Comparison of fecal indicators with pathogenic bacteria and rotavirus in groundwater

Andrew S. Ferguson [a],*, Alice C. Layton [b], Brian J. Mailloux [c], Patricia J. Culligan [a], Daniel E. Williams [b], Abby E. Smartt [b], Gary S. Sayler [b], John Feighery [d], Larry D. McKay [e], Peter S.K. Knappett [e], Ekaterina Alexandrova [c], Talia Arbit [c], Michael Emch [f], Veronica Escamilla [f], Kazi Matin Ahmed [g], Md. Jahangir Alam [g], P. Kim Streatfield [h], Mohammad Yunus [h], Alexander van Geen [i]

[a] Department of Civil Engineering and Engineering Mechanics, Columbia University, NY, USA
[b] Department of Microbiology and Center for Environmental Biotechnology, University of Tennessee, Knoxville, TN, USA
[c] Department of Environmental Sciences, Barnard College, NY, USA
[d] Department of Earth and Environmental Engineering, Columbia University, NY, USA
[e] Department of Earth and Planetary Sciences, University of Tennessee, Knoxville, TN, USA
[f] Department of Geography, University of North Carolina at Chapel Hill, NC, USA
[g] Department of Geology, University of Dhaka, Dhaka, Bangladesh
[h] International Center for Diarrhoeal Research, Bangladesh, Dhaka, Bangladesh
[i] Lamont-Doherty Earth Observatory of Columbia University, Palisades, NY 10964, USA

ARTICLE INFO

ABSTRACT

Groundwater is routinely analyzed for fecal indicators but direct comparisons of fecal indicators to the presence of bacterial and viral pathogens are rare. This study was conducted in rural Bangladesh where the human population density is high, sanitation is poor, and groundwater pumped from shallow tubewells is often contaminated with fecal bacteria. Five indicator microorganisms (*E. coli*, total coliform, F+RNA coliphage, *Bacteroides* and human-associated *Bacteroides*) and various environmental parameters were compared to the direct detection of waterborne pathogens by quantitative PCR in groundwater pumped from 50 tubewells. Rotavirus was detected in groundwater filtrate from the largest proportion of tubewells (40%), followed by *Shigella* (10%), *Vibrio* (10%), and pathogenic *E. coli* (8%). Spearman rank correlations and sensitivity–specificity calculations indicate that some, but not all, combinations of indicators and environmental parameters can predict the presence of pathogens. Culture-dependent fecal indicator bacteria measured on a single date did not predict total bacterial pathogens, but annually averaged monthly measurements of culturable *E. coli* did improve prediction for total bacterial pathogens. A qPCR-based *E. coli* assay was the best indicator for the bacterial pathogens. F+RNA coliphage were neither correlated nor sufficiently sensitive towards rotavirus, but were predictive of bacterial pathogens. Since groundwater cannot be excluded as a significant source of diarrheal disease in Bangladesh and neighboring countries with similar characteristics, the need to develop more effective methods for screening tubewells with respect to microbial contamination is necessary.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Human pathogenic microorganisms, including rotavirus (Lodder et al., 2010; Verheyen et al., 2009), *Shigella* (Faruque et al., 2002), *Vibrio* (Alam et al., 2006) and pathogenic *Escherichia coli* (*E. coli*) (Begum et al., 2007; Momba et al., 2006) are readily detected worldwide in surface water and more recently in low concentrations in groundwater (Borchardt et al., 2003a; Gibson and Schwab, 2011; Hunt et al., 2010; Wampler and Sisson, 2011). The increasing identification of

pathogens in groundwater is troubling because there is a general belief that groundwater does not require treatment before drinking. This is of particular concern in densely populated areas residing on shallow sandy aquifers without adequate sanitation as found in Bangladesh (Leber et al., 2011; Luby et al., 2008; van Geen et al., 2011). The installation of tens of millions of shallow tubewells throughout rural South and Southeast Asia in order to provide groundwater as a primary source of drinking water was expected to greatly reduce the exposure to waterborne pathogens and the incidence of diarrheal disease. However, across the developing world diarrheal disease still remains a leading cause of illness and mortality in children under five (Kosek et al., 2003) and little information is available with

---

respect to the extent that tubewell water may still serve as a source for waterborne pathogens.

The detection of waterborne pathogens is expensive, time consuming, and complex due to pathogen variability. Consequently, indicators of fecal contamination are routinely used by environmental agencies and health organizations to monitor water quality with an emphasis on culture-dependent assays for total coliforms and *E. coli* as fecal indicator bacteria (FIB). In relation to groundwater, the occurrence of total coliforms is interpreted as an indication of contaminated surface water infiltration and the potential occurrence of fecal pathogens (Payment and Locas, 2010). However, the recently documented survival and proliferation of FIB within soils and sediments (Brennan et al., 2010; Pote et al., 2009) and tubewell hand-pumps (Ferguson et al., 2011), make it potentially more difficult to correlate pathogens with FIB that were originally present in recharge water. Also, there is little reason to believe that viral presence will be reflected by FIB, given the different survival and transport characteristics of these two classes of microorganisms. For this reason, coliphage is the proposed surrogate for human enteric viruses (Bushon, 2003). More sophisticated culture-independent (Schriewer et al., 2010), isotope (Boehm et al., 2004), and chemical (Glassmeyer et al., 2005) assays for determining fecal contamination have been proposed to address the limitations associated with traditional culture-dependent assays. However, such methods are impractical in developing countries where the need is greatest for 'ASSURED' (Affordable, Sensitive (few false-negatives), Specific, (few false-positives), User-friendly, Rapid, Equipment-free, and Delivered to those who need it) diagnostic testing of drinking water quality (Urdea et al., 2006).

It is with this need in mind that groundwater samples were collected for parallel analysis of FIB and pathogens from a set of 50 tubewells within a single village of Bangladesh. The original motivation for this work was that shallow low-Arsenic (As) wells might be more prone to contamination with FIB than shallow high-As wells and, therefore, inducing households to switch from a high- to a low-As well might increase their exposure to microbial pathogens (Ahmed et al., 2006; van Geen et al., 2011). The location and timing of sample collection were based on the high levels of FIB previously measured in tubewell water, particularly during the monsoon (Leber et al., 2011; van Geen et al., 2011). These tubewells were also previously characterized for depth, platform quality, the need to prime (or not prime) the well and the human population residing within a 25 m and 50 m radius of each well (van Geen et al., 2011). Groundwater from each well was tested for total coliform, *E. coli*, and F+RNA coliphage using established culture-dependent methods and water quality parameters. Groundwater was also filtered and the filters preserved for analysis in the laboratory by culture-independent methods for pathogenic (*Shigella*, *Vibrio*, *E. coli*, and rotavirus) and non-pathogenic markers (*E. coli*, *Bacteroides* and human-associated *Bacteroides* (Hu*Bacteroides*)).

2. Materials and methods

2.1. Study site

The 50 tubewells selected for sampling between late July and early August, 2009, are located in the village of Bara Haldia (23.370°N; 90.646°E) in Matlab upazila, approximately 45 km southeast of Dhaka (Fig. A1). The location of each monitored well was measured by differential GPS. Consent was obtained from a senior member of the household before all samples were taken, which was approved by the Institutional Review Boards of ICDDRB and Columbia University. The village is located within an embankment built for flood protection and sits on a 3–6 m clay layer underlain by a shallow sandy aquifer. Throughout the Bengal Basin, tubewells are typically constructed out of 2.5 inch diameter PVC pipe with a 5 foot screen and are installed by the traditional hand-flapper (mud circulation)

method. The bottom portion of well annulus where the slotted-PVC filter is located is first filled with coarse sand. The remaining portion of the annulus is packed with sediment removed during drilling and sometimes capped using clay. A grout seal is rarely used for privately installed household wells. Some households install a concrete and brick platform around their tubewell and others do not. Groundwater is typically extracted with a cast-iron hand-pump. The tubewells range between 8 and 37 m (25–120 ft) in depth and were selected from a set of 81 wells within Matlab upazila that had been monitored for total coliforms and *E. coli* quasi-monthly (van Geen et al., 2011). Additional information available from these tubewells, including the average and range of water quality, fecal indicator and pathogen concentration, is summarized in Table 1. The complete dataset is available as supplementary data.

2.2. Field measurements and sampling

Prior to groundwater sample collection approximately twice the volume of water within the entire well casing was purged with the existing tubewell hand-pump at a rate of 30–40 L/min, irrespective of well depth. Wells which required priming were done so using stored water from the same pump. At each site, tubewell water (2–8 L) was filtered for DNA and RNA based molecular methods and culture-dependent F+RNA coliphage. Groundwater was collected, stored and transported at room temperature in 20 mL scintillation vials with PolySeal caps for As, Fe, Mn, and S analysis by high-

Table 1
Summary of tubewell and groundwater characteristics. Platform quality was assigned as either good (0) or cracked and deteriorating (1). Microbial indicator and pathogen data were log transformed gene copies/100 mL. Units are: depth (ft), temperature (°C), EC (mS/cm), ORP (mV), Salinity (ppt), As (ppb), NH_3 (ppm/N), NO_2 (ppm/N), NO_3 (ppm/N), PO_4 (ppm/N), SO_4 (ppm), Br (ppm), Cl (ppm), Cl/Br (ppm), total coliform and culturable *E. coli* (MPN/100 mL) and coliphage (PFU/100 mL). Averaged frequency of culturable *E. coli* detection at >1 MPN/100 mL was calculated annually and for the 2008–2009 monsoon.

