

Evaluation of the association between meteorological and environmental risk factors and child diarrhea and enteropathogen infection prevalence in rural Bangladesh

Jessica A. Grembi, Anna Nguyen, Marie Riviere, Tejas S. Athni, Stephanie Djajadi, Ayse Ercumen, Audrie Lin, Yoshika Crider, Andrew Mertens, John M. Colford, Jr., Benjamin F. Arnold, Mahbubur Rahman, Leanne Unicomb, Stephen P. Luby, Adam Bennett, Jade Benjamin-Chung

1. Background and Scientific Rationale

Meteorological, environmental, and social factors influence the spread of infectious diseases and their associated vectors (Kelly-Hope et al. 2007; Bejon et al. 2014; Gómez-Barroso et al. 2017). In particular, spatiotemporal clustering of enteropathogen infection outbreaks have routinely been observed (Valcour et al. 2016; Glatman-Freedman et al. 2016; Chard et al. 2020; Kelly-Hope et al. 2007), and the influence of meteorological and environmental factors on the spread and persistence of enteropathogens has been well documented (Bates et al. 2007; Pullan et al. 2011). However, the relative importance of these risk factors, and potential interactions among them, are not fully understood for enteropathogens. Evidence on the relationships between enteropathogen infection and meteorological/environmental risk factors is critical to identifying promising interventions to reduce enteric disease, particularly within the context of climate change (Whitmee et al. 2015).

Prior studies have documented associations between meteorological risk factors and selected enteropathogens, but the nature of the association varies by pathogen (Levy et al. 2016). In tropical climates, the monsoon season is associated with increased diarrheal disease and enteropathogen prevalence (Platts-Mills et al. 2015; Black et al. 1982; Luby et al. 2018; Mertens et al. 2019; Grembi et al. 2020). The monsoon season is characterized by increased precipitation and higher temperatures, meteorological factors which are associated with enteropathogen prevalence independently and in combination (Kelly-Hope et al. 2007; Kovats et al. 2004; M.-J. Chen et al. 2012; Levy et al. 2016; Kraay Alicia N. M. et al., n.d.). However, increased precipitation and temperature also lead to higher relative humidity, which impacts both the inactivation rate and growth of different enteropathogens on surfaces, and can alter the transfer efficiency between fomites and hands (Stine et al. 2005; Lopez et al. 2013; Kim et al. 2012). Humidity can also modify the effect of other meteorological variables (i.e., temperature) on enteropathogens independent of precipitation, as shown by Liu et al. for shigellosis (Liu et al. 2019).

Other environmental risk factors, such as soil moisture, hydrologic landscape, vegetation index and land use patterns can influence enteropathogen persistence and transmission. The concentration of pathogenic *E. coli* was positively correlated with soil moisture in household soils in rural Bangladesh (Montealegre et al. 2018). Enteropathogen survival in soil can also be influenced by soil microbial community composition, which is strongly associated with land use (Moynihan et al. 2015). The vegetation index is negatively correlated with rotavirus infection in the Philippines but positively associated with *Schistosoma mansoni* prevalence in Brazil (Jagai et al. 2012; Bavia et al. 2001), again

emphasizing the differential effects of environmental risk factors on specific pathogens. Fecal contamination of surface water has been reported globally (Teklehaimanot et al. 2015; Rosef, Rettedal, and Lågeide 2001), but there have been few studies to directly link proximity to surface water to enteropathogen infection risk except in the case of cholera and *Schistosoma mansoni* (Ali et al. 2002; Arnold et al. 2020). Hydrologic landscape and water flow patterns could accumulate or wash away enteropathogens in the upper layers of soil, but these have not yet been explored as risk factors for enteropathogen infection or diarrheal disease.

The influence of environmental risk factors varies by enteropathogen, even among those with similar transmission routes. Population density was identified as a risk factor for cholera outbreaks, but not for diarrheal illness or soil-transmitted helminth infection (Ali et al. 2002; Jarquin et al. 2016) and norovirus-associated diarrhea is increased in colder temperatures (Greer, Drews, and Fisman 2009), while food-borne salmonellosis increases with temperature (Kovats et al. 2004). Thus, risk factors need to be evaluated for each pathogen individually.

Meteorological and environmental risk factors may also synergistically influence pathogen prevalence and/or transmission. For example, the hydrologic landscape, as characterized by flooding risk and soil moisture content, may interact with precipitation to result in localized enteropathogen hotspots; while such interaction has been documented for liver fluke infection (León et al. 2018; 2019), no prior studies have evaluated the relationship for enteropathogens. Social factors like population density and proximity to roads influence enteropathogen transmission and infection rates, and these relationships can be intensified by precipitation (Eisenberg et al. 2006; Bates et al. 2007; Ali et al. 2002; Borchardt et al. 2003). Proximate land use can also interact with rainfall events to intensify enteropathogen transmission (Walters, Thebo, and Boehm 2011; Poulsen et al. 2018). Population density can influence the fecal contamination of surface waters, thereby increasing the risk of enteropathogen infection from exposure to surface waters (Knappett et al. 2011). Although these risk factors are acknowledged as important for diarrheal disease and environmental transmission of enteropathogens, most previous studies have included only a subset of factors, which precludes the exploration of the full range of potential interactions (Ali et al. 2002; Bates et al. 2007; Chard et al. 2020; Chen et al. 2012; Fleury et al. 2006; Green, Krause, and Wylie 2006; Greer, Drews, and Fisman 2009; Jagai et al. 2012; Pullan et al. 2011).

In order to inform a precision public health approach (Desmond-Hellmann 2016) to reduce diarrheal disease and enteropathogen prevalence, this study aims to identify meteorological and environmental risk factors associated with disease prevalence. Leveraging data from a large intervention trial in rural Bangladesh (Luby et al. 2018; Lin et al. 2018; Ercumen et al. 2019; Benjamin-Chung et al. 2020; Grembi et al. 2020), we will identify spatial and temporally indexed environmental risk factors associated with diarrhea and enteropathogen prevalence in children, including a suite of bacteria, viruses, protozoa, and soil-transmitted helminths (STH). Evidence from this study will inform optimal resource allocation for future diarrheal disease and enteropathogen infection control efforts.

2. Study objectives

1. To assess the spatial distribution of diarrhea and enteropathogen prevalence among children in rural Bangladesh
2. To assess the temporal distribution of diarrhea and enteropathogen prevalence among children in rural Bangladesh
3. To identify meteorological and environmental risk factors associated with diarrhea and enteropathogen prevalence among children in rural Bangladesh
4. To identify interactions between meteorological and environmental risk factors and diarrhea and enteropathogen prevalence among children in rural Bangladesh

3. Study design

We will conduct an observational, exploratory analysis using temporally matched geospatial risk factors and caregiver-reported diarrhea and enteropathogen infection data collected prospectively within the WASH Benefits Bangladesh cluster randomized controlled trial. Details about the six intervention arms and design of the WASH Benefits cluster-randomized trial can be found in Luby et al. 2018. Briefly, the trial included 8 arms: improved water (W), improved sanitation (S), improved handwashing (H), improved nutrition (N), combined WSH, combined N+WSH, and a double-sized control arm. Within rural Gazipur, Mymensingh, Tangail, and Kishoreganj districts of rural Bangladesh, 90 village clusters for each arm (180 for control arm) were randomized within geographic blocks (8 clusters per block) and 7-8 compounds were enrolled per cluster. The trial enrolled a total of 720 village clusters spanning a geographic area of 12,500 km²; the distance between compounds within each cluster ranged from 0-1.2 km (median: 0.3 km). Pregnant women in their 2nd or 3rd trimester were enrolled and their children ("index children") were followed for approximately two years.

