>
<td>HF183</td>
<td>0.65 (0.45, 0.85)</td>
<td>65</td>
<td>0.39 (0.06, 0.87)</td>
<td>-0.01 (-0.19, 0.18)</td>
</tr>
<tr>
<td></td>
<td>Mnif</td>
<td>0.70 (0.55, 0.83)</td>
<td>65</td>
<td>0.54 (0.17, 0.87)</td>
<td>-0.05 (-0.25, 0.14)</td>
</tr>
<tr>
<td>model 2</td>
<td>HF183</td>
<td>0.64 (0.45, 0.84)</td>
<td>65</td>
<td>0.47 (0.16, 0.79)</td>
<td>-0.06 (-0.25, 0.13)</td>
</tr>
<tr>
<td></td>
<td>Mnif</td>
<td>0.71 (0.56, 0.84)</td>
<td>65</td>
<td>0.48 (0.25, 0.73)</td>
<td>0.04 (-0.14, 0.23)</td>
</tr>
<tr>
<td>model 3</td>
<td>HF183</td>
<td>0.72 (0.55, 0.87)</td>
<td>65</td>
<td>0.49 (0.27, 0.72)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mnif</td>
<td>0.71 (0.58, 0.84)</td>
<td>65</td>
<td>0.48 (0.25, 0.73)</td>
<td></td>
</tr>
<tr>
<td>model 4</td>
<td>HF183</td>
<td>0.66 (0.55, 0.83)</td>
<td>65</td>
<td>0.47 (0.16, 0.79)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mnif</td>
<td>0.71 (0.56, 0.84)</td>
<td>65</td>
<td>0.48 (0.25, 0.73)</td>
<td></td>
</tr>
<tr>
<td>household soil</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bootstrap</td>
<td>HF183</td>
<td>0.58 (0.36, 0.80)</td>
<td>71</td>
<td>0.27 (0.18, 0.37)</td>
<td>-0.05 (-0.21, 0.11)</td>
</tr>
<tr>
<td></td>
<td>Mnif</td>
<td>0.65 (0.41, 0.86)</td>
<td>71</td>
<td>0.35 (0.30, 0.53)</td>
<td>-0.05 (-0.21, 0.11)</td>
</tr>
<tr>
<td>model 1</td>
<td>HF183</td>
<td>0.65 (0.40, 0.85)</td>
<td>71</td>
<td>0.23 (0.09, 0.37)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mnif</td>
<td>0.69 (0.52, 0.84)</td>
<td>71</td>
<td>0.29 (0.19, 0.40)</td>
<td></td>
</tr>
<tr>
<td>model 2</td>
<td>HF183</td>
<td>0.67 (0.44, 0.86)</td>
<td>71</td>
<td>0.15 (0.02, 0.35)</td>
<td>-0.01 (-0.12, 0.11)</td>
</tr>
<tr>
<td></td>
<td>Mnif</td>
<td>0.69 (0.52, 0.83)</td>
<td>71</td>
<td>0.24 (0.09, 0.43)</td>
<td></td>
</tr>
<tr>
<td>model 4</td>
<td>HF183</td>
<td>0.72 (0.55, 0.87)</td>
<td>71</td>
<td>0.24 (0.09, 0.43)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mnif</td>
<td>0.71 (0.58, 0.84)</td>
<td>71</td>
<td>0.24 (0.09, 0.43)</td>
<td></td>
</tr>
<tr>
<td>stored water</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bootstrap</td>
<td>HF183</td>
<td>0.72 (0.55, 0.87)</td>
<td>93</td>
<td>0.36 (0.20, 0.52)</td>
<td>-0.30 (-0.57, -0.01)</td>
</tr>
<tr>
<td>model 1</td>
<td>HF183</td>
<td>0.65 (0.40, 0.85)</td>
<td>93</td>
<td>0.23 (0.09, 0.37)</td>
<td>-0.20 (-0.07, 0.46)</td>
</tr>
<tr>
<td>model 2</td>
<td>HF183</td>
<td>0.67 (0.44, 0.86)</td>
<td>93</td>
<td>0.20 (0.06, 0.20)</td>
<td>-0.01 (-0.21, 0.19)</td>
</tr>
<tr>
<td>model 3</td>
<td>HF183</td>
<td>0.69 (0.52, 0.84)</td>
<td>93</td>
<td>0.14 (0.08, 0.23)</td>
<td>-0.01 (-0.21, 0.19)</td>
</tr>
<tr>
<td>model 4</td>
<td>HF183</td>
<td>0.72 (0.55, 0.87)</td>
<td>93</td>
<td>0.14 (0.08, 0.23)</td>
<td>-0.01 (-0.21, 0.19)</td>
</tr>
<tr>
<td>latency compound</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>m5</td>
<td>HF183</td>
<td>0.72 (0.55, 0.87)</td>
<td>113</td>
<td>0.26 (0.08, 0.55)</td>
<td>0.03 (-0.11, 0.18)</td>
</tr>
<tr>
<td></td>
<td>Mnif</td>
<td>0.71 (0.58, 0.84)</td>
<td>113</td>
<td>0.26 (0.08, 0.55)</td>
<td>0.03 (-0.11, 0.18)</td>
</tr>
</tbody>
</table>
S9. Stan code