	Variable	Mean	SD	Min	Max
Tubewell feature	Platform quality	0.36	0.48	0.00	1.00
	Mo. priming	0.07	0.09	0.00	0.38
	Depth	54.76	19.82	25.00	120.00
	Population (25 m)	21.54	11.78	5.00	51.00
	Population (50 m)	53.44	30.40	6.00	130.00
Water chemistry	Temp	26.43	0.34	25.83	27.33
	pH	6.43	0.23	6.02	6.90
	EC	1.43	0.87	0.34	3.93
	ORP	101.82	82.83	−207.70	199.60
	Sal	0.72	0.45	0.16	2.06
	As	87.65	111.55	0.62	435.87
	NH_3	2.98	2.68	0.00	10.00
	NO_2	2.34	3.57	0.00	18.41
	NO_3	2.02	4.65	0.00	25.75
	PO_4	0.33	0.47	0.00	2.86
	SO_4	4.83	20.30	0.00	126.67
	Cl	238.74	260.84	4.44	933.85
	Br	1.25	1.26	0.00	4.72
	ClBr	338.44	170.66	54.98	769.09
Culture dependent indicators	Total coliform	243.48	656.95	0.25	>2419
	E. coli	85.48	434.78	0.25	>2419
	Freq. *E. coli* annual	0.41	0.15	0.14	0.68
	Freq. *E. coli* monsoon	0.43	0.20	0.13	0.88
	F+RNA coliphage	15.41	32.10	0.025	145.95
	m*E. coli*	2.06	1.05	0.60	5.18
Culture independent	*Bacteroides*	3.28	1.53	0.60	7.45
	Hu*Bacteroides*	1.19	1.36	0.60	6.82
Pathogen gene	log ipaH	0.74	0.45	0.60	2.68
	log elt	0.61	0.05	0.60	0.99
	log eae	0.67	0.33	0.60	2.77
	log stx	0.62	0.13	0.60	1.52
	log ompW	0.66	0.20	0.60	1.74
	log rota	1.83	1.38	0.85	5.14

resolution inductively-coupled plasma mass spectrometry (Cheng et al., 2004), and at −20 °C in 15 mL centrifuge tubes (Fisher Scientific, Pittsburgh, PA) for fluoride, chloride, bromide, nitrite, nitrate, sulfate and phosphate analysis using a Dionex DX2000 ion chromatograph in gradient mode equipped with an AS-11HC column (Dionex, Sunnyvale, CA). Samples were analyzed at the tubewell for pH, ORP, specific conductance, and temperature (YSI, Yellowsprings, Ohio), and ammonia (Chemetrics, Calverton, VA).

2.3. Fecal bacteria sampling and MPN analysis

Groundwater samples were collected directly from the hand-pump discharge in duplicate sterile 100 mL bottles containing sodium thiosulfate and stored at 4 °C in the dark until incubation for measuring total coliform and *E. coli*. Samples were analyzed for total coliforms and culture based *E. coli* using the MPN based Colilert™ assay (IDEXX Laboratories, Inc) within 8 h of collection according to the manufacturer's instructions. The MPN solution of Hurley and Roscoe (1983) was used to determine an MPN/100 mL and respective 95% confidence intervals. The detection limit based on duplicate measurements was 0.5 MPN/100 mL. For undiluted samples, the range of the method (>2419 MPN/100 mL) was exceeded by one sample for *E. coli* and two for total coliforms. Two sterile water blanks were analyzed during each sampling period and were always negative for total coliforms or *E. coli*. As part of a larger monitoring study of 113 wells (van Geen et al., 2011), the annually-averaged frequency of culturable *E. coli* detection at >1 MPN/100 mL was calculated for the subset of 50 wells.

2.4. Groundwater filtration

Duplicate groundwater (2–8 L) samples were filtered through two nitrocellulose 0.2 μm filters (Corning, Corning, NY) using a vacuum pump (Gast, Benton Harbor, MI). One filter was utilized for bacterial molecular analyses and the second filter was used for F+RNA coliphage and virus molecular analyses. Groundwater was filtered directly for bacterial analyses. The method of Sinton et al. (1996) was used for virus and F+RNA coliphage samples. Before filtration, the water sample was acidified to pH 2.5–3 with HCl (1:1 v/v) and amended with MgCl2 to a final concentration of 1 mM. Following filtration, half of the acidified filter sample was cut from its housing, placed into a 50 mL tube containing 5 mL beef extract solution and stored at 4 °C in the dark for up to 8 h. The beef extract solution was 0.75 M NaCl, 1% beef extract, and 3% Tween 80. The remaining filters were cut from their housing, placed into separate sterile petri dishes, sealed with parafilm, immediately placed on dry-ice and stored at −20 °C for subsequent molecular analysis. Water samples (50 mL) were also collected from each sample point for use in F+RNA coliphage elution.

2.5. Enumeration of coliphage

E. coli (ATCC#700891) (Debartolomeis and Cabelli, 1991) was used as the specific F+RNA coliphage host in all phage assays. Host cells were grown in tryptic soy broth (Fisher Scientific) with 50 μg/mL ampicillin and streptomycin, and incubated at 37 °C for 4 h. The acidified phage filter, eluted in beef extract, was amended with its corresponding water sample (25 mL) and 150 μL MgCl2 (1.33 M). Water samples were then screened for the presence of F+RNA coliphage using the Easyphage test kit (Scientific Methods, Granger, Indiana) according to manufacturer's instructions. The detection limit for F+RNA coliphage was ½ the lowest count we could achieve or 2.5×10^{-2} PFU/100 mL. F+RNA strain MS2 (ATCC#15597-B1) was used as a positive control.

2.6. Extraction, quantification and cloning of DNA and RNA from filters

DNA and RNA were extracted from all tubewell filter samples. In the laboratory, ¼ to ½ of each filter was sliced into ~1 mm strips and placed into bead beat tubes (Fast DNA Spin Kit for Soil or Fast RNA Spin Kit for Soil, MP Biomedicals, LLC, Solon, OH). Subsequent DNA and RNA extraction were performed according to the manufacturer's procedures and all extractions were stored at −20 °C. DNA and RNA were quantified and diluted to 1–2 ng/μL for quantitative PCR assays.

Fecal indicator (*E. coli* 23S rRNA, *Bacteroides*, and Hu*Bacteroides*) genes were assayed in triplicate using previously described methods (Knappett et al., 2011). In addition, spike controls to survey for PCR inhibition were run for all samples as well as negative controls to survey for cross contamination. Pathogen genes were assayed in duplicate using the primers, probes and master mix types listed in Table A1 and following standard quantitative PCR protocols (Layton et al., 2006). Bacterial standards were made from a relevant gene fragment cloned into PCR4-TOPO cloning vector (Layton et al., 2006). Apart from an initial incubation step of 55 °C for 30 min for the reverse transcriptase step and variations in the annealing temperature as indicated in Table A1, the QRT-PCR and QPCR amplification protocols were identical. The rotavirus standard consisted of a synthesized 200 nucleotide oligomer based on rotavirus VP6 from X94617 contained in a plasmid (GenScript, Piscataway, NJ). Data were calculated only for samples in which both PCR reactions had >1 gene copy and were quantified as copies/μL nucleic acid extract. Gene copies were adjusted to copies/100 mL for tubewell water based on the fraction of the filter extracted, multiplied by the volume of DNA extract and divided by the filtered sample volume. The detection limits for the PCR assays varied due to differences in the amount of tube well water filtered and nucleic acid extracted per tubewell. However, the mean detection limit was ~4 copies/100 mL and ~7 copies/100 mL for DNA and RNA based assays, respectively. Molecular data was log transformed.

Following QPCR or QRT-PCR, PCR products in selected positive wells were purified using the QIAquick PCR Purification Kit (Qiagen Sciences, Germantown, MD). Two μLs of the purified PCR product was reamplified using FastStart PCR master mix (Roche Applied Science, Indianapolis, IN) and the primers used for the qPCR reaction (Table A1) using previously described PCR conditions (Layton et al., 2006). Two μLs of this second PCR product was ligated into the PCR4-TOPO vector, transformed into *E. coli* top10 cells (TOPO TA Cloning Kit, Invitrogen) and plated on to LB media containing 50 μg/mL Kanamycin in accordance to the manufacturer instructions. Plasmids in individual *E. coli* colonies were screened for inserts, sequenced using M13f and M13r primers at the University of Tennessee Molecular Biology Resource Facility and identified using the BLASTN program at National Center for Biotechnology Information (NCBI) (Altschul et al., 1990).

2.7. Statistical analysis

Data was analyzed using PASW software version 18. The correlations between indicator and pathogen concentrations were determined by Spearman rank correlations. Binary outcomes were also calculated for various combinations of pathogens to compensate for the limited number of detections. Binary and multinomial logistic regression (0–1 for pathogens) was used to determine whether field based indicators or environmental parameters used individually or in combination were predictive of the presence or absence of a pathogen or groups of pathogens. Sensitivity and specificity were calculated for each indicator following their comparison with the individual or combined viral and/or bacterial pathogens at a variety of cutoff values. Sensitivity is the fraction of time that an indicator correctly corresponds with the presence of a pathogen in

tubewell water. High sensitivity is required to ensure that a tubewell contaminated with a pathogen is identified. Specificity is the second desirable characteristic of an indicator and represents an indicator's ability to correctly identify tubewells without pathogens. If the specificity is too low, then the number of false positives will be high and the usefulness of the indicator is lost. Water quality parameters may be statistically correlated but not necessarily good indicators of each other because of either low sensitivity or specificity (Pepe et al., 2004).

For combining two indicators into one new parameter, logistic regression was utilized to determine sensitivity and specificity towards different combinations of pathogen groups according to function (Eq. (1)):

$$z = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \tag{1}$$

where β_0 is the intercept, and β_1 and β_2 are the regression coefficients for parameters x_1 and x_2. The logistic function (z) was calculated for every possible paired combination of indicators using a program written in Matlab (version 7.10.0.499, The MathWorks, Inc.).