4. Participants

This analysis will be restricted to children enrolled in one of three cohorts of the WASH Benefits Bangladesh study: the diarrhea cohort, the parasite cohort, or the environmental enteropathy cohort. These data span 3 years and were collected in one of six waves (enrollment, midline, and endline for main cohort; and baseline, midline and endline for the environmental enteropathy [EE] cohort which occurred ~6 months after each main survey visit for children enrolled in this cohort). Each collection wave lasted approximately 1 year (details below). Monthly, approximately 54 clusters were sampled from the main cohort and 14 from the EE cohort in a spatially dependent manner as the study team moved around the study area over the course of a year (dates detailed below).

The diarrhea cohort included index children, other children living within the same compound that were younger than 3 years at study enrollment, and children within the same household born after the index children. 10,887 children were enrolled in the diarrhea cohort. Caregiver-reported diarrhea was collected at enrollment (May 2012-July 2013), and approximately 13 and 26 months after enrollment (Sept 2013-Sept 2014 and Dec 2014-Oct 2015). Data on caregiver-reported diarrhea was collected at three additional times for a subset of 1,787 index children (enrolled in the environmental enteropathy cohort, described below) at approximately 7 (Dec 2012-Jan 2014), 18 (Nov 2013-Nov 2014), and 33 (Mar 2015-Mar 2016) months after enrollment. This cohort will be used for evaluating the diarrhea outcome.

The parasite cohort included all index children and up to 2 other children per enrolled compound aged 3-12 years who lived in the index household or another household in the compound. These additional 2 children were enrolled in the preferential order of sibling of index child, child living in index household, or child living in another household in the compound, depending on availability. The parasite cohort consisted of 7,795 children aged 2-12 years of 4,102 women enrolled for follow-up. Stool samples were collected from 7,187 children approximately 36 months after enrollment, with a median age of 2.7 years (mean: 4.8 years, range: 1.8-12 years) and measured for *Giardia* and STH (Ercumen et al. 2019; Lin et al. 2018; Benjamin-Chung et al. 2020). Additionally, at enrollment stool samples were collected from 705 children aged 18-27 months living in the compound (older sibling or from another household). Bangladesh has a national school-based mass drug administration (MDA) program (mebendazole, biannually) and a separate MDA program (albendazole) for pre-school-aged children. Over 60% of children were reported to have consumed deworming medication in the previous 6 months, with approximately half dewormed through the MDA program. This cohort will be used for STH and *Giardia* outcomes.

The EE cohort included index children from a subsample of clusters that was evenly balanced across the control, WSH, nutrition, and N+WSH arms (allocation ratio 1:1:1:1) (Lin et al. 2020). Clusters in each arm were selected based on logistical feasibility for specimen collection and transport to a central laboratory. Households were visited three times for biological specimen collection; the cohort consisted of 1,645 children at age 3 months, 1,978 children at age 14 months, and 2,049 children at age 28 months. Caregiver reported diarrhea and stool samples were collected from index children at all three timepoints: age 3.0 +/- 1.7 months (n = 1,090), age 14.0 +/- 2.0 months (n = 1,499), and age 28.1 +/- 1.9 months (n = 1,512) (Lin et al. 2020). Enteropathogens were measured from stool collected when index children were approximately 14 months old (Grembi et al. 2020). This cohort will be used for all pathogen outcomes evaluated in children less than 18 months old.

Additional eligibility criteria are described elsewhere (Luby et al. 2018; Lin et al. 2018; Ercumen et al. 2019; Benjamin-Chung et al. 2020; Grembi et al. 2020; Lin et al. 2020).

5. Outcomes

Primary outcomes

Prevalence of:

- Caregiver-reported diarrhea in the past 7 days for children 0-5.5 yrs old
- *Giardia* in children 18 months -12 years old (detected by qPCR)
- Any STH in children 2-12 years old (detected by qPCR)
- Any virus in children 14 months old (detected by qPCR)
- Any parasite in children 14 months old (detected by qPCR)

Secondary outcomes

1. Prevalence of:

- Pathogen-specific STH in children 2-12 years old (detected by qPCR)
- Pathogen-specific STH in children 2-12 years old (detected by Kato-Katz)
- Pathogen-specific viruses, bacteria, and parasites in children 14 months old (detected by qPCR)

2. Quantity of enteropathogens as measured by:

- Mean Cq value for *Giardia* in children 18 months -12 years old (detected by qPCR)
- Mean Cq value for STH in children 2-12 years old (detected by qPCR)
- Mean eggs per gram of stool for *Trichuris trichuria* and hookworm in children 2-12 years old (detected by Kato-Katz)
- Mean Cq value for pathogen-specific viruses, bacteria, and parasites in children 14 months old (detected by qPCR)
- Number of pathogens (in total and by type: virus, bacteria, and parasite) in children 14 months old (detected by qPCR)

Diarrhea was defined as three or more loose or watery stools in a 24-hour period or a single stool with blood. We obtained a total of 31,308 measurements of caregiver-reported diarrhea. This included 3,678 measurements at enrollment (all older children from enrolled compounds, median age 1.7 yrs); 1,131 measurements at 7 months post-enrollment (all index children, median age 3 months); 8,952 measurements at 13 months post-enrollment (4,747 index children aged 9 months old, and 4,205 other children in the same compound aged 2.3 years old); 1531 measurements at 18 months post-enrollment (all index children, median age 14 months); 9,964 measurements at 26 months post-enrollment (4,667 index children aged 1.9 years old and 5,297 other children in the same compound aged 2.9 years old); and 1531 measurements at 33 months post-enrollment (all index children, median age 28 months). Caregiver-reported bruising and abrasion were also assessed as negative control outcomes (Lipsitch, Tchetgen, and Cohen 2010).

Giardia was measured in 6,896 children from the parasite cohort using qPCR (3,717 index children, 2,691 older siblings, and 488 older children from the compound). Soil-transmitted helminths were measured via qPCR in 2,800 children from the parasite cohort (1,494 index children, 1,108 older siblings, and 198 older children from within the compound) that were randomly selected from control, individual improved water, sanitation, handwashing, and combined WSH arms. All samples from the parasite cohort (n = 7,187) were additionally analyzed for STH by double-slide Kato-Katz microscopy. Though we performed Kato-Katz in a larger number of specimens, qPCR has higher sensitivity and specificity than Kato-Katz (Benjamin-Chung et al. 2020). Thus, Kato-Katz data for hookworm and *Trichuris* will be used as secondary outcomes, as described below (*Ascaris* will not be included in this secondary analysis due to concerns about potential misclassification using Kato-Katz (Benjamin-Chung et al. 2020)).

A suite of 34 enteropathogens, including viruses, bacteria and parasites (Table A1), were measured via qPCR in 1,411 fecal samples from EE cohort children at 14 months age; further details can be found in (Grembi et al. 2020). Pathogens detected in >10% of samples from the EE cohort will be included in the single pathogen analysis. Additionally, we will evaluate combined pathogen outcome measures: prevalence of any pathogen type (viruses or parasites; bacteria are not included as >95% of stools had a

bacterial pathogen detected) and the number of co-occurring pathogens (in total and by type: virus, bacteria, and parasite).

6. Risk factors

We will include nine meteorological and environmental risk factors (Table 1). We will extract spatial risk factor values for each study compound by matching the compound's longitude and latitude to the nearest longitude and latitude values in the source data. If the source data do not contain values near to a study compound (defined as within 500m from a household, or for data with lower spatial resolution one unit based on the spatial resolution for the given risk factor), we will assign a NULL value. We will match compounds to vapor pressure deficit (a proxy for humidity) values that are recorded in the same month/year that outcomes were measured.