S9.1. Model 1

// Model 1:
// single sample type, single target
// no correction for sensitivity/specificity
// no compound-varying intercepts

data{
  // sample data
  int<lower = 0> N_samp; // number of sample observations
  int<lower = 0, upper = 1> y_samp[N_samp]; // sample observations
  vector<lower = 0, upper = 1>[N_samp] phase; // survey phase
  vector<lower = 0, upper = 1>[N_samp] treat; // treatment arm
  vector<lower = 0, upper = 1>[N_samp] did; // phase*treat interaction
  int<lower = 0> K; // number of predictors
  matrix[N_samp, K] X; // predictor values

  // prior predictive check?
  int<lower = 0, upper = 1> prior_only; // toggle on to sample priors only
}

parameters{
  // linear model parameters
  real b0; // intercept
  real bP; // phase
  real bT; // treatment
  real bD; // DID
  vector[K] g; // predictor coefficients
}

model{
  // linear model for probability of human contamination
  vector[N_samp] p_samp = inv_logit(b0 + bP*phase + bT*treat + bD*did + X*g);

  // Likelihood
  if(prior_only == 0){
    y_samp ~ binomial(1, p_samp);
  }

  // linear model priors
  b0 ~ normal(0, 1.5);
  bD ~ normal(0, 0.5);
  bT ~ normal(0, 0.5);
  bP ~ normal(0, 0.5);
  g ~ normal(0, 0.5);
}

generated quantities{
  // posterior predictions (or prior predictions, if prior_only == 1)
  // define predicted variables
  int<lower = 0, upper = 1> y_pred[N_samp]; // predicted sample observations
  int<lower = 0> n_pos; // number of predicted positives
  real<lower = 0, upper = 1> p_samp_avg; // mean target prevalence in samples
// calculate human contamination probability
    vector[N_samp] p_samp_sim = inv_logit(b0 + bP*phase + bT*treat + bD*did + X*g);

// predict sample observations
    y_pred = binomial_rng(1, p_samp_sim);
    n_pos = sum(y_pred);

// summarise prevalence calculations
    p_samp_avg = mean(p_samp_sim);
}
S9.2. Model 2

// Model 2:
// single sample type, single target
// diagnostic performance data from this study only
// no compound-varying intercepts

data{
  // sample data
  int<lower = 0> N_samp; // number of sample observations
  int<lower = 0, upper = 1> y_samp[N_samp]; // sample observations
  vector<lower = 0, upper = 1>[N_samp] phase; // survey phase
  vector<lower = 0, upper = 1>[N_samp] treat; // treatment arm
  vector<lower = 0, upper = 1>[N_samp] did; // phase*treat interaction
  int<lower = 0> K; // number of predictors
  matrix[N_samp, K] X; // predictor values

  // sens/spec data
  int<lower = 0> y_spec; // number of true negatives observed
  int<lower = 0> n_spec; // number of non-target samples
  int<lower = 0> y_sens; // number of true positives observed
  int<lower = 0> n_sens; // number of target samples

  // prior predictive check?
  int<lower = 0, upper = 1> prior_only; // toggle on to sample priors only
}

parameters{
  // linear model parameters
  real b0; // intercept
  real bP; // phase
  real bT; // treatment
  real bD; // DID
  vector[K] g; // predictor coefficients

