Etiology of Diarrhea in Bangladeshi Infants in the First Year of Life Analyzed Using Molecular Methods

Mami Taniuchi,1 Shihab U. Sobuz,2 Sharmin Begum,3 James A. Platts-Mills,1 Jie Liu,1 Zhengyu Yang,2 Xin-Qun Wang,2 William A. Petri Jr,1 Rashidul Haque,3 and Eric R. Houpt1

1Division of Infectious Diseases and International Health, Department of Medicine, and 2Department of Public Health Sciences, University of Virginia, Charlottesville; 3Parasitology Laboratory, International Centre for Diarrhoeal Disease Research, Dhaka, Bangladesh

Background. Diarrhea causes enormous morbidity and mortality in developing countries, yet the relative importance of multiple potential enteropathogens has been difficult to ascertain.

Methods. We performed a longitudinal cohort study from birth to 1 year of age in 147 infants in Dhaka, Bangladesh. Using multiplex polymerase chain reaction, we analyzed 420 episodes of diarrhea and 1385 monthly surveillance stool specimens for 32 enteropathogen gene targets. For each infant we examined enteropathogen quantities over time to ascribe each positive target as a probable or less-likely contributor to diarrhea.

Results. Multiple enteropathogens were detected by the first month of life. Diarrhea was associated with a state of overall pathogen excess (mean number of enteropathogen gene targets (±SE), 5.6 ± 0.1 vs 4.3 ± 0.1 in surveillance stool specimens; \( P < .05 \)). After a longitudinal, quantitative approach was applied to filter out less-likely contributors, each diarrheal episode still had an average of 3.3 probable or dominant targets. Enteroaggregative \( E. coli \), \( Campylobacter \), enteropathogenic \( E. coli \), rotavirus, and \( Entamoeba histolytica \) were the most frequent probable contributors to diarrhea. Rotavirus was enriched in moderate to severe diarrheal episodes.

Conclusions. In this community-based study diarrhea seemed to be a multipathogen event and a state of enteropathogen excess above a high carriage baseline.

Keywords. Diarrhea; PCR; rotavirus; enteroaggregative \( e. coli \); \( Campylobacter \).

Diarrhea accounts for 26.1% of childhood deaths in South Asia [1], with a peak incidence in the first year of life [2–4]. Beyond this immediate mortality burden, diarrheal episodes contribute to intestinal barrier dysfunction and malnutrition, which underlie additional mortality [5] and disability-adjusted life-years lost [6]. This large burden of disease continues despite improvements from measures such as oral rehydration solution, antibiotics, cleaner water, sanitation, breast-feeding, and rotavirus vaccination [7–10].

The etiology of diarrhea must be understood to accelerate additional preventive measures. Unfortunately, diarrhea is a nonspecific syndrome defined as ≥3 loose stools in a day and can be caused by a diversity of viruses, bacteria, protozoa, helminths, fungi, as well as non-infectious triggers [11]. Rotavirus is widely accepted as the major diarrheal pathogen in the first year of life [12–14], but the relative importance of enteropathogens thereafter is less clear. Several methods are needed to detect these enteropathogens including culture, immunoassay, microscopy, and polymerase chain reaction (PCR), yet these are generally applied selectively and vary in their sensitivity [9, 10, 15, 16]. Certain bacteria, such as \( Campylobacter \) and \( Shigella \), are difficult to grow, particularly in the global context of widespread antibiotic use. Mixed infections are common but difficult to interpret.
METHODS