We assume water flow accumulation, a hydrologic measure of drainage through or into a specific area, is fairly constant over time and therefore use data obtained from 2006-2007. The Global Surface Water variables for 'occurrence' and 'transition' will be used to calculate the distance from each compound to the surface water source – any, seasonal, or ephemeral – in the same year as sample collection. The abundance/frequency of surface water in proximity to the compound will be estimated by calculating a) the proportion of area covered by surface water within a 250, 500, and 750 meter radius of the household for surface water categories – any, seasonal, or ephemeral –between 1984 and 2015 , b) maximum number of months surface water was present within a 100, 250, 500, or 750 meter radius of the household between October 2014 and October 2015 (this was calculated using the 'seasonality' variable from the Global Surface Water data, with the assumption that annual water seasonality is likely stable).

We will summarize compound-matched daily temperature over the past 30 days, and also 7-day periods preceding outcome data collection including lags up to 3 weeks. Source data will be used to calculate the average, absolute minimum and absolute maximum for 30 or 7 days prior to observation plus 1-, 2-, 3-week lags for the 7-day average.

Compound-matched daily or total weekly precipitation (NOAA/OAR/ESRL dataset, see Table 1) will be calculated with the same lag periods as described for temperature. In addition, we will create a binary variable for heavy rainfall to identify if total precipitation in any 24-h period within the previous 7 days had heavy rainfall ($\geq 80^{\text{th}}$ percentile of total daily precipitation for days where rain was observed, over the entire study period). Finally, we will create an additional categorical variable to describe the rainfall in the previous 60 days to the heavy rainfall variable, defined as "high rainfall" if total precipitation in 60-day period $\geq 66^{\text{th}}$ percentile of all precipitation data, "low rainfall" if total precipitation $\leq 33^{\text{rd}}$ percentile of all precipitation data, and "medium rainfall" otherwise, in accordance with (Mertens et al. 2019). Both the binary heavy rainfall and the categorical 60-d precipitation variables will follow the same lag periods as described for precipitation and temperature above (e.g. the binary heavy rainfall variable will be calculated with 0-, 1-, 2-, & 3-week lags for the diarrhea outcome and 60-d precipitation will be calculated for the 60-days preceding each of those lag periods).

The average enhanced vegetation index is provided at a resolution of 16 days, and we will use the closest date range preceding the date of the outcome measure.

For land use classification, we will match on the year of data collection.

For population density, we will use 2010 data to match compounds with measurements in 2012-14 and 2015 data to match households with measurements in 2015-16.

Table 1. Environmental risk factor definitions and data sources

Spatial risk factor	Source	Description	Unit of measurement	Temporal resolution	Spatial resolution	Link to causal pathway (references included in Background section)
Vapor pressure deficit	Terraclimate (Abatzoglou et al. 2018)	The difference between observed water vapor pressure and the water vapor pressure at full air saturation. Low VPD is associated with humid air, while high VPD is associated with dry air.	kPa	Monthly	4 km (1/24th degree)	- Can influence growth & inactivation rates of enteropathogens on surfaces - Can alter transfer efficiency of pathogens from fomites to hands
Water flow accumulation	WWF HydroSHEDS Hydrologically Conditioned DEM (Lehner, Verdin, and Jarvis 2008)	The number of cells that drain into a particular location. Low values correspond to areas at high elevation that do not see much water drainage. High values correspond to areas of low elevation, where there are many instances where water is flowing from higher elevation sites. Data was compiled in 2006-2007.	Number of hydrologic cells (0-25088 within our study area)	N/A	450 meters (15 arc-seconds)	- Can accumulate or wash away enteropathogens in the upper layers of soil
Distance to surface water	Global Surface Water Explorer (Pekel et al. 2016)	The calculated distance from the child's household to the closest surface water body transition and occurrence variables. The following specific risk factors were created: <ul style="list-style-type: none"> Distance from any surface water (occurrence) Distance from seasonal surface water (transition) Distance from ephemeral surface water (transition) 	meter	Monthly	30 meters	- Increased exposure to potentially contaminated surface water located near the household (during bathing, clothes washing, etc.) - Proximity of surface water used for bathing and cleaning might influence frequency of these activities - Older siblings exposed to nearby surface water could bring enteropathogens to household environment
Abundance and frequency of surface water in close proximity to household	Global Surface Water Explorer (Pekel et al. 2016)	Multiple variables were used from this dataset to create the following risk factors for analysis: <ul style="list-style-type: none"> Number of months between October 2014 and October 2015 that surface water was detected within 100, 250, 500, and 750m of household, as detected by the seasonality variable Surface water occurrence (mean percent of time between 1984 and 2015 that surface water was present within 500m of household) Proportion of area within radius (250, 500, 750m) with surface water detected: <ul style="list-style-type: none"> Any surface water (occurrence) 	months	Monthly	30 meters	- Seasonality of surface water can influence water use, source protection efforts, and can also concentrate enteropathogens in soil/sediments (Knappett et al. 2011; Ehsan et al. 2015; Burton, Gunnison, and Lanza 1987)

		<ul style="list-style-type: none"> Seasonal surface water (transition) Ephemeral surface water (transition) 				
Precipitation (for analyses requiring greater temporal resolution)	NOAA/OAR/ESRL Physical Science Laboratory (Xie et al. 2007; M. Chen et al. 2008)	<p>The precipitation accumulation for the geographical area. Used to create variables for:</p> <ul style="list-style-type: none"> Average 7-day precip, with lags Heavy rainfall in past week (total precipitation in 24h period within previous 7 days >80th percentile daily precipitation over the study period), with lags Rainfall categories (based on 33rd and 66th percentile cutoffs over the study period) for the 60 days preceding heavy rainfall measurement, with lags 	mm; and also factor: heavy rainfall in past week (Y/N), and previous 60-day rainfall category (high, medium, low).	Daily	55 km (0.5 degree)	<ul style="list-style-type: none"> - Can dilute/ wash away environmental enteropathogen contamination - Can transport enteropathogens into household area (e.g., overflow of pit latrine, flooding of nearby rivers/ponds, etc.) - Often important for pathogen transport immediately following extended dry period
Temperature (minimum, maximum, average)	NASA MOD11A1 (Wan, Hook, and Hulley 2015)	A daily daytime and nighttime land surface temperature was extracted for each coordinate point throughout the study period. The absolute minimum temperature, absolute maximum temperature, and average temperature were reported for the 1 day, 7 days, 30 days, and 90 day periods preceding an observation.	°C	Daily	1 km	<ul style="list-style-type: none"> - Can influence pathogen persistence (survival and growth rates) in the environment
Land use classification	NASA MCD12Q1 (M. Friedl 2019) (LC_Type1)	Categorizes land use of an area using the Annual International Geosphere-Biosphere Programme (IGBP) classification.	Factor: mosaic, cropland, savanna, NA (descriptions below)	Yearly	500 m	<ul style="list-style-type: none"> - Can influence pathogen persistence (survival and growth rates) in soil - Croplands fertilized with manure or land used for livestock grazing can introduce zoonotic enteropathogens into soils
Vegetation index	NASA MOD13Q1 (Didan 2015)	<p>A ratio between the red (R) and near infrared (NIR) values, adjusting for canopy background (L), atmospheric resistance (C1, C2), blue values (B), and a constant term (G):</p> $G * ((NIR - R) / (NIR + C1 * R - C2 * B + L))$	Enhanced vegetation index (EVI)	16-day	250 m	<ul style="list-style-type: none"> - Can influence pathogen persistence (survival and growth rates) in soils
Population density	WorldPop (Worldpop 2017)	The number of people per square pixel, with each pixel representing an approximately 100 m x 100 m area.	Persons per 100 m ²	2010, 2015, 2020	100 m	<ul style="list-style-type: none"> - Can increase person-to-person transmission - Can increase fecal contamination of surface waters & the risk of infection from contact with surface water

The International Geosphere-Biosphere Programme (M. Friedl 2019) classifies land use into one of 17 categories; our study area contains three major categories:

- **Cropland/Natural Vegetation Mosaics:** mosaics of small-scale cultivation 40-60% with natural tree, shrub, or herbaceous vegetation.
- **Croplands:** at least 60% of area is cultivated cropland.