  // diagnostic performance parameters
  real<lower=0, upper=1> spec; // specificity
  real<lower=0, upper=1> sens; // sensitivity
}

model{
  // linear model for probability of human contamination
  vector[N_samp] p = inv_logit(b0 + bP*phase + bT*treat + bD*did + X*g);

  // adjust for sens/spec
  vector[N_samp] p_samp = sens * p + (1 - spec) * (1 - p);

  // Likelihoods
  if(prior_only == 0){
    // samples
    y_samp ~ binomial(1, p_samp);

    // validation studies
    y_spec ~ binomial(n_spec, spec);
    y_sens ~ binomial(n_sens, sens);
  }
}
// linear model priors
b0 ~ normal(0, 1.5);
bD ~ normal(0, 0.5);
bT ~ normal(0, 0.5);
bP ~ normal(0, 0.5);
g ~ normal(0, 0.5);

// validation priors
sens ~ beta(3, 2);
spec ~ beta(3, 2);
}
generated quantities{
// posterior predictions (or prior predictions, if prior_only == 1)
// define predicted variables
int<lower = 0, upper = 1> y_pred[N_samp]; // predicted sample observations
int<lower = 0> n_pos; // number of predicted positives
real<lower = 0, upper = 1> p_avg; // mean human contamination prevalence
real<lower = 0, upper = 1> p_samp_avg; // mean target prevalence in samples

// calculate human contamination probability
vector[N_samp] p_sim = inv_logit(b0 + bP*phase + bT*treat + bD*did + X*g);

// adjust for sens/spec
vector[N_samp] p_samp_sim = sens * p_sim + (1 - spec) * (1 - p_sim);

// predict sample observations
y_pred = binomial_rng(1, p_samp_sim);
n_pos = sum(y_pred);

// summarise prevalence calculations
p_avg = mean(p_sim);
p_samp_avg = mean(p_samp_sim);
}
S9.3. Model 3

// Model 3:
// single sample type, single target
// diagnostic performance meta-analysis
// no compound-varying intercepts

data{
// sample data
int<lower = 0> N_samp; // number of sample observations
int<lower = 0, upper = 1> y_samp[N_samp]; // sample observations
vector<lower = 0, upper = 1>[N_samp] phase; // survey phase
vector<lower = 0, upper = 1>[N_samp] treat; // treatment arm
vector<lower = 0, upper = 1>[N_samp] did; // phase*treat interaction
int<lower = 0> K; // number of predictors
matrix[N_samp, K] X; // predictor values

// sens/spec data
int<lower = 0> J_spec;
int<lower = 0> y_spec[J_spec];
int<lower = 0> n_spec[J_spec];
int<lower = 0> J_sens;
int<lower = 0> y_sens[J_sens];
int<lower = 0> n_sens[J_sens];

// prior predictive check?
int<lower = 0, upper = 1> prior_only; // toggle on to sample priors only
}

parameters{
// linear model parameters
real b0; // intercept
real bP; // phase
real bT; // treatment
real bD; // DID
vector[K] g; // predictor coefficients

// sens/spec meta-analysis parameters
real mu_logit_spec; // mean spec on the logit scale
real mu_logit_sens;
real<lower = 0> sigma_logit_spec; // spec SD on logit scale
real<lower = 0> sigma_logit_sens; // non-centered parameterization of logit-transformed sens/spec
vector<offset = mu_logit_spec, multiplier = sigma_logit_spec>[J_spec] logit_spec;
vector<offset = mu_logit_sens, multiplier = sigma_logit_sens>[J_sens] logit_sens;
}

transformed parameters{
// recover sens/spec on probability scale
vector[J_spec] spec = inv_logit(logit_spec);
vector[J_sens] sens = inv_logit(logit_sens);
}

model{
// linear model for probability of human contamination
vector[N_samp] p = inv_logit(b0 + bP*phase + bT*treat + bD*did + X*g);
}
// adjust for sens/spec
vector[N_samp] p_samp = sens[1] * p + (1 - spec[1]) * (1 - p);

// Likelihoods
if(prior_only == 0){
  // samples
  y_samp ~ binomial(1, p_samp);
  // validation studies
  y_spec ~ binomial(n_spec, spec);
  y_sens ~ binomial(n_sens, sens);
}