Study Design
The study was conducted from January 2008 to August 2009 in the Mirpur region of Dhaka. Details of the birth cohort have been described elsewhere [22], and this work focused on molecular testing of stool samples for enteropathogens. Briefly, 147 infants (77 male and 70 female) from the Mirpur neighborhood were enrolled in the first week after birth and followed up thereafter 2 times per week via home visits by field research assistants, with no attrition during the first year of life. Diarrhea was defined as ≥3 unformed or abnormal stools within a 24-hour period. Diarrhea specimens were collected from the home or in the study field clinic. We required that all stool samples be delivered from field to clinic to the laboratory within 6 hours of collection while maintaining cold chain. From 689 diarrheal episodes recorded by the surveillance questionnaire, 420 specimens were obtained that met this collection window (from 145 infants). Stool samples was not collectable for the remaining 269 episodes because they were of short duration and fell outside the biweekly visits (mean duration ±SD for these episodes vs the 420 episodes with collected specimens, 3.2 ± 2.5 vs 5.5 ± 4.0 days; mean age of infant, 147 ± 106 vs 182 ± 104 days of life; P < .05). All 420 collected diarrheal specimens were tested with all assays, and only 1 specimen was collected per episode. We considered diarrheal episodes independent if separated from another episode by 3 diarrhea-free days. The severity of diarrhea was determined using a modified Ruuska-Vesikari score [22, 23].

Monthly surveillance stool specimens were collected from all infants at home, and we required these specimens to be collected ≥7 days before or after a diarrheal episode.

To obtain comparators to Bangladesh, we collected stool specimens from 0–1-year-olds in Virginia, including 18 specimens obtained from infants in a daycare facility or well-baby outpatient clinic in the University of Virginia Pediatric Department, as well as 13 diarrheal stool specimens from the Virginia State Laboratory collected during routine outbreak investigations. Informed consent was obtained from all participants’ parent or guardian. This study was approved by the institutional ethics committees at the University of Virginia and the International Centre for Diarrhoeal Disease Research, Bangladesh.

Molecular Diagnostics
Stool specimens were stored at −80°C until testing. DNA extraction was performed using the QIAamp DNA Stool Mini Kit (Qiagen), following a modified protocol described elsewhere, including bead beating or a freeze-thaw technique to lyse organisms [17, 24]. DNA was stored at −20°C until use. RNA was extracted using the QuickGene RNA Tissue Kit II on the Fujifilm QuickGene-810 system (Fujifilm) [20] and stored at −80°C. Nucleic acid was amplified with sequence-specific primers via a series of panels for viruses, bacteria, protozoa, helminths, and fungi, described elsewhere [17–21]. For this work we added Trichuris trichiura (18S ribosomal RNA gene) [25] and STp, a heat-stable enterotoxin gene to detect enterotoxigenic Escherichia coli (ETEC) as these have been reported to cause diarrhea in this community [26]. The protocol entailed several multiplex PCR reactions for 32 targets to interrogate 29 organisms. Either eae alone or eae plus bfpA was considered indicative of enteropathogenic E. coli (EPEC), either stx1 or stx2 indicative of Shiga toxin-producing E. coli (STEC), either aatA or aacC indicative of enteropathogenic E. coli (EAE) [27], and either LT, STh, or STp indicative of ETEC, and ipaH interrogated both enteroinvasive E. coli (EIEC) Shigella spp. Amplicons from the PCR reactions were detected on the Bioplex 200 (Bio-Rad) with microspheres coupled with specific probes. Luminex data were reported as median fluorescent intensity corrected for background bead fluorescent (cMFI) (cMFI = [MFItarget − MFIbackground]/MFIbackground). Positive controls (DNA template from reference organisms or clinical samples) and negative controls (nuclease-free water) were included in every run. For a run to be valid, the positive and negative controls had to yield signal above and below, respectively, the cMFI cutoffs described elsewhere [17–21]. All Entamoeba histolytica positive samples underwent a secondary real-time duplex E. histolytica and Entamoeba dispar PCR assay to ensure no cross-reaction with E. dispar [28]. Giardia had been previously tested with real-time PCR, and those results were used [24]. Samples were tested in both Bangladesh and Virginia. To assure cross-comparability, the Luminex panels underwent validation at both laboratories using analytic specimens. Accuracy results across 19 targets evaluated showed a mean (±SD) sensitivity and specificity values of 90.9 ± 11.9 and 97.8 ± 3.1, respectively, for the University of Virginia laboratory and 89.8 ± 17.8 and 97.3 ± 3.6 for the Bangladeshi laboratory (difference not significant).