- **Savannas:** tree cover 10-30% (canopy >2m).
- Less than 40 observations are present in our dataset for the following land use classifications, which will be dropped due to sparsity concerns.
 - **Mixed Forests:** dominated by neither deciduous nor evergreen (40-60% of each) tree type (canopy >2m). Tree cover >60%.
 - **Deciduous Broadleaf Forests:** dominated by deciduous broadleaf trees (canopy >2m). Tree cover >60%.
 - **Urban and Built-up Lands:** at least 30% impervious surface area including building materials, asphalt and vehicles.
 - **Permanent Wetlands:** permanently inundated lands with 30-60% water cover and >10% vegetated cover.

7. Statistical methods

Tests of spatial clustering

We will use Kulldorff's spatial scan statistic to detect hot spots of prevalence in the study among all observations and stratifying by child age and season (rainy season is generally April - October but will be empirically defined for each year using daily precipitation data, see Figure A1) (Kulldorff and Nagarwalla 1995). We will define the upper limit of the scan window such that no cluster detected contains more than 25% of the study population. The null hypothesis is that the prevalence within a given scanning window is the same as the prevalence outside the window.

Associations between risk factors and each outcome

The goal of this analysis is to test the associations between risk factors and outcomes while accounting for spatial correlation. To allow for potential non-linear relationships between risk factors and outcomes, we will fit generalized additive mixed models (GAMM) (Hastie and Tibshirani 1986; Wood 2006). We will fit GAMMs with a binomial family with a logit link for binary outcomes and a zero-inflated negative binomial family with a log link for count outcomes. We will include environmental risk factors as continuous variables with the exception of land use and heavy precipitation, which are categorical variables (see definitions above). We will use restricted maximum likelihood (REML) estimation to fit smoothing parameters. We will use cubic regression splines with shrinkage-to-zero (Marra and Wood 2011) for smooth terms to minimize overfitting and correlation between covariates. If we detect spatial autocorrelation, we will also include a bidimensional thin plate spline function of latitude and longitude for each study compound and assess model fit using the Akaike information criteria (Wood 2006). To account for repeated measures among individuals in the same compound for diarrhea, STH, and *Giardia* outcomes, we will include random intercepts for study compound. 7% of children had repeated measures for *Giardia*, but measures were taken over 2 years apart at ages when *Giardia* risk differs substantially, so we will not include child-level random effects. To account for repeated measures among individuals in the same cluster in the EE cohort, we will include random intercepts for study cluster.

We will pool across all intervention arms from the original study to maximize statistical power and will include study arm as a covariate to adjust for any effects of study interventions on each outcome, except

for the diarrhea outcome where we will conduct subgroup analyses separately within the control arm or intervention arms (all grouped together) due to the significant influence of the interventions on diarrhea in a seasonally-dependent manner.

We defined potential confounders for each risk factor using directed acyclic graphs (Table 2). In addition, we defined covariates that we believe are strongly associated with the outcome. These include child age, sex, and household wealth quartile. Additional covariates associated with certain outcomes include antibiotic use in the past week for bacterial pathogens and anthelmintic use in the previous two weeks for STH. We will center continuous covariates, including risk factors. We will adjust for covariates in each model that are associated with the outcome (likelihood ratio test p-value < 0.2). To reduce the possibility of empirical positivity violations (Petersen et al. 2012), we will only fit models for analyses in which the number of outcome events per variable is at least 10 (Peduzzi et al. 1996).

Public holidays can be associated with increased travel, mass gatherings, and altered food preparation practices that can influence enteropathogen transmission (Collender et al. 2019; Schwab and Armah 2019). We have defined a holiday variable to identify samples collected within 1 week following a public holiday (binary variable if sample was collected during Ramadan or within 1 week after Bengali New Year, Eid al-Fitr, or Eid al-Adha). As there was not sufficient variation in this variable in our dataset to include as a covariate, we will conduct a sensitivity analysis removing samples collected following a holiday. Public holidays can be associated with increased travel, mass gatherings, and altered food preparation practices that can influence enteropathogen transmission (Collender et al. 2019; Schwab and Armah 2019). We have defined a holiday variable to identify samples collected within 1 week following a public holiday (binary variable if sample was collected during Ramadan or within 1 week after Bengali New Year, Eid al-Fitr, or Eid al-Adha). As there was not sufficient variation in this variable in our dataset to include as a covariate, we will conduct a sensitivity analysis removing samples collected following a holiday.

We will include smooth terms for each continuous risk factor and covariate and linear fixed effects for categorical risk factors (land use) and covariates. To account for potential seasonality, we will include fixed effects for the season of outcome measurement. We will check for collinearity between risk factors and confounders, and exclude those with a correlation coefficient >0.7 (Dormann et al. 2013).

Table 2. Potential confounders to be adjusted for in statistical models based on directed acyclic graphs for all outcomes

Risk factor	Potential Confounder
Vapor pressure deficit	Monthly precipitation
Water flow accumulation	Monthly precipitation, season, number of animals in compound
Distance from surface water	Water flow accumulation, number of animals in compound

Frequency of surface water in close proximity to household	Monthly precipitation
Precipitation	None
Temperature	None
Land use classification	Household wealth quartile, parental education levels
Normalized difference vegetation index	Household wealth quartile, parental education levels, number of animals in the compound, vapor pressure deficit, water flow accumulation
Population density	Household wealth quartile, parental education levels, number of people in household

Assessment of effect modification

We will fit models to identify potential effect modifiers. As this analysis is exploratory, the list is not comprehensive and we may choose to add to it or to exclude analyses if data is too sparse. For each pair of risk factors, we will evaluate the feasibility of assessing effect modification by variables on a continuous scale (e.g., continuous by continuous interaction) or discretizing variables depending on their empirical distributions.

We will assess whether child age (grouped by ≤ 18 months or > 18 months for diarrhea and 0-5 years or 6-12 years for *Giardia* and STH) is an effect modifier of the association with each risk factor. In addition, we will explore interactions between the following risk factors and each outcome:

- Precipitation + water flow accumulation
- Precipitation + temperature
- Precipitation + prior (60 days) dry period
- Population density + precipitation
- Population density + frequency of surface water in close proximity to household

We will also test whether each environmental risk factor modifies the effect of WASH interventions on diarrhea and enteropathogen infection. We will assess potential effect modification for six treatment arms of the WASH Bangladesh trials (excluding the nutrition-only intervention arm). If there is data sparsity in treatment-risk factor distributions, we will consider pooling across intervention arms.

For continuous risk factors we will fit GAMs for each outcome including an interaction term between treatment arm and each environmental variable and calculate simultaneous 95% confidence intervals (Marra and Wood 2012).

For categorical risk factors we will fit targeted maximum likelihood estimation (TMLE) models for each intervention and outcome stratifying by each risk factor level and estimate relative risk and risk differences.