3. Results

3.1. Fecal indicators and pathogens in tubewell water

Culturable *E. coli*, total coliforms, and coliphage (F+RNA) were detectable in groundwater from 40%, 92%, and 64% of the 50 tubewells sampled over 8 days during the monsoon, with average concentrations shown in Table 1. Although concrete or brick platforms were either cracked or non-existent for 18 out of the 50 sampled wells, culturable *E. coli* was not detected in most (n = 14) of these wells. Using WHO risk categories based on culturable *E. coli* (WHO, 2008), water pumped from 30 (60%) tubewells met the WHO guideline for drinking water (<1 MPN/100 mL), 14 (28%) were within the low risk category (1–10 MPN/100 mL), and the remaining 6 (12%) in the intermediate to very high risk category (10 to >1000 MPN/100 mL) (Table 2). Within each of the three risk categories defined by culturable *E. coli*, the mean of total coliform concentrations increases from 9 to 220 and 1470 MPN/100 mL, respectively. The mean concentrations of coliphage (F+RNA) increased from 4.9 to 25.6 and 38.9 PFU/100 mL across the same risk categories.

Molecular *E. coli* 23S rRNA genes (m*E coli*) were detected in 43 (86%) tubewell samples. Mean concentrations of m*E coli* increased from 1.4×10^2 to 4.1×10^2 and 4.1×10^4 genes/100 mL across the same three risk categories (Table 2). *Bacteroides* and Hu*Bacteroides* were detected in 43 (86%) and 13 (23%) tubewell samples, respectively. Across the same three risk categories defined by culturable *E. coli*, mean concentrations increased from 7.3×10^3 to 1.3×10^4

and 5.7×10^6 genes/100 mL for *Bacteroides,* and from 19.1 to 9.6×10^2 and 1.3×10^6 genes/100 mL for Hu*Bacteroides* (Table 2).

Of the 50 groundwater samples collected, 23 (46%) were positive for at least one microbial pathogen. Rotavirus was the dominant pathogen detected in 20 (40%) tubewell samples, followed by *Shigella* (10%), *Vibrio* (10%), and pathogenic *E. coli* (8%) (Table 2). Even when combined, bacterial pathogens were detected in a lower proportion of wells (24%) compared to rotavirus. Only three tubewells containing bacterial pathogen genes (*ipa*H, *elt*A, *stx*1, *vtx*1, *eae*, or the non-toxigenic Vibrio *ompW*) did not contain rotavirus (Fig. 1a and b). Pathogenic *E. coli* and *Shigella*, on the other hand, were never detected in the same tubewell.

Within the low risk category defined by culturable *E. coli* (n = 30), rotavirus was present in 40% of samples, *Shigella* in 10% of samples, *Vibrio* in 13% of samples, whereas pathogenic *E. coli* were not detected (Table 2). Within the intermediate risk category (n = 14), the proportion of samples with detectable pathogens increases for rotavirus (57%) and pathogenic *E. coli* (21%) but not for *Vibrio* (7%) and *Shigella* (7%). The mean of rotavirus concentrations was not markedly different between samples containing <1 and 1–10 MPN/100 mL culturable *E. coli* at 8.5 and 9.8×10^3 genes/100 mL, respectively. Rotavirus and *Vibrio* were not detected in any of the 6 samples within the highest risk category whereas *Shigella* and pathogenic *E. coli* were detected in only one well sample (17%). No pathogens of any type were detected in the only sample in the very high risk category (>1000 MPN/100 mL *E. coli*). The proportion of wells affected by various pathogens within risk categories defined by culturable *E. coli* did not change markedly when considering only those wells with a good platform.

Using the frequency of culturable *E. coli* detection averaged over a year (van Geen et al., 2011) to categorize indicator and pathogen occurrence instead of *E. coli* measured on the day of pathogen sampling only, half of the wells (52%) were in the 25%–50% detection frequency range (Table 2). The remaining wells were equally split between the >50% and <25% categories. The mean concentrations of total coliform, culturable *E. coli*, and m*E. coli* in the 50 tubewells increased with the annual frequency of detection categories. Total bacterial pathogens were not detected in the <25% annual detection frequency category. Detection was highest for F+RNA coliphage, *Bacteroides*, Hu*Bacteroides*, and total bacterial pathogen in the 25%–50% range of detection frequency. Rotavirus was detected in all three frequency categories (Table 2).

3.2. Verification of molecular assays

To verify the presence of pathogenicity gene sequences in tubewell water, attempts were made to clone representative PCR products from positive tubewells. Although the *Shigella ipa*H and *E. coli elt*A

Table 2
Detection of fecal indicators, bacterial pathogens and rotavirus categorized against no criteria, WHO risk category limits, and annually-averaged *E. coli* detection frequency (>1 MPN/100 mL).

| Criteria	Descriptor	Num. wells	Mean						Percentage of wells positive			
			E. coli (MPN/ 100 mL)	Total coliforms (MPN/100 mL)	F+RNA (PFU/ 100 mL)	m*E. coli* (genes/ 100 mL)	*Bacteroides* (genes/ 100 mL)	Hu*Bacteroides* (genes/ 100 mL)	Pathogenic *E. coli*	*Shigella*	*Vibrio*	Rotavirus
None	All wells	50	85.5	243.5	15.4	5147.8	692,788.1	154,412.7	8	10	10	40
WHO risk categories	10 to 1000	6	701.0	1470.0	38.9	41,207.7	5,707,285.0	1,284,441.5	17	17	0	0
	1 to 10	14	4.3	220.3	25.6	414.9	12,574.9	958.2	21	7	7	57
	<1	30	0.3	9.0	4.9	144.5	7321.5	19.1	0	10	13	40
Annually-averaged *E. coli*	>50%	13	315.5	672.5	9.0	13,500.7	468,037.7	80,926.7	8	0	8	23
	25–50%	26	6.4	125.8	25.0	3117.1	1,094,590.4	256,481.5	12	19	15	46
	<25%	11	0.6	14.5	2.9	76.0	8687.6	6.6	0	0	0	45
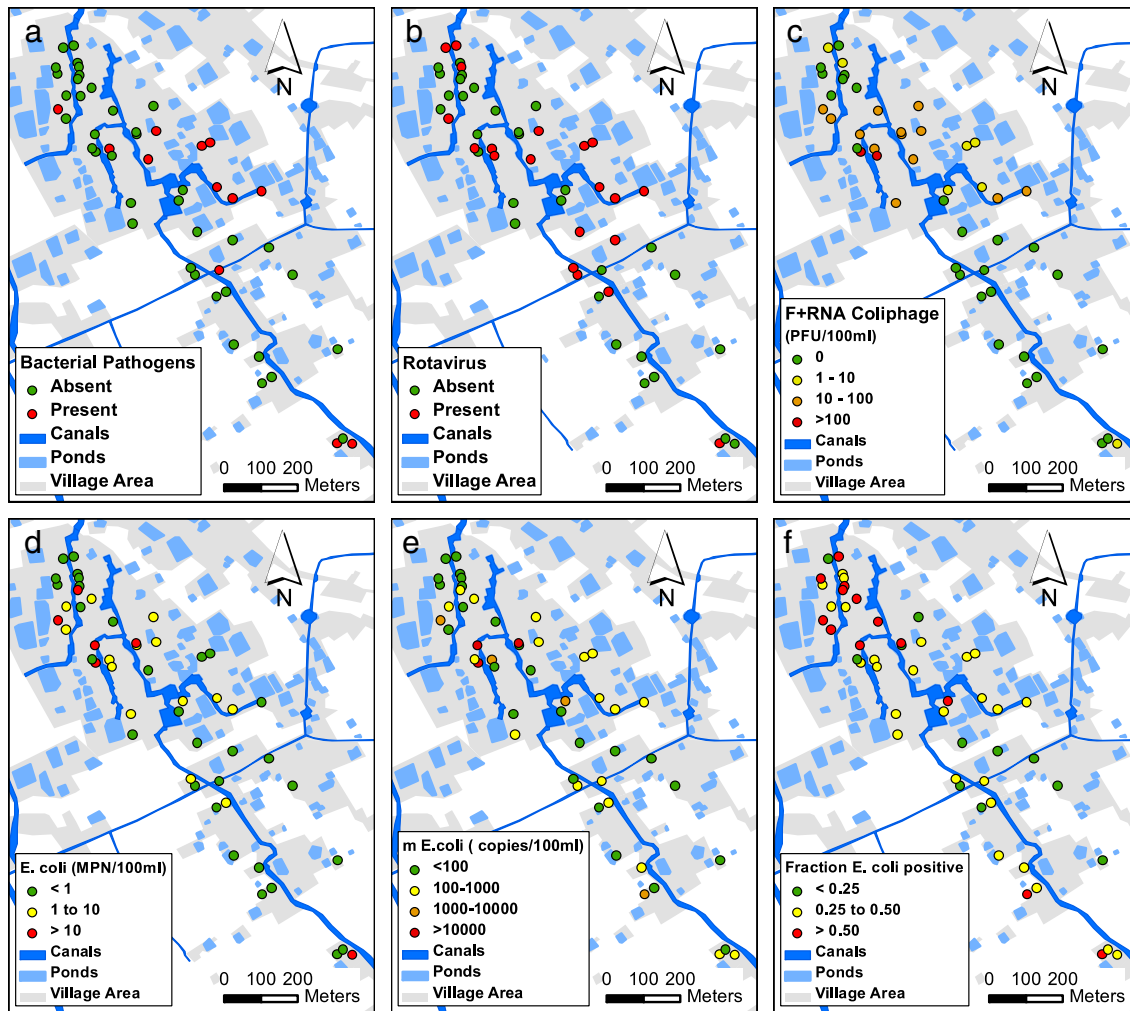
Fig. 1. Spatial distribution of microbial indicators (a) total bacterial pathogens (pathogenic *E. coli*, *Shigella*, and *Vibrio* spp.), (b) rotavirus, (c) F+RNA coliphage, (d) culturable *E. coli*, (e) m*E. coli*, and (f) annually-averaged *E. coli* detection frequency (>1 MPN/100 mL) from tubewells in Bara Haldia.

were cloned from surface water samples (data not shown), cloning of these genes from groundwater samples was unsuccessful. This may reflect their very low concentration in the groundwater samples. In contrast, rotavirus gene sequences could be cloned from a subset of positive well samples (9 total). The gene products from rotavirus were then sequenced to verify their presence in the well water samples. The majority of these sequences (37 out of 39) were 99% similar to G12 rotavirus (EU839971). The other 2 sequences were only 90% similar to the G12-like sequences. The second QRT-PCR assay used in this study was designed based on 37 G12-like sequences from tubewell samples to more specifically target the G12 VP6. This assay confirmed the presence of rotavirus in 7 of the 15 tubewell samples testing positive using the SYBR green assay. Thus, the presence of rotavirus was verified by two separate quantitative methods and by the identification of sequences from 9 different tubewell samples.