Quantitative Approach
To allow comparisons of enteropathogen quantities between targets (because cMFI ranges are target specific), we normalized all cMFI values for diarrheal stool specimens to the background carriage levels detected in surveillance stool specimens. Conservatively, we divided the range of cMFI values from positive surveillance tool specimens into quartiles and then assigned each positive diarrheal specimen the appropriate quartile ranking of...
1–4, or 5 if the cMFI value was higher than the highest value detected in surveillance (>100% percentile). Next, we categorized a positive target as a probable contributor to an episode of diarrhea if it was a first detection or detected at a higher quartile than any prior surveillance stool specimen from the infant. Likewise, we categorized a positive target as a less-likely contributor if it was detected at a similar or lower level than a prior surveillance stool specimen from the infant. Within all probable contributors, those at the highest quartile were considered dominant contributors. For the E. coli targets, we used the average number of probable targets associated with each type of diarrheagenic E. coli.

Statistical Analysis
The primary analytic approach was multivariable logistic regression to assess significant associations between candidate targets and diarrhea. To account for correlation between multiple observations from the same infant, the Huber and White sandwich estimator [29, 30] was used to determine the variance-covariance matrix. We assumed a “working independence model” to obtain estimates of the coefficient and obtained unbiased robust estimates of variances and covariances of these estimates by adjusting for the correlation between multiple observations from cluster samples. We prespecified potential risk predictors based on the significant results of univariate analyses and then used backward elimination to identify robust predictors for inclusion in the final multivariable logistic regression model. A Wald $\chi^2$ test was used to assess the significance of each predictor, and the odds ratio (OR) was used to quantify its effect. Internal model validation was determined by bootstrap model validation [31] to assess how accurately the tested models would predict outcomes for a new sample of data. A bootstrapped corrected C index or area under the receiver operator characteristic curve was used as a measure of overall predictive discrimination, defined in this study as the ability to separate diarrhea samples from surveillance samples. A receiver operator characteristic curve area of 0.5 indicates no discrimination, and 1.0 indicates perfect discrimination.

Because we were also interested in assessing associations among infants with a higher diarrhea severity index (>6) and among those who with a first episode of diarrhea, we conducted subgroup analyses to predict the probability of developing diarrhea. A nonparametric Wilcoxon 2-sample test was used to determine whether there was any difference in number of pathogens between Virginia and Dhaka cohorts for diarrhea or surveillance samples. A linear mixed-effects regression model [32] was used to test for any difference in the number of pathogens between diarrhea and surveillance samples at different months of life for the Dhaka cohort. For this quasi-continuous outcome, the normality assumption was examined by standard diagnostics techniques.

Nonparametric Wilcoxon signed rank tests were used to examine differences between the percentages of positive results in diarrhea and surveillance stool samples for a specific pathogen or target over the entire study year. Bonferroni correction was used to adjust for the multiple comparisons in the number of pathogens detected in stool samples and in the percentage of samples positive for pathogens, and was implemented by multiplying the $P$ value by the number of comparisons of interest. For ordinal rankings of probable, dominant, or less-likely pathogens, and for number, duration, and severity of diarrheal episodes, we used means ± SDs. Unless otherwise indicated, means ± SEs were used elsewhere. All $P$ values were 2 sided, and differences were considered statistically significant at $P < .05$. All analyses were performed using either SAS 9.2 (SAS Institute) or R 2.9 (R Foundation for Statistical Computing; 2011) software.

RESULTS
Enteropathogen Burden in the First Year of Life in Dhaka
The mean number of diarrhea episodes was 4.7 ± 2.3 per infant per year, and specimens were collected for 420 (60.6%) of 693 reported diarrheal episodes within the time windows for acceptability. The mean duration of these episodes was 5.5 ± 4.0 days, and 91% of episodes were mild (severity score ≤6; mean severity score, 4.4 ± 1.6). The analysis included 1385 surveillance stool samples, after the exclusion of 358 samples obtained in close proximity to a diarrheal episode (within 7 days). The number of enteropathogens detected in nondiarrheal stool samples in the first year of life was striking (median, 3 in Dhaka vs 0 in Virginia; interquartile range [IQR], 2–4 vs 0–1; mean, 3.3 vs 0.5; nonparametric Wilcoxon 2-sample test, $P < .05$). This enteropathogen burden was evident immediately, by the first month of life (Figure 1). By comparison, in Virginia very few enteropathogens were detected in either diarrheal or surveillance stool samples (median, 0; IQR, 0–1; n = 31), with only rotavirus, norovirus GI and GII, Salmonella, and eae detected in diarrheal samples.