References

- Abatzoglou, John T., Solomon Z. Dobrowski, Sean A. Parks, and Katherine C. Hegewisch. 2018. "TerraClimate, a High-Resolution Global Dataset of Monthly Climate and Climatic Water Balance from 1958–2015." *Scientific Data* 5 (1): 170191. <https://doi.org/10.1038/sdata.2017.191>.
- Ali, Mohammad, Michael Emch, J.P. Donnay, Mohammad Yunus, and R.B. Sack. 2002. "Identifying Environmental Risk Factors for Endemic Cholera: A Raster GIS Approach." *Health & Place* 8 (3): 201–10. [https://doi.org/10.1016/S1353-8292\(01\)00043-0](https://doi.org/10.1016/S1353-8292(01)00043-0).
- Arnold, Benjamin F., Henry Kanyi, Sammy M. Njenga, Fredrick O. Rawago, Jeffrey W. Priest, W. Evan Secor, Patrick J. Lammie, Kimberly Y. Won, and Maurice R. Odiere. 2020. "Fine-Scale Heterogeneity in Schistosoma Mansonii Force of Infection Measured through Antibody Response." *Proceedings of the National Academy of Sciences* 117 (37): 23174–81. <https://doi.org/10.1073/pnas.2008951117>.
- Bates, S. J., J. Trostle, W. T. Cevallos, A. Hubbard, and J. N. S. Eisenberg. 2007. "Relating Diarrheal Disease to Social Networks and the Geographic Configuration of Communities in Rural Ecuador." *American Journal of Epidemiology* 166 (9): 1088–95. <https://doi.org/10.1093/aje/kwm184>.
- Bavia, M. E., J. B. Malone, L. Hale, A. Dantas, L. Marroni, and R. Reis. 2001. "Use of Thermal and Vegetation Index Data from Earth Observing Satellites to Evaluate the Risk of Schistosomiasis in Bahia, Brazil." *Acta Tropica*, A Global Network on Schistosomiasis Information Systems and Control of Snail-Borne Diseases, 79 (1): 79–85. [https://doi.org/10.1016/S0001-706X\(01\)00105-X](https://doi.org/10.1016/S0001-706X(01)00105-X).
- Bejon, Philip, Tom N Williams, Christopher Nyundo, Simon I Hay, David Benz, Peter W Gething, Mark Otiende, et al. 2014. "A Micro-Epidemiological Analysis of Febrile Malaria in Coastal Kenya Showing Hotspots within Hotspots." *ELife* 2014 (3): 2130–2130. <https://doi.org/10.7554/eLife.02130>.
- Benjamin-Chung, Jade, Nils Pilotte, Ayse Ercumen, Jessica R. Grant, Jacqueline R. M. A. Maasch, Andrew M. Gonzalez, Ashanta C. Ester, et al. 2020. "Comparison of Multi-Parallel QPCR and Double-Slide Kato-Katz for Detection of Soil-Transmitted Helminth Infection among Children in Rural Bangladesh." *PLOS Neglected Tropical Diseases* 14 (4): e0008087. <https://doi.org/10.1371/journal.pntd.0008087>.
- Black, Robert E, Kenneth H Brown, Stan Becker, A.R.M. Abdul Alim, and Imdadul Huq. 1982. "Longitudinal Studies of Infectious Diseases and Physical Growth of Children in Rural Bangladesh: II. Incidence of Diarrhea and Association with Known Pathogens." *American Journal of Epidemiology* 115 (3): 315–24. <https://doi.org/10.1093/oxfordjournals.aje.a113308>.
- Borchardt, Mark A., Po Huang Chyou, Edna O. DeVries, and Edward A. Belongia. 2003. "Septic System Density and Infectious Diarrhea in a Defined Population of Children." *Environmental Health Perspectives* 111 (5): 742–48. <https://doi.org/10.1289/ehp.5914>.
- Burton, G. A., D. Gunnison, and G. R. Lanza. 1987. "Survival of Pathogenic Bacteria in Various Freshwater Sediments." *Applied and Environmental Microbiology* 53 (4): 633–38.
- Chard, Anna N., Karen Levy, Kelly K. Baker, Kevin Tsai, Howard H. Chang, Vonethalom Thongpaseuth, Jeticia R. Sistrunk, and Matthew C. Freeman. 2020. "Environmental and Spatial Determinants of Enteric Pathogen Infection in Rural Lao People's Democratic Republic: A Cross-Sectional Study." Edited by Xiao-Nong Zhou. *PLOS Neglected Tropical Diseases* 14 (4): e0008180–e0008180. <https://doi.org/10.1371/journal.pntd.0008180>.

- Chen, Mingyue, Wei Shi, Pingping Xie, Viviane B. S. Silva, Vernon E. Kousky, R. Wayne Higgins, and John E. Janowiak. 2008. "Assessing Objective Techniques for Gauge-Based Analyses of Global Daily Precipitation." *Journal of Geophysical Research: Atmospheres* 113 (D4). <https://doi.org/10.1029/2007JD009132>.
- Chen, Mu-Jean, Chuan-Yao Lin, Yi-Ting Wu, Pei-Chih Wu, Shih-Chun Lung, and Huey-Jen Su. 2012. "Effects of Extreme Precipitation to the Distribution of Infectious Diseases in Taiwan, 1994-2008." *PloS One* 7 (6): e34651–e34651. <https://doi.org/10.1371/journal.pone.0034651>.
- Collender, Philip A, Christa Morris, Rose Glenn-Finer, Andrés Acevedo, Howard H Chang, James A Trostle, Joseph N S Eisenberg, and Justin V Remais. 2019. "Mass Gatherings and Diarrheal Disease Transmission Among Rural Communities in Coastal Ecuador." *American Journal of Epidemiology* 188 (8): 1475–83. <https://doi.org/10.1093/aje/kwz102>.
- Desmond-Hellmann, Sue. 2016. "Progress Lies in Precision." *Science (New York, N.Y.)* 353 (6301): 731–731. <https://doi.org/10.1126/science.aai7598>.
- Didan, K. 2015. "MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006." NASA EOSDIS Land Processes DAAC. <https://doi.org/10.5067/MODIS/MOD13Q1.006>.
- Dormann, Carsten F., Jane Elith, Sven Bacher, Carsten Buchmann, Gudrun Carl, Gabriel Carré, Jaime R. García Marquéz, et al. 2013. "Collinearity: A Review of Methods to Deal with It and a Simulation Study Evaluating Their Performance." *Ecography* 36 (1): 27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>.
- Ehsan, Amimul M., Thomas Geurden, Stijn Casaert, Sonia M. Parvin, Taohidul M. Islam, Uddin M. Ahmed, Bruno Levecke, Jozef Vercruysse, and Edwin Claerebout. 2015. "Assessment of Zoonotic Transmission of Giardia and Cryptosporidium between Cattle and Humans in Rural Villages in Bangladesh." *PLOS ONE* 10 (2): e0118239. <https://doi.org/10.1371/journal.pone.0118239>.
- Eisenberg, Joseph N S, William Cevallos, Karina Ponce, Karen Levy, Sarah J Bates, James C Scott, Alan Hubbard, et al. 2006. "Environmental Change and Infectious Disease: How New Roads Affect the Transmission of Diarrheal Pathogens in Rural Ecuador." www.pnas.org/cgi/doi/10.1073/pnas.0609431104.
- Ercumen, Ayse, Jade Benjamin-Chung, Benjamin F. Arnold, Audrie Lin, Alan E. Hubbard, Christine Stewart, Zahidur Rahman, et al. 2019. "Effects of Water, Sanitation, Handwashing and Nutritional Interventions on Soil-Transmitted Helminth Infections in Young Children: A Cluster-Randomized Controlled Trial in Rural Bangladesh." Edited by Susana Vaz Nery. *PLoS Neglected Tropical Diseases* 13 (5): e0007323–e0007323. <https://doi.org/10.1371/journal.pntd.0007323>.
- Fleury, Manon, Dominique F. Charron, John D. Holt, O. Brian Allen, and Abdel R. Maarouf. 2006. "A Time Series Analysis of the Relationship of Ambient Temperature and Common Bacterial Enteric Infections in Two Canadian Provinces." *International Journal of Biometeorology* 50 (6): 385–91. <https://doi.org/10.1007/s00484-006-0028-9>.
- Glatman-Freedman, Aharon, Zalman Kaufman, Eran Kopel, Ravit Bassal, Diana Taran, Lea Valinsky, Vered Agmon, et al. 2016. "Near Real-Time Space-Time Cluster Analysis for Detection of Enteric Disease Outbreaks in a Community Setting." *Journal of Infection* 73 (2): 99–106. <https://doi.org/10.1016/J.JINF.2016.04.038>.
- Gómez-Barroso, Diana, Emely García-Carrasco, Zaida Herrador, Policarpo Ncogo, María Romay-Barja, Martín Eka Ondo Mangue, Gloria Nseng, et al. 2017. "Spatial Clustering and Risk Factors of Malaria Infections in Bata District, Equatorial Guinea." *Malaria Journal* 16 (1): 146–146. <https://doi.org/10.1186/s12936-017-1794-z>.
- Green, ChrisG, DennisO Krause, and JohnL Wylie. 2006. "Spatial Analysis of Campylobacter Infection in the Canadian Province of Manitoba." *International Journal of Health Geographics* 5 (1): 2–2. <https://doi.org/10.1186/1476-072X-5-2>.
- Greer, Amy L., Steven J. Drews, and David N. Fisman. 2009. "Why 'Winter' Vomiting Disease? Seasonality, Hydrology, and Norovirus Epidemiology in Toronto, Canada." *EcoHealth* 6 (2): 192–99. <https://doi.org/10.1007/s10393-009-0247-8>.