3.3. Correlation of pathogens with fecal indicators and environmental parameters

The data indicate some statistically significant Spearman rank correlations of pathogens with fecal indicators or environmental parameters but none are very strong correlations. Spearman rank correlations indicate that the presence of rotavirus was not correlated to the

log-transformed concentration of any fecal indicator (Table 3). However, the presence of rotavirus was correlated with the combined presence of three bacterial pathogens, pathogenic *E. coli*, *Shigella*, and *Vibrio* and the presence of only pathogenic *E. coli* and *Shigella* combined. With respect to environmental parameters, only the negative correlation between rotavirus detection and the level of nitrate in well water was significant. Within this data set, population density and platform quality were not correlated with the presence of rotavirus in well water.

Pathogenic *E. coli* detection was positively correlated with the detection of culturable *E. coli*, total coliform, and F+RNA coliphage (Table 3). However, pathogenic *E. coli* detection was not significantly correlated with the concentration of molecular *E. coli*. The likelihood of detecting pathogenic *E. coli*, and only this pathogen, declines significantly with increasing well depth. The detection of pathogenic *E. coli* was not correlated with any other environmental parameters, including platform quality and population density.

The presence of *Shigella* was significantly correlated with the concentration of molecular *E. coli*, but not with the concentration of culturable *E. coli*, total coliforms, or F+RNA coliphage (Table 3). There was a significant positive correlation between the detection of *Shigella* and ORP and negative correlation between *Shigella* detection and ammonia concentrations. *Vibrio* was the only pathogen for which the likelihood of detection increases for wells

Table 3

Spearman rank correlations between indicators and environmental parameters towards fecal pathogens. Only significant correlations at the [*] <0.05 and [**] <0.01 levels are shown.

	Variable	Pathogenic E. coli	Shigella	Vibrio	Pathogenic E. coli + Shigella	Total bacterial pathogens	Rotavirus
Tubewell feature	Platform			0.30[*]			
	Depth	− 0.32[*]			− 0.29[*]		
Water chemistry	ORP		0.37[*]		0.40[**]		
	As				− 0.30[*]		
	NH$_3$		− 0.46[**]		− 0.47[**]	− 0.41[**]	
	NO$_3$						− 0.36[*]
Indicator	mE. coli		0.32[*]		0.44[**]	0.40[**]	
	Total coliform	0.35[*]			0.37[**]	0.32[*]	
	E. coli	0.30[*]			0.28[*]		
	F+RNA	0.34[*]			0.47[**]	0.36[*]	
Pathogen	Pathogenic E. coli + Shigella						0.36[**]
	Total bacterial						0.40[**]

with a broken or missing platform. The presence of *Vibrio* was not significantly correlated with any other indicator or environmental parameter.

The combined presence of pathogenic *E. coli* and *Shigella* improved the significance for five out of the previous seven individual pathogen-indicator correlations (Table 3). There was an additional significant negative correlation between the combined detection of pathogenic *E. coli* or *Shigella* and As concentration. The presence of *Vibrio* combined with pathogenic *E. coli* and *Shigella* reduced the number of significant correlations from eight to four. The presence of bacterial pathogens and rotavirus combined was not significantly correlated with any fecal indicator or environmental parameter.

4. Discussion

4.1. Temporal and spatial context for the pathogen data

Culturable levels of *E. coli* in shallow tubewells vary both geographically (Fig. 1) and seasonally within Matlab and in another region of Bangladesh (Leber et al., 2011; van Geen et al., 2011). The annually-averaged frequency of culturable *E. coli* detection was $38 \pm 15\%$ for the 30 wells with <1 MPN/100 mL *E. coli* when sampling for pathogens, $43 \pm 13\%$ in the 14 wells within the low risk category, and $57 \pm 8\%$ in the six remaining wells in intermediate to very high risk category. During the monsoon of 2008 and 2009 the averaged frequency of culturable *E. coli* detection was $34 \pm 18\%$ for the 30 wells with <1 MPN/100 mL *E. coli* when sampling for pathogens, $53 \pm 13\%$ in the 14 wells within the low risk category, and $65 \pm 20\%$ in the six remaining wells in intermediate to very high risk category. Levels of culturable *E. coli* measured in tubewell water on the 8 days when samples were collected for pathogens were therefore broadly consistent with the monitoring of the same wells over a more extended period.

Spatial patterns of culturable *E. coli* in shallow groundwater have been related to variations in local population density (van Geen et al., 2011). The population residing within 25 m of each of the 50 wells sampled for pathogens averages 22 and ranges from 5 to 51. Linear regression for these 50 wells indicates a statistically significant increase from, on average, 30 to 60% in the frequency of culturable *E. coli* detection during the monsoon season across the full range of population density within Bara Haldia ($p < 0.05$). The relationship between shallow groundwater contamination and presumed variations in source strength related to population density at the surface helps explain why all but 3 of the wells with detectable levels of culturable *E. coli* were located in the northern, more densely populated, portion of the village (Fig. 1d). Five out of 6 wells in the intermediate to very high risk categories were located in northern Bara Haldia. Wells positive for F+RNA coliphage (Fig. 1c) and m*E. coli* (Fig. 1e) were also more prevalent in the northern half. The context provided by the culturable *E. coli*,

m*E.coli* and F+RNA coliphage data suggests that pathogen concentrations measured on a single occasion in 50 wells of Bara Haldia likely reflect recurring patterns of monsoon contamination that are linked to local but relatively stable environmental conditions.

4.2. Evidence of downward transport for some pathogens

Spatial patterns of culturable *E. coli* in shallow groundwater have been linked to the rate of local groundwater recharge, using groundwater As as a proxy for the age of shallow groundwater (Leber et al., 2011; van Geen et al., 2011). Groundwater dating using the ^{3}H–^{3}He method indicates a broad increase in As concentrations with the age of groundwater (i.e. the elapsed time since the infiltrating water was isolated from the atmosphere) and with progressively more reducing conditions in the aquifer (Stute et al., 2007). Within the subset of 50 wells sampled for pathogens within Bara Haldia, however, there was no statistically significant relationship between the As content of shallow groundwater and either the levels of culturable *E. coli* measured in parallel with pathogens or the frequency of culturable *E. coli* detection during the monsoon of 2008 and 2009. On the other hand, both the frequency of culturable *E. coli* over the entire year (Pearson correlation $r = -0.39$, $p < 0.01$), and detection of pathogenic *E. coli* and *Shigella* combined (*rho*: -0.30, $p < 0.05$) (Table 2) are negatively correlated with As concentrations in the same well water samples and therefore the local hydrogeology. Ammonia levels measured in the 50 wells provide some support for the modulation of a downward pathway of contamination by the local hydrogeology. In Bara Haldia and in other areas of Bangladesh where As contamination is an issue, concentrations of ammonia and As are often positively correlated (Harvey et al., 2002). This is because the dominant source of ammonia to groundwater is the natural degradation of organic matter, rather than sewage (Dowling et al., 2002) and the impact of this source is diluted in areas where groundwater is relatively young. For the 50 wells sampled in Bara Haldia, concentrations of ammonia are negatively correlated with the detection of *Shigella* individually and combined with pathogenic *E. coli* and *Vibrio* (Table 3). The positive correlation between ORP and the presence of *Shigella* alone, as well as pathogenic *E. coli* and *Shigella* combined, is also consistent with the notion that less reduced, more rapidly recharged aquifers are more likely to be contaminated with pathogenic bacteria. The relationships between the presence of bacterial pathogens and several environmental parameters therefore support the notion that areas of rapid recharge, e.g. sandy unconfined aquifers that are not capped with an impermeable clay layer (Aziz et al., 2008), are particularly likely to be contaminated with fecal matter.

It is presently unclear why the presence of rotavirus appears to be unrelated to any of the environmental parameters suggesting

downward transport of bacterial pathogens. The negative correlation between the presence of rotavirus and nitrate is hard to explain in terms of local recharge because nitrate levels are elevated in only the most recently recharged groundwater in reducing aquifers (Andersen et al., 2007). Although there is a statistically significant correlation between the presence of rotavirus and that of all bacterial pathogens combined, the relation to environmental parameters may be obscured by differences in resistance to environmental factors and groundwater transport among species and strains (Borchardt et al., 2003a, 2003b; Taylor et al., 2004). Rotavirus may also contaminate shallow aquifers along a pathway that differs from that followed by bacterial pathogens.