// linear model priors
b0 ~ normal(0, 1.5);
bD ~ normal(0, 0.5);
bT ~ normal(0, 0.5);
bP ~ normal(0, 0.5);
g ~ normal(0, 0.5);

// validation priors
logit_spec ~ normal(mu_logit_spec, sigma_logit_spec);
logit_sens ~ normal(mu_logit_sens, sigma_logit_sens);
sigma_logit_spec ~ normal(0, .5);
sigma_logit_sens ~ normal(0, .5);
mu_logit_spec ~ normal(.5, 1);
mu_logit_sens ~ normal(.5, 1);
}

generated quantities{
  // posterior predictions (or prior predictions, if prior_only == 1)
  // define predicted variables
  int<lower = 0, upper = 1> y_pred[N_samp];  // predicted sample observations
  int<lower = 0> n_pos;  // number of predicted positives
  real<lower = 0, upper = 1> p_avg;  // mean human contamination prevalence
  real<lower = 0, upper = 1> p_samp_avg;  // mean target prevalence in samples

  // calculate human contamination probability
  vector[N_samp] p_sim = inv_logit(b0 + bP*phase + bT*treat + bD*did + X*g);
  // adjust for sens/spec
  vector[N_samp] p_samp_sim = sens[1] * p_sim + (1 - spec[1]) * (1 - p_sim);

  // predict sample observations
  y_pred = binomial_rng(1, p_samp_sim);
  n_pos = sum(y_pred);

  // summarise prevalence calculations
  p_avg = mean(p_sim);
  p_samp_avg = mean(p_samp_sim);
}
S9.4. Model 4

// Model 4:
// single sample type, two targets
// diagnostic performance meta-analysis
// no compound-varying intercepts

data{
  // sample data
  int<lower = 0> N_samp;  // number of sample observations
  int<lower = 0, upper = 1> y_hf[N_samp];  // HF183 observations
  int<lower = 0, upper = 1> y_mn[N_samp];  // Mnif observations
  vector<lower = 0, upper = 1>[N_samp] phase;  // survey phase
  vector<lower = 0, upper = 1>[N_samp] treat;  // treatment arm
  vector<lower = 0, upper = 1>[N_samp] did;  // phase*treat interaction
  int<lower = 0> K;  // number of predictors
  matrix[N_samp, K] X;  // predictor values

  // sens/spec data
  int<lower = 0> J_spec_hf;
  int<lower = 0> J_sens_hf;
  int<lower = 0> J_spec_mn;
  int<lower = 0> J_sens_mn;

  // prior predictive check?
  int<lower = 0, upper = 1> prior_only;  // toggle on to sample priors only
}

parameters{
  // linear model parameters
  real b0;  // intercept
  real bP;  // phase
  real bT;  // treatment
  real bD;  // DID
  vector[K] g;  // predictor coefficients

  // sens/spec meta-analysis parameters
  real mu_logit_spec_hf;  // mean spec for HF183 on the logit scale
  real mu_logit_sens_hf;
  real mu_logit_spec_mn;
  real mu_logit_sens_mn;
  real<lower = 0> sigma_logit_spec_hf;  // spec SD on logit scale
  real<lower = 0> sigma_logit_sens_hf;
  real<lower = 0> sigma_logit_spec_mn;
  real<lower = 0> sigma_logit_sens_mn;
  // non-centered parameterization of logit-transformed sens/spec for each target
  vector<offset = mu_logit_spec_hf, multiplier = sigma_logit_spec_hf>[J_spec_hf] logit_spec_hf;
}
vector<offset = mu_logit_sens_hf, multiplier = sigma_logit_sens_hf>[J_sens_hf]
logit_sens_hf;
vector<offset = mu_logit_spec_mn, multiplier = sigma_logit_spec_mn>[J_spec_mn]
logit_spec_mn;
vector<offset = mu_logit_sens_mn, multiplier = sigma_logit_sens_mn>[J_sens_mn]
logit_sens_mn;
}
transformed parameters{
// recover sens/spec on probability scale
    vector[J_spec_hf] spec_hf = inv_logit(logit_spec_hf);
    vector[J_sens_hf] sens_hf = inv_logit(logit_sens_hf);
    vector[J_spec_mn] spec_mn = inv_logit(logit_spec_mn);
    vector[J_sens_mn] sens_mn = inv_logit(logit_sens_mn);
}
model{
// linear model for probability of human contamination
    vector[N_samp] p = inv_logit(b0 + bP*phase + bT*treat + bD*did + X*g);