- Grembi, Jessica A., Audrie Lin, Md Abdul Karim, Md Ohedul Islam, Rana Miah, Benjamin F. Arnold, Elizabeth T. Rogawski McQuade, et al. 2020. "Effect of Water, Sanitation, Handwashing and Nutrition Interventions on Enteropathogens in Children 14 Months Old: A Cluster-Randomized Controlled Trial in Rural Bangladesh." *The Journal of Infectious Diseases*.
<https://doi.org/10.1093/infdis/jiaa549>.
- Hastie, Trevor, and Robert Tibshirani. 1986. "Generalized Additive Models." *Statistical Science* 1 (3): 297–310. <https://doi.org/10.1214/ss/1177013604>.
- Jagai, Jyotsna S., Rajiv Sarkar, Denise Castronovo, Deepthi Kattula, Jesse McEntee, Honorine Ward, Gagandeep Kang, and Elena N. Naumova. 2012. "Seasonality of Rotavirus in South Asia: A Meta-Analysis Approach Assessing Associations with Temperature, Precipitation, and Vegetation Index." Edited by Cécile Viboud. *PLoS ONE* 7 (5): e38168–e38168.
<https://doi.org/10.1371/journal.pone.0038168>.
- Jarquin, Claudia, Benjamin F. Arnold, Fredy Muñoz, Beatriz Lopez, Victoria M. Cuéllar, Andrew Thornton, Jaymin Patel, et al. 2016. "Population Density, Poor Sanitation, and Enteric Infections in Nueva Santa Rosa, Guatemala." *The American Journal of Tropical Medicine and Hygiene* 94 (4): 912–19. <https://doi.org/10.4269/ajtmh.15-0555>.
- Kelly-Hope, Louise A., Wladimir J Alonso, Dinh Thiem Vu, Duc Anh Dang, Gia Canh Do, Hyejon Lee, David L Smith, and Mark A Miller. 2007. "Geographical Distribution and Risk Factors Associated with Enteric Diseases in Vietnam." *American Journal of Tropical Medicine and Hygiene* 76 (4): 706–12. <https://doi.org/10.4269/ajtmh.2007.76.706>.
- Kim, Su Jung, Jiyeon Si, Jung Eun Lee, and Gwangpyo Ko. 2012. "Temperature and Humidity Influences on Inactivation Kinetics of Enteric Viruses on Surfaces." <https://doi.org/10.1021/es3032105>.
- Knappett, Peter S. K., Veronica Escamilla, Alice Layton, Larry D. McKay, Michael Emch, Daniel E. Williams, R. Huq, et al. 2011. "Impact of Population and Latrines on Fecal Contamination of Ponds in Rural Bangladesh." *Science of The Total Environment* 409 (17): 3174–82.
<https://doi.org/10.1016/j.scitotenv.2011.04.043>.
- Kovats, R S, S J Edwards, S Hajat, B G Armstrong, K L Ebi, and B Menne. 2004. "The Effect of Temperature on Food Poisoning: A Time-Series Analysis of Salmonellosis in Ten European Countries." *Epidemiology and Infection* 132 (3): 443–53.
<https://doi.org/10.1017/s0950268804001992>.
- Kraay Alicia N. M., Man Olivia, Levy Morgan C., Levy Karen, Ionides Edward, and Eisenberg Joseph N. S. n.d. "Understanding the Impact of Rainfall on Diarrhea: Testing the Concentration-Dilution Hypothesis Using a Systematic Review and Meta-Analysis." *Environmental Health Perspectives* 128 (12): 126001. <https://doi.org/10.1289/EHP6181>.
- Kulldorff, Martin, and Neville Nagarwalla. 1995. "Spatial Disease Clusters: Detection and Inference." *Statistics in Medicine* 14 (8): 799–810. <https://doi.org/10.1002/sim.4780140809>.
- Lehner, Bernhard, Kristine Verdin, and Andy Jarvis. 2008. "New Global Hydrography Derived From Spaceborne Elevation Data." *Eos, Transactions American Geophysical Union* 89 (10): 93–94.
<https://doi.org/10.1029/2008EO100001>.
- León, Tomás M., Vichian Plermkamon, Kittiwet Kuntiyawichai, Banchob Sripana, and Robert C. Spear. 2019. "Hydrology-Informed Metapopulation Modeling of Liver Fluke Transmission in the Lawa Lake Complex of Northeast Thailand." *BioRxiv*, March, 569913. <https://doi.org/10.1101/569913>.
- León, Tomás M., Travis C. Porco, Christina S. Kim, Sasithorn Kaewkes, Wanlop Kaewkes, Banchob Sripana, and Robert C. Spear. 2018. "Modeling Liver Fluke Transmission in Northeast Thailand: Impacts of Development, Hydrology, and Control." *Acta Tropica* 188 (December): 101–7.
<https://doi.org/10.1016/j.actatropica.2018.08.008>.
- Levy, Karen, Andrew P. Woster, Rebecca S. Goldstein, and Elizabeth J. Carlton. 2016. "Untangling the Impacts of Climate Change on Waterborne Diseases: A Systematic Review of Relationships between Diarrheal Diseases and Temperature, Rainfall, Flooding, and Drought." *Environmental Science & Technology* 50 (10): 4905–22. <https://doi.org/10.1021/acs.est.5b06186>.