4.3. Sensitivity and selectivity of fecal indicators

Although there are some significant correlations between the detection of certain pathogens and water quality, these might be fairly site-specific and difficult to generalize. This section therefore focuses on the effectiveness of using six indicators of fecal contamination, individually or in pairs, to predict the presence of microbial pathogens. Indicator and pathogen data were converted to a binary outcome and classified to determine the maximum sensitivity and specificity by modifying a cutoff value. The cutoff value was chosen to first maximize sensitivity (minimize false negatives) and then maximize specificity (minimize false positives). Although the trade-offs between sensitivity and selectivity can be balanced differently (Harwood et al., 2005), only those indicators that capture all wells containing a pathogen, or combination of pathogens, and correctly identify ≥50% of wells without pathogens are considered here.

The analysis shows that the effective fecal indicators for predicting the presence of pathogenic E. coli, detected in 4 out of 50 wells, are molecular E. coli, total coliform, culturable E. coli, and F+RNA coliphage (Table A2a). The number is reduced from four to three indicators if molecular E. coli, which currently requires PCR analysis, is not included. For Shigella (5 wells positive), molecular E. coli, total coliform, and F+RNA coliphage are again effective indicators but culturable E. coli is not. Only molecular E. coli is an effective predictor of the presence of pathogenic E. coli or Shigella combined (9 wells) as well as the presence of any of the bacterial pathogens (12 wells) (Table A2b). F+RNA coliphage is also an effective predictor of the presence of pathogenic E. coli or Shigella combined. Using the same criteria, none of the individual fecal indicators are effective for predicting the presence of Vibrio (5 wells) or rotavirus (20 wells). Similarly, no indicator is individually effective for predicting the presence of any of the bacterial pathogens as well as rotavirus combined (23 wells).

Paired combinations of fecal indicators can improve sensitivity and specificity above that achieved using an indicator exclusively (Table A3). For pathogenic E. coli, three different combinations of paired fecal indicators increased specificity by up to 14% (total coliform and F+RNA coliphage) above that achieved using total coliform exclusively. This number is reduced to two if you discount PCR based methodologies. For Vibrio spp. (18% increase), combined pathogenic E. coli and Shigella (6% increase), and total bacterial pathogens (10% increase), one combination of paired indicators improved specificity in each case. Other than for pathogenic E. coli, all combinations of paired indicators require PCR based analysis, which in resource-limited areas is not a viable option in the context of ASSURED diagnostic testing (Urdea et al., 2006). Combinations of paired fecal indicators were not effective for predicting the presence of Shigella, rotavirus, or total pathogens at 100% sensitivity and ≥50% specificity.

4.4. Practical and health implications

The frequency (40%) and concentrations (up to 1.4×10^5 genes/100 mL) of human rotavirus measured in Bara Haldia were more than 10 fold higher than those reported for shallow groundwater in Wisconsin (0.5% (Borchardt et al., 2003a)), unspecified drinking water samples in Benin (~2% (Verheyen et al., 2009)), and drinking water samples during an epidemic in France (~2% (Gratacap-Cavallier et al., 2000)). The frequency of occurrence of rotavirus in shallow wells of Bara Haldia was similar to that for a surface water source in the Netherlands 48% (Lodder et al., 2010). Rotavirus is widely believed to be transmitted by person-to-person contact (Anderson and Weber, 2004), but our data suggest that shallow groundwater might be a source even if the risk of infection cannot be directly determined from molecular assays (Gassilloud et al., 2003). The presence of more than one rotavirus genotype in the well water samples is not surprising because rotavirus has considerable genomic diversity (Dennehy, 2008). In addition, the dominant circulating strains can change over time (Rahman et al., 2007). The dominant type found in this study was G12, first isolated on the Asian subcontinent from stool samples between 2004 and 2006 and might be increasing in frequency (Paul et al., 2008; Ramani et al., 2007). A more systematic study of potential rotavirus infection from drinking groundwater seems warranted, especially given the predominance of rotavirus-caused cases of diarrhea reported in the Matlab hospital (Tanaka et al., 2007; Zaman et al., 2009).

Bacterial pathogenicity genes were correlated with the presence of rotavirus with 75% of wells positive for one of these genes also positive for rotavirus. The mechanism for co-contamination of wells is unclear, but could result from the observations that multiple pathogens can frequently be isolated from patients with diarrhea (Albert et al., 1999) or could be indicative of shallow wells subject to fecal contamination. This suggests that surveillance for rotavirus alone may identify drinking water wells impacted by pathogen contamination. Thus rotavirus may serve as a conservative indicator of fecal contamination because rotavirus can remain for months after introduction into groundwater in the lab (Espinosa et al., 2008) and also may be preferentially transported in groundwater relative to bacteria (Taylor et al., 2004).

The main freshwater FIB (culturable E. coli) used by regulatory agencies and international health organizations to determine groundwater quality was significantly correlated and exhibited no false negatives (100% sensitivity) towards only one bacterial group (pathogenic E. coli). Contingency analysis also showed a statistically significant relationship between the annually-averaged E. coli detection frequencies for the 50 wells and the presence of total bacterial pathogens, (p<0.01) (Fig. A2a and c). The available data indicates that wells with a low (<25%) annually-averaged frequency of E. coli detection are less likely to have bacterial pathogens than wells with moderate (25–50%) and high (>50%) frequency of E. coli detection. This still provides some justification for repeated E. coli measurements as a way of screening wells for microbial contamination. However, annually-averaged E. coli detection frequency was not predictive for the presence or absence of rotavirus (Fig. A2b and d).

Total coliforms exhibited the highest sensitivity (100%) towards rotavirus and agreed with previous work on the viral contamination of groundwater (Payment and Locas, 2010) and the screening of six wastewater reclamation facilities in the US (Harwood et al., 2005). However, as total coliforms were present in 92% of tubewells the low cutoff value required meant their specificity was low, exhibited by 89% false positive rate (Table A2a). No bacterial indicator, exclusive or in paired combination, effectively distinguished tubewells positive for rotavirus, possibly because of different survival and transport characteristics between bacteria and viruses (Borchardt et al., 2003a; Taylor et al., 2004). Interestingly, the common fecal indicator virus (F+RNA coliphage) used by regulatory agencies did not significantly correlate or provide high sensitivity or specificity towards rotavirus (74% and 36%, respectively). Lodder et al. (2010) also discounted the use of somatic and F-specific coliphage as indicators of source water quality, highlighting that currently the most effective approach of identifying pathogenic viral contamination is through direct detection. The need for assured

diagnostic testing (Urdea et al., 2006) indicates that alternate phage, coliphage host, and indicators for enteric viruses should be investigated (Harwood et al., 2005).

Based on our study, groundwater cannot be excluded as a significant source of diarrheal disease in Bangladesh and neighboring countries with similar characteristics. The most promising individual indicators for detecting bacterial pathogens in groundwater are F+RNA coliphage and molecular *E. coli*, but both require more effort and expertise than *E. coli* detection by culturing. F+RNA coliphage is not an effective indicator of rotavirus, which is present in groundwater pumped from a particularly high proportion of wells.

Acknowledgments

Appendix A. Supplementary data

The complete dataset can be found online at http://dx.doi.org/10.1594/PANGAEA.774444.

References

Ahmed MF, Ahuja S, Alauddin M, Hug SJ, Lloyd JR, Pfaff A, et al. Ensuring safe drinking water in Bangladesh. Science 2006;314:1687–8.

Alam M, Sultana M, Nair GB, Sack RB, Sack DA, Siddique AK, et al. Toxigenic *Vibrio cholerae* in the aquatic environment of Mathbaria, Bangladesh. Appl Environ Microbiol 2006;72:2849–55.

Albert MJ, Faruque ASG, Faruque SM, Sack RB, Mahalanabis D. Case–control study of enteropathogens associated with childhood diarrhea in Dhaka, Bangladesh. J Clin Microbiol 1999;37:3458–64.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403–10.

Andersen MS, Baron L, Gudbjerg J, Gregersen J, Chapellier D, Jakobsen R, et al. Discharge of nitrate-containing groundwater into a coastal marine environment. J Hydrol 2007;336:98-114.

Anderson EJ, Weber SG. Rotavirus infection in adults. Lancet Infect Dis 2004;4:91–9.

Aziz Z, van Geen A, Stute M, Versteeg R, Horneman A, Zheng Y, et al. Impact of local recharge on arsenic concentrations in shallow aquifers inferred from the electromagnetic conductivity of soils in Araihazar, Bangladesh. Water Resour Res 2008;44.

Begum Y, Talukder K, Nair G, Khan S, Svennerholm A, Sack R, et al. Comparison of enterotoxigenic *Escherichia coli* isolated from surface water and diarrhoeal stool samples in Bangladesh. Can J Microbiol 2007;53:19–26.

Boehm AB, Shellenbarger GG, Paytan A. Groundwater discharge: potential association with fecal indicator bacteria in the surf zone. Environ Sci Technol 2004;38: 3558–66.

Borchardt MA, Bertz PD, Spencer SK, Battigelli DA. Incidence of enteric viruses in groundwater from household wells in Wisconsin. Appl Environ Microbiol 2003a;69:1172–80.

Borchardt MA, Bertz PD, Spencer SK, Battigelli DA. Incidence of Enteric Viruses in Groundwater from Household Wells in Wisconsin, 69. ; 2003b. p. 1172–80.

Brennan FP, O'Flaherty V, Kramers G, Grant J, Richards KG. Long-term persistence and leaching of *Escherichia coli* in temperate maritime soils. Appl Environ Microbiol 2010;76:1449–55.

Bushon RN. Fecal indicator viruses: U.S. geological survey techniques of water-resources investigations; 2003. book 9, chap. A7, section 7.2, November.

Cheng Z, Zheng Y, Mortlock R, van Geen A. Rapid multi-element analysis of groundwater by high-resolution inductively coupled plasma mass spectrometry. Anal Bioanal Chem 2004;379:513–8.

Debartolomeis J, Cabelli VJ. Evaluation of an *Escherichia coli* host strain for enumeration of F male-specific bacteriophages. Appl Environ Microbiol 1991;57:1301–5.