// adjust for sens/spec
    // by convention the first sens/spec element represents this current study
    vector[N_samp] p_hf = sens_hf[1] * p + (1 - spec_hf[1]) * (1 - p);
    vector[N_samp] p_mn = sens_mn[1] * p + (1 - spec_mn[1]) * (1 - p);

// Likelihoods
    if(prior_only == 0){
        // samples
        y_hf ~ binomial(1, p_hf);
        y_mn ~ binomial(1, p_mn);

        // validation studies
        y_spec_hf ~ binomial(n_spec_hf, spec_hf);
        y_sens_hf ~ binomial(n_sens_hf, sens_hf);
        y_spec_mn ~ binomial(n_spec_mn, spec_mn);
        y_sens_mn ~ binomial(n_sens_mn, sens_mn);
    }

// linear model priors
    b0 ~ normal(0, 1.5);
    bD ~ normal(0, 0.5);
    bT ~ normal(0, 0.5);
    bP ~ normal(0, 0.5);
    g ~ normal(0, 0.5);

// validation priors
    logit_spec_hf ~ normal(mu_logit_spec_hf, sigma_logit_spec_hf);
    logit_sens_hf ~ normal(mu_logit_sens_hf, sigma_logit_sens_hf);
    sigma_logit_spec_hf ~ normal(0, .5);
    sigma_logit_sens_hf ~ normal(0, .5);
    mu_logit_spec_hf ~ normal(.5, 1);
    mu_logit_sens_hf ~ normal(.5, 1);
    logit_spec_mn ~ normal(mu_logit_spec_mn, sigma_logit_spec_mn);
    logit_sens_mn ~ normal(mu_logit_sens_mn, sigma_logit_sens_mn);
    sigma_logit_spec_mn ~ normal(0, .5);
    sigma_logit_sens_mn ~ normal(0, .5);
    mu_logit_spec_mn ~ normal(.5, 1);
mu_logit_sens_mn ~ normal(.5, 1);
}

generated quantities{
// posterior predictions (or prior predictions, if prior_only == 1)
// define predicted variables
    int<lower = 0, upper = 1> y_pred_hf[N_samp]; // predicted HF183 observations
    int<lower = 0> n_pos_hf; // number of predicted HF183 positives
    real<lower = 0, upper = 1> p_samp_avg_hf; // mean HF183 prevalence in samples
    int<lower = 0, upper = 1> y_pred_mn[N_samp]; // predicted Mnif observations
    int<lower = 0> n_pos_mn; // number of predicted Mnif positives
    real<lower = 0, upper = 1> p_samp_avg_mn; // mean Mnif prevalence in samples
    real<lower = 0, upper = 1> p_avg; // mean human contamination prevalence

// calculate human contamination probability
    vector[N_samp] p_sim = inv_logit(b0 + bP*phase + bT*treat + bD*did + X*g);

// adjust for sens/spec
    vector[N_samp] p_hf_sim = sens_hf[1] * p_sim + (1 - spec_hf[1]) * (1 - p_sim);
    vector[N_samp] p_mn_sim = sens_mn[1] * p_sim + (1 - spec_mn[1]) * (1 - p_sim);

// predict sample observations
    y_pred_hf = binomial_rng(1, p_hf_sim);
    n_pos_hf = sum(y_pred_hf);
    y_pred_mn = binomial_rng(1, p_mn_sim);
    n_pos_mn = sum(y_pred_mn);

// summarise prevalence calculations
    p_avg = mean(p_sim);
    p_samp_avg_hf = mean(p_hf_sim);
    p_samp_avg_mn = mean(p_mn_sim);
S9.5. Model 5

// Model 5:
// three sample types, two targets
// diagnostic performance meta-analysis
// compound-varying intercept, type-varying intercept

data{
  int<lower = 0> N_type; // number of sample types considered
}

// compound data
int<lower = 0> J_comp; // number of unique compounds
int<lower = 0> N_comp; // number of compound observations
int<lower = 1, upper = J_comp> comp[N_comp]; // compound index
int<lower = 0> K_comp; // number of compound-level predictors
matrix[N_comp, K_comp] X_comp; // compound-level predictors