- Lin, Audrie, Shahjahan Ali, Benjamin F Arnold, Md Ziaur Rahman, Mohammad Alauddin, Jessica Grembi, Andrew N Mertens, et al. 2020. "Effects of Water, Sanitation, Handwashing, and Nutritional Interventions on Environmental Enteric Dysfunction in Young Children: A Cluster-Randomized, Controlled Trial in Rural Bangladesh." *Clinical Infectious Diseases* 70 (5): 738–47. <https://doi.org/10.1093/cid/ciz291>.
- Lin, Audrie, Ayse Ercumen, Jade Benjamin-Chung, Benjamin F Arnold, Shimul Das, Rashidul Haque, Sania Ashraf, et al. 2018. "Effects of Water, Sanitation, Handwashing, and Nutritional Interventions on Child Enteric Protozoan Infections in Rural Bangladesh: A Cluster-Randomized Controlled Trial." *Clinical Infectious Diseases* 67 (10): 1515–22. <https://doi.org/10.1093/cid/ciy320>.
- Lipsitch, Marc, Eric Tchetgen Tchetgen, and Ted Cohen. 2010. "Negative Controls: A Tool for Detecting Confounding and Bias in Observational Studies." *Epidemiology (Cambridge, Mass.)* 21 (3): 383–88. <https://doi.org/10.1097/EDE.0b013e3181d61eeb>.
- Liu, Zhidong, Yanyu Liu, Ying Zhang, Jiahui Lao, Jing Zhang, Hui Wang, and Baofa Jiang. 2019. "Effect of Ambient Temperature and Its Effect Modifiers on Bacillary Dysentery in Jinan, China." *Science of The Total Environment* 650 (February): 2980–86. <https://doi.org/10.1016/J.SCITOTENV.2018.10.053>.
- Lopez, Gerardo U, Charles P Gerba, Akrum H Tamimi, Masaaki Kitajima, Sheri L Maxwell, and Joan B Rose. 2013. "Transfer Efficiency of Bacteria and Viruses from Porous and Nonporous Fomites to Fingers under Different Relative Humidity Conditions Downloaded From." *Applied and Environmental Microbiology* 79: 5728–34. <https://doi.org/10.1128/AEM.01030-13>.
- Luby, Stephen P., Mahbubur Rahman, Benjamin F. Arnold, Leanne Unicomb, Sania Ashraf, Peter J. Winch, Christine P. Stewart, et al. 2018. "Effects of Water Quality, Sanitation, Handwashing, and Nutritional Interventions on Diarrhoea and Child Growth in Rural Bangladesh: A Cluster Randomised Controlled Trial." *The Lancet Global Health* 6 (1718): 30490–94. [https://doi.org/10.1016/S2214-109X\(17\)30490-4](https://doi.org/10.1016/S2214-109X(17)30490-4).
- M. Friedl, D. Sulla-Menashe. 2019. "MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V006." NASA EOSDIS Land Processes DAAC. <https://doi.org/10.5067/MODIS/MCD12Q1.006>.
- Marra, Giampiero, and Simon N. Wood. 2011. "Practical Variable Selection for Generalized Additive Models." *Computational Statistics & Data Analysis* 55 (7): 2372–87. <https://doi.org/10.1016/j.csda.2011.02.004>.
- . 2012. "Coverage Properties of Confidence Intervals for Generalized Additive Model Components: Coverage Properties of GAM Intervals." *Scandinavian Journal of Statistics* 39 (1): 53–74. <https://doi.org/10.1111/j.1467-9469.2011.00760.x>.
- Mertens, Andrew, Kalpana Balakrishnan, Padmavathi Ramaswamy, Paramasivan Rajkumar, Prabhakar Ramaprabha, Natesan Durairaj, Alan E. Hubbard, Ranjiv Khush, John M. Colford, and Benjamin F. Arnold. 2019. "Associations between High Temperature, Heavy Rainfall, and Diarrhea among Young Children in Rural Tamil Nadu, India: A Prospective Cohort Study." *Environmental Health Perspectives* 127 (4): 047004–047004. <https://doi.org/10.1289/EHP3711>.
- Montealegre, Maria Camila, Subarna Roy, Franziska Böni, Muhammed Iqbal Hossain, Tala Navab-Daneshmand, Lea Caduff, A. S. G. Faruque, Mohammad Aminul Islam, and Timothy R. Julian. 2018. "Risk Factors for Detection, Survival, and Growth of Antibiotic-Resistant and Pathogenic Escherichia Coli in Household Soils in Rural Bangladesh." *Applied and Environmental Microbiology* 84 (24). <https://doi.org/10.1128/AEM.01978-18>.
- Moynihan, Emma L., Karl G. Richards, Fiona P. Brennan, Sean F. Tyrrel, and Karl Ritz. 2015. "Enteropathogen Survival in Soil from Different Land-Uses Is Predominantly Regulated by Microbial Community Composition." *Applied Soil Ecology* 89 (May): 76–84. <https://doi.org/10.1016/j.apsoil.2015.01.011>.

- Peduzzi, Peter, John Concato, Elizabeth Kemper, Theodore R. Holford, and Alvan R. Feinstein. 1996. "A Simulation Study of the Number of Events per Variable in Logistic Regression Analysis." *Journal of Clinical Epidemiology* 49 (12): 1373–79. [https://doi.org/10.1016/S0895-4356\(96\)00236-3](https://doi.org/10.1016/S0895-4356(96)00236-3).
- Pekel, Jean-François, Andrew Cottam, Noel Gorelick, and Alan S. Belward. 2016. "High-Resolution Mapping of Global Surface Water and Its Long-Term Changes." *Nature* 540 (7633): 418–22. <https://doi.org/10.1038/nature20584>.
- Petersen, Maya L, Kristin E Porter, Susan Gruber, Yue Wang, and Mark J van der Laan. 2012. "Diagnosing and Responding to Violations in the Positivity Assumption." *Statistical Methods in Medical Research* 21 (1): 31–54. <https://doi.org/10.1177/0962280210386207>.
- Platts-Mills, James A., Sudhir Babji, Ladaporn Bodhidatta, Jean Gratz, Rashidul Haque, Alexandre Havt, Benjamin J J McCormick, et al. 2015. "Pathogen-Specific Burdens of Community Diarrhoea in Developing Countries: A Multisite Birth Cohort Study (MAL-ED)." *The Lancet Global Health* 3 (9): 564–75. [https://doi.org/10.1016/S2214-109X\(15\)00151-5](https://doi.org/10.1016/S2214-109X(15)00151-5).
- Poulsen, Melissa N., Jonathan Pollak, Deborah L. Sills, Joan A. Casey, Sara G. Rasmussen, Keeve E. Nachman, Sara E. Cosgrove, Dalton Stewart, and Brian S. Schwartz. 2018. "Residential Proximity to High-Density Poultry Operations Associated with Campylobacteriosis and Infectious Diarrhea." *International Journal of Hygiene and Environmental Health* 221 (2): 323–33. <https://doi.org/10.1016/J.IJHEH.2017.12.005>.
- Pullan, Rachel L, Peter W Gething, Jennifer L Smith, Charles S Mwandawiro, Hugh J.W. Sturrock, Caroline W Gitonga, Simon I Hay, and Simon Brooker. 2011. "Spatial Modelling of Soil-Transmitted Helminth Infections in Kenya: A Disease Control Planning Tool." *PLoS Neglected Tropical Diseases* 5 (2). <https://doi.org/10.1371/journal.pntd.0000958>.
- Rosef, Olav, Grethe Rettedal, and Lars Lågeide. 2001. "Thermophilic Campylobacters in Surface Water: A Potential Risk of Campylobacteriosis." *International Journal of Environmental Health Research* 11 (4): 321–27. <https://doi.org/10.1080/09603120120081791>.
- Schwab, Benjamin, and Ralph Armah. 2019. "Can Food Safety Shortfalls Disrupt 'Ag for Nutrition' Gains? Evidence from Eid al-Adha." *Food Policy* 83 (February): 170–79. <https://doi.org/10.1016/j.foodpol.2019.01.002>.
- Stine, Scott W, † Inhong Song, Christopher Y Choi, and Charles P Gerba. 2005. "Effect of Relative Humidity on Preharvest Survival of Bacterial and Viral Pathogens on the Surface of Cantaloupe, Lettuce, and Bell Peppers." *Journal of Food Protection* 68 (7): 1352–58. <https://doi.org/10.4315/0362-028X-68.7.1352>.
- Teklehaimanot, Giorgis Z., B. Genthe, I. Kamika, and M. N. B. Momba. 2015. "Prevalence of Enteropathogenic Bacteria in Treated Effluents and Receiving Water Bodies and Their Potential Health Risks." *Science of The Total Environment* 518–519 (June): 441–49. <https://doi.org/10.1016/j.scitotenv.2015.03.019>.
- Valcour, James E., Dominique F. Charron, Olaf Berke, Jeff B. Wilson, Tom Edge, and David Waltner-Toews. 2016. "A Descriptive Analysis of the Spatio-Temporal Distribution of Enteric Diseases in New Brunswick, Canada." *BMC Public Health* 16 (1): 204–204. <https://doi.org/10.1186/s12889-016-2779-5>.
- Walters, Sarah P., Anne L. Thebo, and Alexandria B. Boehm. 2011. "Impact of Urbanization and Agriculture on the Occurrence of Bacterial Pathogens and Stx Genes in Coastal Waterbodies of Central California." *Water Research* 45 (4): 1752–62. <https://doi.org/10.1016/J.WATRES.2010.11.032>.
- Wan, Z., S. Hook, and G. Hulley. 2015. "MOD11A1 MODIS/Terra Land Surface Temperature/Emissivity Daily L3 Global 1km SIN Grid V006." NASA EOSDIS Land Processes DAAC. <https://doi.org/10.5067/MODIS/MOD11A1.006>.
- Whitmee, Sarah, Andy Haines, Chris Beyrer, Frederick Boltz, Anthony G. Capon, Braulio Ferreira de Souza Dias, Alex Ezech, et al. 2015. "Safeguarding Human Health in the Anthropocene Epoch:

- Report of The Rockefeller Foundation–Lancet Commission on Planetary Health.” *The Lancet* 386 (10007): 1973–2028. [https://doi.org/10.1016/S0140-6736\(15\)60901-1](https://doi.org/10.1016/S0140-6736(15)60901-1).
- Wood, Simon N. 2006. *Generalized Additive Models: An Introduction with R*. Texts in Statistical Science. Boca Raton, FL: Chapman & Hall/CRC.
- Worldpop. 2017. “Bangladesh 100m Population.” Universtiy of Southampton. <https://doi.org/10.5258/SOTON/WP00533>.
- Xie, Pingping, Mingyue Chen, Song Yang, Akiyo Yatagai, Tadahiro Hayasaka, Yoshihiro Fukushima, and Changming Liu. 2007. “A Gauge-Based Analysis of Daily Precipitation over East Asia.” *Journal of Hydrometeorology* 8 (3): 607–26. <https://doi.org/10.1175/JHM583.1>.

Appendix

Table A1. Real time PCR assays on the TaqMan Array Card and associated gene targets. Assays have been described previously and extensively validated (Liu 2013; Liu 2016a, Liu 2016b). Nucleic acid was extracted with the QIAamp Fast DNA Stool mini kit (Qiagen, Hilden, Germany) with pre-treatment steps that included bead beating. AgPath One Step RT-PCR reagents were used for qPCR reactions, which were performed on ViiA 7 systems. Quantification cycles (C_q s) are the PCR cycle values at which fluorescence from amplification exceeds the background, which acts as an inverse metric of quantity of nucleic acid. Valid results required proper functioning of controls (the negative results of a sample are valid only when its external control MS2 is positive, $C_q < 35$; the positive results are valid only when the corresponding extraction blank is negative for the relevant targets, $C_q > 35$), and excluded data flagged by the real time PCR software, i.e., BADROX in combination with NOISE or SPIKE.

Pathogen	Gene
Virus	
Adenovirus 40/41	Fiber gene
Astrovirus	Capsid
Norovirus GI/GII	GI ORF1-2 and GII ORF1-2
Rotavirus	<i>NSP3</i>
Sapovirus	<i>RdRp</i>
Bacteria	
Enteraggregative <i>Escherichia coli</i> (EAEC)*	<i>aaiC</i> , <i>aatA</i>
Enteropathogenic <i>E. coli</i> (EPEC)*	<i>bfpA</i> , <i>eae</i>
Enterotoxigenic <i>E. coli</i> (ETEC)*	<i>LT</i> , <i>STh</i> , <i>STp</i>
Shiga toxin-producing <i>E. coli</i> (STEC)*	<i>stx1</i> , <i>stx2</i>
<i>Aeromonas</i>	Aerolysin
<i>Bacteroides fragilis</i>	EGBF
<i>Campylobacter</i> spp.	<i>cpn60</i>
<i>Campylobacter jejuni/coli</i>	<i>cadF</i>
<i>Clostridium difficile</i>	<i>tcdA</i> , <i>tcdB</i>
<i>Helicobacter pylori</i>	<i>ureC</i>
<i>Plesiomonas shigelloides</i>	<i>gyrB</i>
<i>Salmonella enterica</i>	<i>ttr</i>
<i>Shigella</i> spp./Enteroinvasive <i>E. coli</i> (EIEC)	<i>ipaH</i>
<i>Vibrio cholerae</i>	<i>hlyA</i>
Fungi	
<i>Encephalitozoon intestinalis</i>	SSU rRNA
<i>Enterocytozoon bienersi</i>	<i>ITS</i>
Protozoa	
<i>Cryptosporidium</i> spp.	18S rRNA
<i>Entamoeba histolytica</i>	18S rRNA
<i>Entamoeba</i> spp.	18S rRNA
<i>Giardia</i> spp.	18S rRNA
<i>Cyclospora cayetanensis</i>	18S rRNA
<i>Cystoisospora belli</i>	18S rRNA
Helminth	
<i>Ancylostoma duodenale</i>	<i>ITS2</i>
<i>Ascaris lumbricoides</i>	<i>ITS1</i>
<i>Blastocystis</i> spp.	18S rRNA
<i>Hymenolepis nana</i>	<i>ITS1</i>
<i>Necator americanus</i>	<i>ITS2</i>
<i>Strongyloides stercoralis</i>	Dispersed repetitive sequence
<i>Schistosoma</i> spp.	<i>ITS</i>
<i>Trichuris trichiura</i>	18S rRNA
Controls	
MS2	<i>MS2g1</i>
PhHV	<i>gB</i>

* *E. coli* pathotypes were defined as follows: EAEC (*aaiC*, or *aatA*, or both), atypical EPEC (*eae* without *bfpA*, *stx1*, and *stx2*), typical EPEC (*bfpA* and *eae*), ETEC (*STh*, *STp*, or *LT*), STEC (*eae* without *bfpA* and with *stx1*, *stx2*, or both).

Figure A1. Empirically established annual rainy season during the study period. Rainy season was defined as the period when the rolling 5-day average rainfall was $\geq 10\text{mm/day}$. The colored lines depict the beginning (green) and end (red) of the rainy season. Note: 2016 rainy season was truncated to the end of the study period.