Dennehy P. Rotavirus vaccines: an overview. Clin Microbiol Rev 2008;21:198–208.

Dowling CB, Poreda RJ, Basu AR, Peters SL, Aggarwal PK. Geochemical study of arsenic release mechanisms in the Bengal Basin groundwater. Water Resour Res 2002;38.

Espinosa AC, Mazari-Hiriart M, Espinosa R, Maruri-Avidal L, Mendez E, Arias CF. Infectivity and genome persistence of rotavirus and astrovirus in groundwater and surface water. Water Res 2008;42:2618–28.

Faruque SM, Khan R, Kamruzzaman M, Yamasaki S, Ahmad QS, Azim T, et al. Isolation of *Shigella dysenteriae* type 1 and *S. flexneri* strains from surface waters in Bangladesh: comparative molecular analysis of environmental *Shigella* isolates versus clinical strains. Appl Environ Microbiol 2002;68:3908–13.

Ferguson AS, Mailloux BJ, Ahmed KM, van Geen A, McKay LD, Culligan PJ. Hand-pumps as reservoirs for microbial contamination of well water. J Water Health 2011;9(4): 708–17.

Gassilloud B, Schwartzbrod L, Gantzer C. Presence of viral genomes in mineral water: a sufficient condition to assume infectious risk? Appl Environ Microbiol 2003;69: 3965–9.

Gibson KE, Schwab KJ. Detection of bacterial indicators and human and bovine enteric viruses in surface water and groundwater sources potentially impacted by animal and human wastes in Lower Yakima Valley, Washington. Appl Environ Microbiol 2011;77:355–62.

Glassmeyer ST, Furlong ET, Kolpin DW, Cahill JD, Zaugg SD, Werner SL, et al. Transport of chemical and microbial compounds from known wastewater discharges: potential for use as indicators of human fecal contamination. Environ Sci Technol 2005;39:5157–69.

Gratacap-Cavallier B, Genoulaz O, Brengel-Pesce K, Soule H, Innocenti-Francillard P, Bost M, et al. Detection of human and animal rotavirus sequences in drinking water. Appl Environ Microbiol 2000;66:2690–2.

Harvey CF, Swartz CH, Badruzzaman AB, Keon-Blute N, Yu W, Ali MA, et al. Arsenic mobility and groundwater extraction in Bangladesh. Science 2002;298: 1602–6.

Harwood VJ, Levine AD, Scott TM, Chivukula V, Lukasik J, Farrah SR, et al. Validity of the indicator organism paradigm for pathogen reduction in reclaimed water and public health protection. Appl Environ Microbiol 2005;71:3163–70.

Hunt RJ, Borchardt MA, Richards KD, Spencer SK. Assessment of sewer source contamination of drinking water wells using tracers and human enteric viruses. Environ Sci Technol 2010;44:7956–63.

Hurley MA, Roscoe ME. Automated statistical-analysis of microbial enumeration by dilution series. J Appl Bacteriol 1983;55:159–64.

Knappett PSK, Layton A, McKay LD, Williams D, Mailloux BJ, Huq MR, et al. Efficacy of hollow-fiber ultrafiltration for microbial sampling in groundwater. Ground Water 2011;49:53–65.

Kosek M, Bern C, Guerrant RL. The global burden of diarrhoeal disease, as estimated from studies published between 1992 and 2000. Bull World Health Organ 2003;81:197–204.

Layton A, McKay L, Williams D, Garrett V, Gentry R, Sayler G. Development of *Bacteroides* 16S rRNA gene TaqMan-based real-time PCR assays for estimation of total, human, and bovine fecal pollution in water. Appl Environ Microbiol 2006;72: 4214–24.

Leber J, Rahman MM, Ahmed KM, Mailloux B, van Geen A. Contrasting Influence of geology on *E-coli* and arsenic in aquifers of Bangladesh. Ground Water 2011;49:111–23.

Lodder WJ, van den Berg H, Rutjes SA, Husman AMD. Presence of enteric viruses in source waters for drinking water production in the Netherlands. Appl Environ Microbiol 2010;76:5965–71.

Luby SP, Gupta SK, Sheikh MA, Johnston RB, Ram PK, Islam MS. Tubewell water quality and predictors of contamination in three flood-prone areas in Bangladesh. J Appl Microbiol 2008;105:1002–8.

Momba MN, Malakate VK, Theron J. Abundance of pathogenic *Escherichia coli*, *Salmonella typhimurium* and *Vibrio cholerae* in Nkonkobe drinking water sources. J Water Health 2006;4:289–96.

Paul S, Kobayashi N, Nagashima S, Ishino M, Watanabe S, Alam M, et al. Phylogenetic analysis of rotaviruses with genotypes G1, G2, G9 and G12 in Bangladesh: evidence for a close relationship between rotaviruses from children and adults. Arch Virol 2008;153:1999–2012.

Payment P, Locas A. Pathogens in water: value and limits of correlation with microbial indicators. Ground Water 2010;49:4-11.

Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. Am J Epidemiol 2004;159:882–90.

Pote J, Haller L, Kottelat R, Sastre V, Arpagaus P, Wildi W. Persistence and growth of faecal culturable bacterial indicators in water column and sediments of Vidy Bay, Lake Geneva, Switzerland. J Environ Sci (China) 2009;21:62–9.

Rahman M, Sultana R, Ahmed G, Nahar S, Hassan Z, Saiada F, et al. Prevalence of G2P[4] and G12P[6] rotavirus, Bangladesh. Emerg Infect Dis 2007;13:18–24.

Ramani S, Banerjee I, Primrose Gladstone B, Sarkar R, Selvapandian D, Le Fevre AM, et al. Geographic information systems and genotyping in identification of rotavirus G12 infections in residents of an urban slum with subsequent detection in hospitalized children: emergence of G12 genotype in South India. J Clin Microbiol 2007;45:432–7.

Schriewer A, Miller WA, Byrne BA, Miller MA, Oates S, Conrad PA, et al. Presence of Bacteroidales as a predictor of pathogens in surface waters of the central California coast. Appl Environ Microbiol 2010;76:5802–14.

Sinton LW, Finlay RK, Reid AJ. A simple membrane filtration-elution method for the enumeration of F-RNA, F-DNA and somatic coliphages in 100-ml water samples. J Microbiol Methods 1996;25:257–69.

Stute M, Zheng Y, Schlosser P, Horneman A, Dhar RK, Datta S, et al. Hydrological control of As concentrations in Bangladesh groundwater. Water Resour Res 2007;43. (-).

Tanaka G, Faruque ASG, Luby SP, Malek MA, Glass RI, Parashar UD. Deaths from rotavirus disease in Bangladeshi children — estimates from hospital-based surveillance. Pediatr Infect Dis J 2007;26:1014–8.

Taylor R, Cronin A, Pedley S, Barker J, Atkinson T. The implications of groundwater velocity variations on microbial transport and wellhead protection — review of field evidence. FEMS Microbiol Ecol 2004;49:17–26.

Urdea M, Penny LA, Olmsted SS, Giovanni MY, Kaspar P, Shepherd A, et al. Requirements for high impact diagnostics in the developing world. Nature 2006;444(Suppl. 1):73–9.

van Geen A, Ahmed KM, Akita Y, Alam MJ, Culligan PJ, Emch M, et al. Fecal contamination of shallow tubewells in Bangladesh inversely related to arsenic. Environ Sci Technol 2011;45:1199–205.

Verheyen J, Timmen-Wego M, Laudien R, Boussaad I, Sen S, Koc A, et al. Detection of adenoviruses and rotaviruses in drinking water sources used in rural areas of Benin, West Africa. Appl Environ Microbiol 2009;75:2798–801.

Wampler PJ, Sisson AJ. Spring flow, bacterial contamination, and water resources in Rural Haiti. Environ Earth Sci 2011;62(8):1619–28.

WHO. Guidelines for drinking-water quality, incorporating first and second addenda to third edition. Recommendations, vol. 1; 2008. Geneva.

Zaman K, Yunus M, Faruque ASG, El Arifeen S, Hossain I, Azim T, et al. Surveillance of rotavirus in a rural diarrhoea treatment centre in Bangladesh, 2000–2006. Vaccine 2009;27:F31–4.

Supplementary Data (Appendix A) for:

Comparison of fecal indicators with pathogenic bacteria and rotavirus in groundwater

Andrew S. Ferguson[1]*, Alice C. Layton[2], Brian J, Mailloux[3], Patricia J. Culligan[1], Daniel E. Williams[2], Abby E. Smartt[2], Gary S. Sayler[2], John Feighery[4], Larry D. McKay[5], Peter S.K. Knappett[5], Ekaterina Alexandrova[3], Talia Arbit[3], Michael Emch[6], Veronica Escamilla[6], Kazi Matin Ahmed[7], Md. Jahangir Alam[7], P. Kim Streatfield[8], Mohammad Yunus[8], Alexander van Geen[9]

[1]Department of Civil Engineering and Engineering Mechanics, Columbia University, NY, USA

[2]Department of Microbiology and Center for Environmental Biotechnology, University of Tennessee, Knoxville, Tennessee, USA.

[3]Department of Environmental Sciences, Barnard College, New York, USA.

[4]Department of Earth and Environmental Engineering, Columbia University, New York, USA.

[5]Department of Earth and Planetary Sciences, University of Tennessee, Knoxville, TN, USA.

[6]Department of Geography, University of North Carolina at Chapel Hill, North Carolina, USA.

[7]Department of Geology, University of Dhaka, Dhaka, Bangladesh.

[8]International Center for Diarrhoeal Research, Bangladesh, Dhaka, Bangladesh.

[9]Lamont-Doherty Earth Observatory of Columbia University, Palisades, NY 10964, USA.

*Corresponding Author: Email: asf747@gmail.com

24   Table A1. Quantitative PCR primer and probes used to target specific genes and organisms in

25   tubewell water samples.