In Dhaka, the number of enteropathogens was significantly higher in diarrheal than surveillance stool specimens at all time points during the first year of life (linear mixed-effects regression model; Bonferroni adjusted $P < .05$) except months 6 and 10. Overall, there was a significant difference in the number of pathogens between surveillance and diarrheal specimens (estimated mean difference, $−1.25$; 95% CI, $−1.42$ to $−1.09$; Bonferroni adjusted $P < .05$). At least 1 pathogen was detected in 100% of diarrheal specimens. Coinfections were the norm (96.2% of specimens), with a median of 5 (IQR, 3–6; mean, 4.5) enteropathogens detected per diarrheal specimen.

Specific Enteropathogens Detected and Their Association With Diarrhea
We first analyzed causes of diarrhea by comparing the distribution of enteropathogens found in diarrhea versus surveillance
stool specimens (Figure 2). After Bonferroni correction, we detected a higher rate of rotavirus, *Campylobacter*, *Trichuris*, sapovirus, *Cyclospora*, *Ancylostoma*, and *E. histolytica* in diarrheal stool specimens (in order of statistical significance). Conversely, the EAEC marker *aatA* was found more frequently in surveillance specimens (Bonferroni adjusted *P* < .05). By univariate analysis, 16 targets were significantly associated with diarrhea as shown in Figure 3 (*P* < .05), and *aatA* (EAEC) was the only target negatively associated with diarrhea (*P* < .05). By multivariable analysis, *Cyclospora* (odds ratio [OR], 7.3; 95% confidence interval [CI], 3.1–18.4), rotavirus (OR, 6.3; 95% CI, 4.1–9.8), sapovirus (OR, 4.2; 95% CI, 2.2–8.1), *Trichuris* (OR, 2.8; 95% CI, 2.0–4.1), *Ancylostoma* (OR, 2.4; 95% CI, 1.4–4.0), *E. histolytica* (OR, 2.4; 95% CI, 1.7–3.5), *Campylobacter* (OR, 2.1; 95% CI, 1.6–2.7), EAEC (*aaiC*) (OR, 1.5; 95% CI, 1.1–2.0), and ETEC (*STh*) (OR, 1.4; 95% CI, 1.1–1.8) remained significantly associated with diarrhea (*P* < .05). The bootstrapped corrected C index of 0.75 from the internal model validation indicated that our multivariable model had good predictive discrimination. When only the first episode of diarrhea was analyzed against preceding surveillance stool samples, the association with diarrhea remained and increased for rotavirus (OR, 8.5; 95% CI, 3.5–20.1), *Trichuris* (OR, 8.6; 95% CI, 2.9–25.6), and *STh* (OR, 2.6; 95% CI, 1.2–5.6).

**Quantitative Approach to the Microbial Etiology of Diarrhea**

The average cMFI, a measure of PCR amplicon quantity [17–21], was significantly higher in diarrheal than in surveillance specimens for *eae* (median, 26.2 for diarrheal vs 20.8 for surveillance specimens; IQR, 16.1–50.7 vs 11.8–31.9; mean, 27.6 vs 10.6), nonparametric Wilcoxon 2-sample test, Bonferroni adjusted *P* < .05), *Salmonella* (median, 20.2 vs 8.3; IQR, 10.6–41.8 vs 3.6–11.2; mean, 27.6 vs 10.6), and *aaiC* (median, 99.4 vs 90.6; IQR, 40.6–136.4 vs 21.1–128.5; mean, 103.4 vs 83.2), suggesting that diarrhea may result from excess enteropathogen loads of EPEC, *Salmonella*, and EAEC, respectively. We also examined total enteropathogen intensity in each diarrheal specimen by summing the enteropathogen quartiles, and this intensity greatly exceeded that of surveillance specimens (median, 14 for diarrheal vs 11 for surveillance specimens; IQR, 10–18 vs 6–15;
mean, 14.1 vs 10.8; nonparametric Wilcoxon 2-sample test, \( P < .05 \)).