// hw sample data
int<lower = 0> N_hw; // number of compound observations
int<lower = 0, upper = 1> y_hw_hf[N_hw]; // hw HF183 observations
// int<lower = 0, upper = 1> y_hw_mn[N_hw]; // no hw Mnif observations
int<lower = 1, upper = J_comp> comp_hw[N_hw]; // hw compound index
int<lower = 0> K_hw; // number of hw sample-level predictors
matrix[N_hw, K_hw] X_hw; // hw predictors

// ds sample data
int<lower = 0> N_ds; // number of compound observations
int<lower = 0, upper = 1> y_ds_hf[N_ds]; // ds HF183 observations
int<lower = 0, upper = 1> y_ds_mn[N_ds]; // ds Mnif observations
int<lower = 1, upper = J_comp> comp_ds[N_ds]; // ds compound index
int<lower = 0> K_ds; // number of ds sample-level predictors
matrix[N_ds, K_ds] X_ds; // ds predictors

// ls sample data
int<lower = 0> N_ls; // number of compound observations
int<lower = 0, upper = 1> y_ls_hf[N_ls]; // ls HF183 observations
int<lower = 0, upper = 1> y_ls_mn[N_ls]; // ls Mnif observations
int<lower = 1, upper = J_comp> comp_ls[N_ls]; // ls compound index
int<lower = 0> K_ls; // number of ls sample-level predictors
matrix[N_ls, K_ls] X_ls; // ls predictors

// sens/spec data
int<lower = 0> J_sens_hf;
int<lower = 0> y_sens_hf[J_sens_hf];
int<lower = 0> n_sens_hf[J_sens_hf];
int<lower = 0> J_sens_mn;
int<lower = 0> y_sens_mn[J_sens_mn];
int<lower = 0> n_sens_mn[J_sens_mn];

// prior predictive check?
int<lower = 0, upper = 1> prior_only;
}

parameters{
  // compound prevalence linear model parameters
  vector[K_comp] b_comp; // compound-level coefficients
  real mu_comp; // mean of compound-varying intercept
  real<lower = 0> sigma_comp; // SD of compound-varying intercept
  vector<offset = mu_comp, multiplier = sigma_comp>[J_comp] a_comp;

  // sample-level parameters
  vector[K_hw] g_hw; // hw predictor coefficients
  vector[K_ds] g_ds; // ds predictor coefficients
  vector[K_ls] g_ls; // ls predictor coefficients
  real<lower = 0> sigma_type; // SD of sample differences
  vector<multiplier = sigma_type>[N_type] a_type; // sample type-varying intercept

  // sens/spec meta-analysis parameters
  real mu_logit_spec_hf;
  real mu_logit_sens_hf;
  real mu_logit_spec_mn;
  real mu_logit_sens_mn;
  real<lower = 0> sigma_logit_spec_hf;
  real<lower = 0> sigma_logit_sens_hf;
  real<lower = 0> sigma_logit_spec_mn;
  real<lower = 0> sigma_logit_sens_mn;
  vector<offset = mu_logit_spec_hf, multiplier = sigma_logit_spec_hf>[J_spec_hf] logit_spec_hf;
  vector<offset = mu_logit_sens_hf, multiplier = sigma_logit_sens_hf>[J_sens_hf] logit_sens_hf;
  vector<offset = mu_logit_spec_mn, multiplier = sigma_logit_spec_mn>[J_spec_mn] logit_spec_mn;
  vector<offset = mu_logit_sens_mn, multiplier = sigma_logit_sens_mn>[J_sens_mn] logit_sens_mn;
}

transformed parameters{
  vector[J_spec_hf] spec_hf = inv_logit(logit_spec_hf);
  vector[J_sens_hf] sens_hf = inv_logit(logit_sens_hf);
  vector[J_spec_mn] spec_mn = inv_logit(logit_spec_mn);
  vector[J_sens_mn] sens_mn = inv_logit(logit_sens_mn);
}

model{
  // linear model for compound contamination
  vector[N_comp] logit_p_comp = a_comp[comp] + X_comp * b_comp;

  // linear models for sample-type specific prevalence
  vector[N_hw] p_hw = inv_logit(logit_p_comp[comp_hw] + a_type[1] + X_hw * g_hw);
  vector[N_ds] p_ds = inv_logit(logit_p_comp[comp_ds] + a_type[2] + X_ds * g_ds);
  vector[N_ds] p_ls = inv_logit(logit_p_comp[comp_ls] + a_type[3] + X_ls * g_ls);