26

Target organism (Relevance)	Assay Gene Target (Assay type and annealing temperature)	Oligonucleotide Sequences	Reference
Shigella and EIEC *E. coli* (Dysentery type *E. coli* and *Shigella*)	*ipa*H (Fluorogenic Probe, 60°C)	*Ipa*H U1f- 5' CCTTTTCCGCGTTCCTTG A-3' *Ipa*H L1r- 5'- CGGAATCCGGAGGTATTG C-3' *Ipa*H Taq- 5'-CGCCTTTCCGATACCGTCTCTGCA-3'	(von Seidlein et al., 2006)
ETEC *E. coli* (Enterotoxigenic *E. coli* strains)	*elt*A (Heat labile toxin LT) (Fluorogenic Probe, 60°C)	*Elt*311f- 5' TCTGAATATAGCTCCGGCAGA-3' *Elt*414r-5' CAACCTTGTGGTGCATGATGA-3' *Elt*383Taqr -5' TTCTCTCCAAGCTTGGTGATCCGGT-3'	(Persson et al., 2007)Modified
Shigella and EIEC *E. coli*, EPEC (*Shigella* and multiple types of *E. coli* pathogens)	*stx*1 or *vtx*1 (shiga toxin 1) (Fluorogenic Probe, 60°C)	Ib*Stx*1m64f- 5'-GCAAAGACGTATGTAGATTCGCT-3' Ib*Stx*1m209r 5'-ATCTATCCCTCTGACATCAACTG-3' Ib*Stx*1mTaq119f 5-ATGTCATTCGCTCTGCAATAGGTACTC-3'	(Persson et al., 2007) (Ibekwe et al., 2002) Modified
EPEC,VTEC, A/EEC *E. coli* (*Shigella* and multiple types *of E. coli* pathogens)	*eae* (Fluorogenic Probe, 60°C)	PS*eae*1001f- 5'-GGCGAATACTGGCGAGACTA-3' PS*eae*1123r -5' –ACGGATATCGAAGCCATTTG-3' PS*eae*taq1035f-5'-GTTAACGGCTATTTCCGCATGAGCG-3'	(Persson et al., 2007) (Ibekwe et al., 2002) Modified
Vibrio	*ompW* (SYBR green 60°C)	OmpW f-5'-CACCAAGAAGGTGACTTTATTGTG-3' OmpWr-5'-GGTTTGTCGAATTAGCTTCACC-3'	(Nandi et al., 2000)
Rotavirus (RNA)	VP6 (Reverse transcriptase SYBR green 60°C) G12VP6 (Reverse transcriptase Fluorogenic Probe 61°C)	RotavirusA.ext.fwd 5'-CTTTAAAACGAAGTCTTCRACATG-3' RotavirusA.ext.rev 5'-TARATTACCAATTCCTCCAGTTTG- 3' G12RV23f -5' -GGAGGTTCTGTATTCATTGTCA3' G12RV176r - 5' CC AATTCCTCCAGTTTGAAAG 3' G12RV55fTaq 5'-AAAGATGCTAGGGATAAGATTGTTGAAGGTAC-3'	Modified (Logan et al., 2006) This Study

Table A2. Sensitivity and specificity for (a) individual fecal pathogens and (b) grouped fecal pathogens against microbial indicators and environmental characteristics. For calculations and determination of cutoff thresholds all parameters were ranked low to high except those labeled [a] which were ranked high to low. Cutoff values are optimized to achieve maximum sensitivity and specificity >0%. A sensitivity of 100% and a specificity of ≥50% are highlighted in grey.

A2a	Variable	Range (min/max)	Pathogenic *E. coli*			*Shigella*			*Vibrio*			Rotavirus		
			Cutoff	Sens	Spec	Cutoff	Sens	Spec	Cutoff	Sens	Spec	Cutoff	Sens	Spec
Tubewell feature	Priming	0/0.38	0.03	0.75	0.48	0.03	0.60	0.47	0.06	0.40	0.73	0.03	0.65	0.53
	depth[a]	25/120	51.50	1.00	0.52	54.00	0.80	0.49	62.50	1.00	0.19	82.50	1.00	0.14
	pop25	5/51	12.50	1.00	0.24	16.50	1.00	0.51	7.00	1.00	0.10	7.00	0.94	0.11
	pop50	6/130	39.00	1.00	0.41	24.50	1.00	0.14	24.50	1.00	0.14	11.50	1.00	0.10
Water chemistry	Temp	25.83/27.33	26.06	1.00	0.18	26.08	1.00	0.21	26.01	1.00	0.12	25.89	1.00	0.07
	pH	6.02/6.88	6.24	1.00	0.26	6.06	1.00	0.3	6.06	1.00	0.03	6.06	1.00	0.04
	EC	0.34/3.93	0.46	0.50	0.02	0.97	1.00	0.47	0.84	0.80	0.37	0.46	0.89	0.03
	ORP	-207.7/199.6	126.75	1.00	0.52	138.60	1.00	0.67	87.80	0.60	0.30	-54.85	0.95	0.07
	Sal	0.16/2.06	0.22	0.50	0.02	0.48	1.00	0.44	0.41	0.80	0.33	0.22	0.89	0.03
	As[a]	0.62/435.87	228.59	1.00	0.11	36.01	1.00	0.56	85.85	1.00	0.37	428.93	1.00	0.03
	NH_3[a]	0/10	4.25	1.00	0.23	1.25	1.00	0.72	3.50	1.00	0.33	9.00	0.95	0.07
	NO_2	0/18.41	2.99	0.50	0.78	4.25	1.00	0.84	3.26	1.00	0.83	0.34	0.70	0.36
	NO_3	0/25.75	0.54	0.50	0.67	0.30	1.00	0.29	0.27	1.00	0.23	0.12	0.85	0.14
	PO_4	0/2.86	0.32	1.00	0.50	0.32	1.00	0.49	0.31	1.00	0.45	0.31	0.77	0.44
	SO_4	0/126.67	1.42	0.50	0.74	0.03	1.00	0.46	0.13	0.50	0.47	0.03	0.50	0.39
	Br	0/4.72	0.45	1.00	0.38	0.51	1.00	0.47	0.40	1.00	0.36	0.33	0.85	0.32
	Cl	4.44/933.85	47.30	1.00	0.41	67.26	1.00	0.45	45.81	1.00	0.38	5.48	1.00	0.04
	ClBr	54.98/769.09	217.69	1.00	0.30	327.66	1.00	0.44	234.32	1.00	0.33	101.20	1.00	0.17
Indicators	m*E. coli*	0.6/5.18	2.25	1.00	0.65	2.31	1.00	0.69	1.84	1.00	0.45	0.63	0.95	0.21
	Bacteroides	0.6/7.45	2.27	1.00	0.23	2.17	0.80	0.17	1.36	1.00	0.17	1.36	0.90	0.18
	Hu*Bacteroides*	0.6/6.82	0.50	1.00	0.02	0.83	0.6	0.79	0.50	1.00	0.02	0.71	0.37	0.79
	Total coliform	-0.60/3.48	1.63	1.00	0.74	0.81	1.00	0.43	0.81	0.80	0.38	-0.45	1.00	0.11
	E. coli	-0.60/3.48	0.45	1.00	0.70	-0.30	0.60	0.52	-0.30	0.40	0.50	-0.45	0.58	0.57
	F+RNA	-1.60/2.16	0.17	1.00	0.60	0.57	1.00	0.69	0.17	0.60	0.57	-1.27	0.74	0.36

S3

A2b	Variable	Range (min/max)	Pathogenic *E. coli* +*Shigella*			Total bacterial pathogens			Total pathogens		
			Cutoff	Sens	Spec	Cutoff	Sens	Spec	Cutoff	Sens	Spec
Tubewell feature	Priming	0/0.38	0.03	0.67	0.49	0.03	0.67	0.50	0.03	0.65	0.56
	depth[a]	25/120	54.00	0.89	0.54	62.50	0.92	0.19	82.50	0.96	0.12
	pop25	5/51	12.50	1.00	0.27	7.00	1.00	0.12	7.00	0.95	0.12
	pop50	6/130	24.50	1.00	0.15	24.50	1.00	0.17	11.50	1.00	0.11
Water chemistry	Temp	25.83/27.33	26.06	1.00	0.21	26.01	1.00	0.14	25.89	1.00	0.08
	pH	6.02/6.88	6.06	1.00	0.03	6.06	1.00	0.03	6.06	1.00	0.05
	EC	0.34/3.93	0.46	0.78	0.03	0.46	0.83	0.03	0.46	0.91	0.04
	ORP	-207.7/199.6	126.75	1.00	0.59	-54.85	0.83	0.03	-54.85	0.91	0.04
	Sal	0.16/2.06	0.22	0.78	0.03	0.22	0.83	0.3	0.22	0.91	0.04
	As[a]	0.62/435.87	228.59	1.00	0.13	228.59	1.00	0.14	428.93	1.00	0.04
	NH$_3$[a]	0/10	4.25	1.00	0.28	4.25	1.00	0.28	9.00	0.95	0.08
	NO$_2$	0/18.41	2.99	0.67	0.80	2.99	0.80	0.85	0.34	0.77	0.40
	NO$_3$	0/25.75	0.30	0.80	0.28	0.27	0.86	0.24	0.26	0.88	0.20
	PO$_4$	0/2.86	0.32	1.00	0.51	0.31	1.00	0.52	0.31	0.81	0.50
	SO$_4$	0/126.67	0.03	0.80	0.46	0.03	0.71	0.45	0.03	0.53	0.40
	Br	0/4.72	0.45	1.00	0.42	0.40	1.00	0.41	0.33	0.88	0.36
	Cl	4.44/933.85	47.30	1.00	0.44	45.81	1.00	0.44	5.48	1.00	0.04
	ClBr	54.98/769.09	217.69	1.00	0.33	217.69	1.00	0.36	101.20	1.00	0.19
Indicators	m*E. coli*	0.6/5.18	2.25	1.00	0.74	1.84	1.00	0.54	0.63	0.96	0.24
	Bacteroides	0.6/7.45	2.17	0.89	0.18	1.36	0.92	0.17	1.36	0.86	0.16
	Hu*Bacteroides*	0.6/6.82	0.83	0.56	0.82	0.83	0.42	0.80	0.71	0.41	0.84
	Total coliform	-0.60/3.48	0.81	1.00	0.47	0.82	0.92	0.49	0.41	0.91	0.32
	E. coli	-0.60/3.48	-0.3	0.78	0.58	-0.30	0.67	0.57	-0.45	0.64	0.60
	F+RNA	-1.60/2.16	0.18	1.00	0.68	0.18	0.83	0.69	-1.27	0.73	0.36

Table A3. Sensitivity and specificity for paired fecal indicators against individual and grouped fecal pathogens. Only combinations which improved specificity above that for individual fecal indicators are shown along with their % increase.