Next, we examined enteropathogen quantities over time, using each infant as his or her own internal control. Figure 4 shows a representative infant who had 3 diarrheal episodes in the first year of life (4.3 ± 0.6 enteropathogens in diarrheal vs 2.6 ± 1.5 in surveillance specimens). Using our definitions of probable and less-likely contributors (see Methods), we identified probable and less-likely targets (≥1 of each) in 94.0% and 77.9% of diarrheal episodes, respectively. Across the cohort, 3.3 ± 1.9 probable and 2.3 ± 1.9 less-likely targets were identified in each diarrheal specimen. The distribution of these categories by pathogen is shown in Figure 5. The most common pathogens judged to be probable contributors to diarrhea were EAEC (12.1% of all probable detections), Campylobacter (11.3%), EPEC (10.2%), rotavirus (7.9%), ETEC (6.8%), and E. histolytica (6.6%). Among moderate or severe episodes (severity score >6), the pathogen distribution shifted toward rotavirus (13.7% of all probable detections), E. histolytica (10.8%), and Cryptosporidium (8.5%) and away from EAEC (9.9%), Campylobacter (8.9%), and EPEC (6.4%).

**DISCUSSION**

The importance of this work is the demonstration that childhood diarrhea in this Bangladeshi setting seems to be
polymicrobial and may be better conceptualized as a syndrome of enteropathogen excess rather than a single infection. Amidst the enteropathogen milieu of this region, where carriage of multiple enteropathogen targets was the norm, diarrhea reliably occurred when a quantitative threshold was exceeded. For instance, 95% of the time when enteropathogen target cMFI quartiles exceeded a sum of 21 units, it was in the setting of diarrhea.

By analyzing the cohort longitudinally and accounting for prior pathogen exposure and burden, we were able to reduce the total enteropathogen targets detected in a diarrheal specimen from a mean of 5.6 to 3.3 “probable” contributors. This means that ≥3 enteropathogens seemed to act in concert to cause diarrhea.

Our analysis had to make several major assumptions, which were guided by how we would clinically approach the infant represented in Figure 4 if presented with all the data. First, we assumed that higher-burden infections were more likely to cause diarrhea. This quantitative principle has been proposed by molecular guidelines for infectious disease causation, which posit that putative pathogen nucleic acid sequence should correlate temporally and quantitatively with disease [33]. Others have found such quantitation suggestive in implicating diarrheagenic pathogens in other settings, such as EPEC in Peru [4].

Second, we analyzed the cohort longitudinally, accounting for each subject’s history of diarrhea and prior enteropathogen exposure. This makes a major assumption that primary infections are more likely to be symptomatic than secondary ones, which we acknowledge downplays the scenario of reinfection with a prior pathogen of new virulent subtype. That said, we think it supportive that the longitudinal quantitative approach (Figure 5), via a rather different analytical method, identified a list of important enteropathogens similar to that identified by the standard OR approach (Figure 3). For example, when we combine the multivariate ORs with pathogen prevalence to determine the adjusted attributable fraction [34], we find Campylobacter, EAEC, EPEC, and rotavirus to be the pathogens of highest burden—the same top 4 that we found with our longitudinal quantitative approach. The advantage to the longitudinal quantitative approach is that it implicates pathogens in every individual diarrheal episode, whereas the OR or attributable fraction approach only describes the population.