  // adjust for sens/spec
  vector[N_hw] p_hw_hf = sens_hf[1] * p_hw + (1 - spec_hf[1]) * (1 - p_hw);
  vector[N_hw] p_hw_mn = sens_mn[1] * p_hw + (1 - spec_mn[1]) * (1 - p_hw);
  vector[N_ds] p_ds_hf = sens_hf[1] * p_ds + (1 - spec_hf[1]) * (1 - p_ds);
  vector[N_ds] p_ds_mn = sens_mn[1] * p_ds + (1 - spec_mn[1]) * (1 - p_ds);
vector[N_ls] p_ls_hf = sens_hf[1] * p_ls + (1 - spec_hf[1]) * (1 - p_ls);
vector[N_ls] p_ls_mn = sens_mn[1] * p_ls + (1 - spec_mn[1]) * (1 - p_ls);

// Likelihoods
if(prior_only == 0){
// samples
  y_hw_hf ~ binomial(1, p_hw_hf);
  // y_hw_mn ~ binomial(1, p_hw_mn); no Mnif for HW samples
  y_ds_hf ~ binomial(1, p_ds_hf);
  y_ds_mn ~ binomial(1, p_ds_mn);
  y_ls_hf ~ binomial(1, p_ls_hf);
  y_ls_mn ~ binomial(1, p_ls_mn);

// validation studies
  y_spec_hf ~ binomial(n_spec_hf, spec_hf);
  y_sens_hf ~ binomial(n_sens_hf, sens_hf);
  y_spec_mn ~ binomial(n_spec_mn, spec_mn);
  y_sens_mn ~ binomial(n_sens_mn, sens_mn);
}

// sample priors
// compound-level
  a_comp ~ normal(mu_comp, sigma_comp);
  mu_comp ~ normal(0, 1.5);
  sigma_comp ~ normal(0, 0.5);
  b_comp ~ normal(0, 0.5);

// sample-level
  a_type ~ normal(0, sigma_type);
  sigma_type ~ normal(0, 0.5);
  g_hw ~ normal(0, 0.5);
  g_ds ~ normal(0, 0.5);
  g_ls ~ normal(0, 0.5);

// validation priors
  logit_spec_hf ~ normal(mu_logit_spec_hf, sigma_logit_spec_hf);
  logit_sens_hf ~ normal(mu_logit_sens_hf, sigma_logit_sens_hf);
  sigma_logit_spec_hf ~ normal(0, .5);
  sigma_logit_sens_hf ~ normal(0, .5);
  mu_logit_spec_hf ~ normal(.5, 1);
  mu_logit_sens_hf ~ normal(.5, 1);
  logit_spec_mn ~ normal(mu_logit_spec_mn, sigma_logit_spec_mn);
  logit_sens_mn ~ normal(mu_logit_sens_mn, sigma_logit_sens_mn);
  sigma_logit_spec_mn ~ normal(0, .5);
  sigma_logit_sens_mn ~ normal(0, .5);
  mu_logit_spec_mn ~ normal(.5, 1);
  mu_logit_sens_mn ~ normal(.5, 1);
}

generated quantities{
// simulated sample outcome containers
  int y_hw_hf_sim[N_hw];
  int y_hw_mn_sim[N_hw];
  int y_ds_hf_sim[N_ds];
  int y_ds_mn_sim[N_ds];
  int y_ls_hf_sim[N_ls];
  int y_ls_mn_sim[N_ls];
int n_pos_hw_hf;
int n_pos_hw_mn;
int n_pos_ds_hf;
int n_pos_ds_mn;
int n_pos_ls_hf;
int n_pos_ls_mn;
real<lower = 0, upper = 1> p_avg_comp;
real<lower = 0, upper = 1> p_avg_hw;
real<lower = 0, upper = 1> p_avg_ds;
real<lower = 0, upper = 1> p_avg_ls;
real<lower = 0, upper = 1> p_samp_avg_hw_hf;
real<lower = 0, upper = 1> p_samp_avg_hw_mn;
real<lower = 0, upper = 1> p_samp_avg_ds_hf;
real<lower = 0, upper = 1> p_samp_avg_ds_mn;
real<lower = 0, upper = 1> p_samp_avg_ls_hf;
real<lower = 0, upper = 1> p_samp_avg_ls_mn;