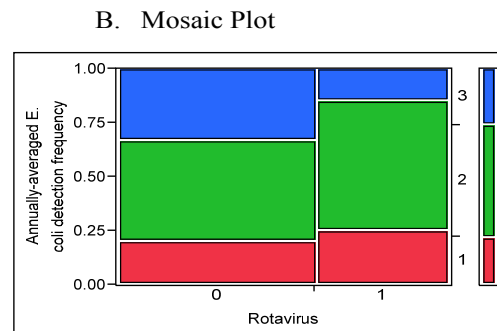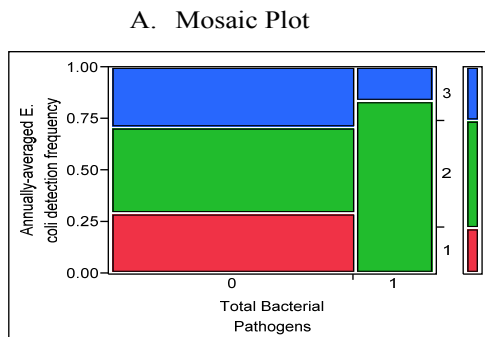| Pathogen	Variable	Sens	Spec	% increase
Pathogenic *E. coli*	log Total coliform and F+RNA coliphage	1	0.84	13.51
	log Total coliform and m*E. coli*	1	0.80	8.11
	log Total coliform and *E. coli*	1	0.78	5.41
Vibrio	log m*E. coli* and log *Bacteroides*	1	0.58	18.16
Pathogenic *E. coli* + *Shigella*	log *Bacteroides* and log F+RNA coliphage	1	0.78	5.54
Total bacterial pathogens	log m*E. coli* and log *Bacteroides*	1	0.59	10.00
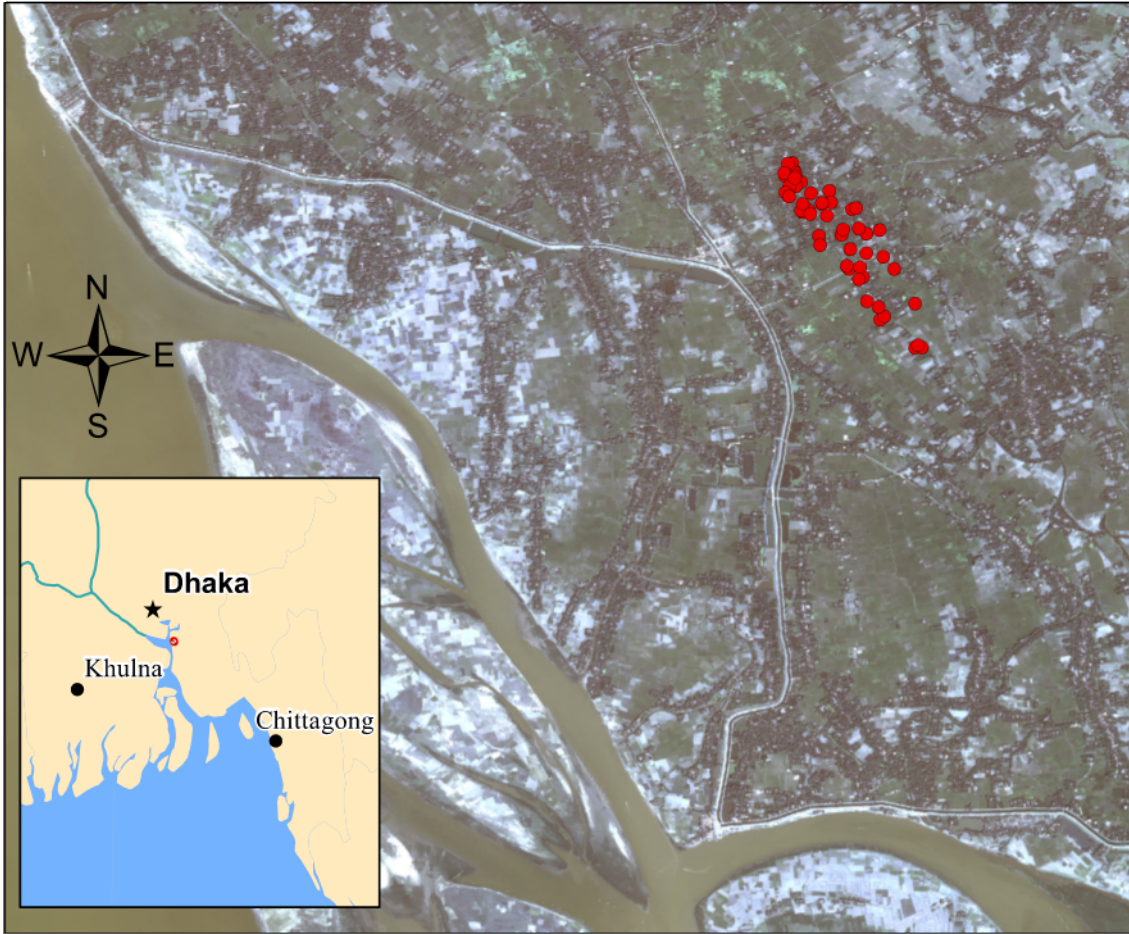
Fig. A1. Country map showing the location of the study site Bara Haldia, and close-up region showing the tubewell sample locations and Meghna River.

A. Mosaic Plot

B. Mosaic Plot

C. Contingency Table

Count Total % Col % Row %	1	2	3	
0	11 22.00 100.00 28.95	16 32.00 61.54 42.11	11 22.00 84.62 28.95	38 76.00
1	0 0.00 0.00 0.00	10 20.00 38.46 83.33	2 4.00 15.38 16.67	12 24.00
	11 22.00	26 52.00	13 26.00	50

Tests

N	DF	-LogLike	RSquare (U)
50	2	4.6495576	0.0909

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	9.299	0.0096*
Pearson	6.984	0.0304*

Warning: 20% of cells have expected count less than 5, ChiSquare suspect.

D. Contingency Table

Count Total % Col % Row %	1	2	3	
0	6 12.00 54.55 20.00	14 28.00 53.85 46.67	10 20.00 76.92 33.33	30 60.00
1	5 10.00 45.45 25.00	12 24.00 46.15 60.00	3 6.00 23.08 15.00	20 40.00
	11 22.00	26 52.00	13 26.00	50

Tests

N	DF	-LogLike	RSquare (U)
50	2	1.1040003	0.0216

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	2.208	0.3315
Pearson	2.098	0.3503

Fig. A2. Contingency analysis of annually-averaged *E. coli* detection frequency by total bacterial pathogens and rotavirus. The categories for annually averaged *E. coli* detection frequency were 1= <25%, 2= 25-50% and 3 = >50%. The categories for total bacterial pathogens and rotavirus were 0 = absent and 1 = present. A) Mosaic Plot of annually-averaged *E. coli* detection frequency by total bacterial pathogens, B) Mosaic Plot of annually-averaged *E. coli* detection frequency by rotavirus C) Contingency Table of annually-averaged *E. coli* detection frequency by total bacterial pathogens, D) Contingency Table of annually-averaged *E. coli* detection frequency by rotavirus (analysis done using JMP 8.0, SAS institute, Cary, NC).

REFERENCES

Ibekwe AM, Watt PM, Grieve CM, Sharma VK, Lyons SR. Multiplex Fluorogenic Real-Time PCR for Detection and Quantification of *Escherichia coli* O157:H7 in Dairy Wastewater Wetlands. Appl. Environ. Microbiol. 2002; 68: 4853-4862.

Logan C, O'Leary JJ, O'Sullivan N. Real-Time Reverse Transcription-PCR for Detection of Rotavirus and Adenovirus as Causative Agents of Acute Viral Gastroenteritis in Children. J. Clin. Microbiol. 2006; 44: 3189-3195.

Nandi B, Nandy RK, Mukhopadhyay S, Nair GB, Shimada T, Ghose AC. Rapid Method for Species-Specific Identification of Vibrio cholerae Using Primers Targeted to the Gene of Outer Membrane Protein OmpW. J. Clin. Microbiol. 2000; 38: 4145-4151.

Persson S, Olsen KEP, Scheutz F, Krogfelt KA, Gerner-Smidt P. A method for fast and simple detection of major diarrhoeagenic *Escherichia coli* in the routine diagnostic laboratory. Clinical Microbiology and Infection 2007; 13: 516-524.

von Seidlein L, Kim DR, Ali M, Lee H, Wang X, Thiem VD, et al. A Multicentre Study of *Shigella* Diarrhoea in Six Asian Countries: Disease Burden, Clinical Manifestations, and Microbiology. PLoS Med 2006; 3: e353.