The implication of this study is that single-pathogen approaches may make limited gains in the management of diarrhea in such communities. Interventions against rotavirus may make the greatest gains, because this was by many measures the most important pathogen in this study (with a high number of episodes, a high OR for diarrhea, the highest OR for first episodes, very few unlikely contributions, and higher burdens among the more severe cases). However, the effect of rotavirus vaccination on reducing all-cause diarrhea in such settings is usually 10%–30%. [12–14]. This low rate is not surprising to us, given these findings, nor is it a disappointment. Subsequent single-pathogen approaches targeting lesser pathogens may be
less efficacious. It is also possible that if a single pathogen facilitates diarrhea caused by other pathogens, a proposed mechanism of protection by the LT ETEC vaccine [35], then certain single-pathogen approaches could have broader effects. In any event, our view is that multipathogen measures, such as water, sanitation, and hygiene improvements, should be prioritized.

Another clinically relevant result from our data was the differing ratios of probable to less-likely detections for the different enteropathogens. This has implications for how diagnostics should be interpreted clinically in these settings. Our study would suggest that detection of rotavirus, Shigella or enteroinvasive E. coli, Cryptosporidium, and E. histolytica is particularly relevant because >90% of detections seemed to be probably contributing to diarrhea. To take a different example, the EPEC gene eae yielded a diarrhea-associated OR of 1.7, of only modest statistical significance, and nearly half of all detections were judged less likely to be contributing to diarrhea. Were all detections interpreted as causal and treated with an antibiotic, many unnecessary treatments would result. Quantitatively, however, eae was detected at higher burdens in diarrheal specimens. Therefore, we would hypothesize that treating those highest-burden eae-positive diarrhea episodes would have greater prospects for success. Ultimately, intervention trials such as this are needed to test the validity of these quantitative diagnostics and the etiology of diarrhea.

We were surprised to find helminth DNA at a 5%–10% rate, particularly Trichuris, and surprised that it was highly diarrhea associated. Helminths were rarely seen with microscopy (only 0.6% Ascaris and 0.1% Trichuris in surveillance stool samples), which is not surprising because this was a study of 0–1-year-olds, and the helminth life cycles require multiple infections and several weeks to months to produce eggs [36]. Trichuris is highly prevalent in this region of Dhaka in older children, and its detection was peculiar in that it was frequently a probable contributor to diarrhea but rarely dominant. This means it was frequently observed with many more abundant copathogens, and we therefore speculate that it was a marker for a fecal-oral contamination event.

Our work had some limitations. One major assumption was that any detected nucleic acids was potentially pathogenic, in other words, that it reflected the presence of viable pathogens and was not merely killed material or benign transit. Of course, this is a caveat for any non-culture-based method. Our study emphasized quantitation to guard against this risk. Furthermore, carriage of enteropathogen nucleic acids was by no means universal in infants, because it was rare in Virginia. Moreover, although we tested for 29 enteropathogens, we acknowledge that no enteropathogen work-up from stool samples can be entirely complete. Other pathogens could be added, but it is unlikely that this would diminish our findings; more likely,
it would add to the polymicrobial milieu. Finally, this was a community-based study of mostly mild diarrhea episodes and results may differ among severe or hospital-treated diarrheal cases. In summary, our quantitative, longitudinal, molecular analysis of diarrheal stool samples suggests that multiple enteropathogens act in concert to cause childhood diarrhea in Dhaka, Bangladesh.

**Notes**

**Acknowledgments.** We thank the families who participated in the birth cohort, the Parasitology Laboratory at the International Centre for Diarrhoeal Disease Research, Bangladesh (sample collection and processing), Florida Nazib (sample archiving), Jaco Verweij (*Trichuris* assay design), Firdausi Qadri (advice on ETEC detection), Dorothy Bunyan (help with the Virginia study), Denise Toney (provision of Virginia diarrheal samples), Uma Nayak and Aleya Ferdous (data management), and Mark Conaway (review of statistical methods).

**Financial support.** This work was supported by the National Institute of Allergy and Infectious Diseases, National Institutes of Health (grant U01 AI075396 to E. R. H. and R01 AI043596 to W. A. P.) and the Bill and Melinda Gates Foundation.

**Potential conflicts of interest.** All authors: No reported conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

**References**