// simulate type-specific probabilities
vector[N_comp] logit_p_comp_sim = a_comp[comp] + X_comp * b_comp;
vector[N_comp] p_comp_sim = inv_logit(logit_p_comp_sim);
vector[N_hw] p_hw_sim = inv_logit(logit_p_comp_sim[comp_hw] + a_type[1] + X_hw * g_hw);
vector[N_ds] p_ds_sim = inv_logit(logit_p_comp_sim[comp_ds] + a_type[2] + X_ds * g_ds);
vector[N_ls] p_ls_sim = inv_logit(logit_p_comp_sim[comp_ds] + a_type[3] + X_ls * g_ls);

// adjust simulations for sens/spec
vector[N_hw] p_hw_hf_sim = sens_hf[1] * p_hw_sim + (1 - spec_hf[1]) * (1 - p_hw_sim);
vector[N_hw] p_hw_mn_sim = sens_mn[1] * p_hw_sim + (1 - spec_mn[1]) * (1 - p_hw_sim);
vector[N_ds] p_ds_hf_sim = sens_hf[1] * p_ds_sim + (1 - spec_hf[1]) * (1 - p_ds_sim);
vector[N_ds] p_ds_mn_sim = sens_mn[1] * p_ds_sim + (1 - spec_mn[1]) * (1 - p_ds_sim);
vector[N_ls] p_ls_hf_sim = sens_hf[1] * p_ls_sim + (1 - spec_hf[1]) * (1 - p_ls_sim);
vector[N_ls] p_ls_mn_sim = sens_mn[1] * p_ls_sim + (1 - spec_mn[1]) * (1 - p_ls_sim);

// simulate sample observations
y_hw_hf_sim = binomial_rng(1, p_hw_hf_sim);
y_hw_mn_sim = binomial_rng(1, p_hw_mn_sim);
y_ds_hf_sim = binomial_rng(1, p_ds_hf_sim);
y_ds_mn_sim = binomial_rng(1, p_ds_mn_sim);
y_ls_hf_sim = binomial_rng(1, p_ls_hf_sim);
y_ls_mn_sim = binomial_rng(1, p_ls_mn_sim);

// summarize simulated samples
n_pos_hw_hf = sum(y_hw_hf_sim);
n_pos_hw_mn = sum(y_hw_mn_sim);
n_pos_ds_hf = sum(y_ds_hf_sim);
n_pos_ds_mn = sum(y_ds_mn_sim);
n_pos_ls_hf = sum(y_ls_hf_sim);
n_pos_ls_mn = sum(y_ls_mn_sim);

// mean prevalence predictions
\[
\begin{align*}
p_{\text{avg comp}} &= \text{mean}(p_{\text{comp sim}}); \\
p_{\text{avg hw}} &= \text{mean}(p_{\text{hw sim}}); \\
p_{\text{avg ds}} &= \text{mean}(p_{\text{ds sim}}); \\
p_{\text{avg ls}} &= \text{mean}(p_{\text{ls sim}}); \\
p_{\text{samp avg hw hf}} &= \text{mean}(p_{\text{hw hf sim}}); \\
p_{\text{samp avg hw mn}} &= \text{mean}(p_{\text{hw mn sim}}); \\
p_{\text{samp avg ds hf}} &= \text{mean}(p_{\text{ds hf sim}}); \\
p_{\text{samp avg ds mn}} &= \text{mean}(p_{\text{ds mn sim}}); \\
p_{\text{samp avg ls hf}} &= \text{mean}(p_{\text{ls hf sim}}); \\
p_{\text{samp avg ls mn}} &= \text{mean}(p_{\text{ls mn sim}}); \\
\end{align*}
\]
S10. Supplementary References


(6) Kildare, B. J.; Leutenegger, C. M.; McSwain, B. S.; Bambic, D. G.; Rajal, V. B.; Wuertz, S. 16S RRNA-Based Assays for Quantitative Detection of Universal, Human-, Cow-, and


(20) Somnark, P.; Chyerochana, N.; Mongkolsuk, S.; Sirikananchana, K. Performance Evaluation of Bacteroidales Genetic Markers for Human and Animal Microbial Source Tracking in


