Association Between Stool Enteropathogen Quantity and Disease in Tanzanian Children Using TaqMan Array Cards: A Nested Case-Control Study

James A. Platts-Mills, Jean Gratz, Esto Mduma, Erling Svensen, Caroline Amour, Jie Liu, Athanasia Maro, Queen Saidi, Ndeallilia Swai, Happiness Kumburu, Benjamin J. J. McCormick, Gibson Kibi, and Eric R. Houpt*

Division of Infectious Diseases and International Health, Department of Medicine, University of Virginia, Charlottesville, Virginia; Haydom Lutheran Hospital, Haydom, Tanzania; Kilimanjaro Clinical Research Institute, Moshi, Tanzania; Fogarty International Center, National Institutes of Health, Bethesda, Maryland

Abstract. Etiologic studies of diarrhea are limited by uneven diagnostic methods and frequent asymptomatic detection of enteropathogens. Polymerase chain reaction-based stool pathogen quantification may help distinguish clinically significant infections. We performed a nested case-control study of diarrhea in infants from a community-based birth cohort in Tanzania. We tested 71 diarrheal samples and pre-diarrheal matched controls with a laboratory-developed TaqMan Array Card for 19 enteropathogens. With qualitative detection, no pathogens were significantly associated with diarrhea. When pathogen quantity was considered, rotavirus (odds ratio [OR] = 2.70 per log_{10} increase, \( P < 0.001 \)), astrovirus (OR = 1.49, \( P = 0.01 \)), and Shigella/enteroinvasive Escherichia coli (OR = 1.47, \( P = 0.04 \)) were associated with diarrhea. Enterotoxigenic \( E. \) coli (0.15 SD decline in length-for-age z score after 3 months per log_{10} increase, \( P < 0.001 \)) and \( Campylobacter \) jejuni/C. coli (0.11 SD decline, \( P = 0.003 \)) in pre-diarrheal stools were associated with poor linear growth. Quantitative analysis can help refine the association between enteropathogens and disease in endemic settings.

INTRODUCTION

Acute infectious diarrhea remains the world’s second leading cause of death in children under 5 years old, and in Africa, it is responsible for one-quarter of all deaths in this age group.1 Diarrhea incidence and mortality are highest in the first year of life.2,3 Studies of diarrhea in low- and middle-income countries are complicated by the diversity of required diagnostic methods for enteropathogens as well as high rates of asymptomatic detection.4,5 Furthermore, in most such studies, the etiology of a large percentage of episodes is not determined, likely because of poor test sensitivity. Nucleic acid amplification tests allow the use of a single highly sensitive diagnostic modality for the detection of a wide range of enteropathogens; however, the increased sensitivity comes at the cost of an increase in the background rate of detection for many pathogens.6,7 In this context, there is some evidence that quantification adds important additional information about the relationship between specific enteropathogens and disease.8–11 However, such an approach has not previously been used for an etiologic study involving more than one pathogen. Our laboratory has developed an arrayed single-plex polymerase chain reaction (PCR) assay for sensitive quantitative detection of 19 enteropathogens.5 The primary aim of this study was to assess whether enteropathogen quantification in stool samples could improve our understanding of the association between pathogen detection and disease by comparing pathogen quantity in diarrheal cases and matched asymptomatic controls.

MATERIALS AND METHODS

**Study population.** Samples were selected from an ongoing community-based birth cohort initiated in December of 2009 of 262 children in Haydom, Tanzania as part of an ongoing study on the Etiology, Risk Factors and Interactions of Enteric Infections and Malnutrition and the Consequences for Child Health and Development (MAL-ED). All subjects were enrolled by day 17 of life. In-home surveillance visits for symptoms of diarrhea were performed two times per week. A diarrhea episode was defined as three or more loose stools in 24 hours or visible blood in stool. Monthly surveillance stool samples as well as diarrheal samples were collected. Diarrheal samples could be collected during or up to 48 hours after cessation of diarrhea. All mothers were breastfeeding at the time of the diarrheal episode. At the time of this study, a total of 88 episodes of diarrhea had been captured in subjects less than 1 year of age. For 77 of these diarrheal episodes, a pre-diarrheal surveillance sample was available from the same individual for use as a matched asymptomatic control, which was collected no more than 8 weeks (range = 7–56 days) before the onset of diarrhea. In each case, the subject was diarrhea-free for at least 1 week before and after collection of the surveillance sample. Of these samples, 71 diarrheal samples and 71 matched controls from 61 unique subjects had stool available for nucleic acid extraction and testing. For 10 sample pairs that corresponded to second diarrheal episodes from a previously included subject, the median number of days between the first episode of diarrhea and the second pre-diarrheal control sample was 72 (range = 10–245). Informed consent was obtained from the parents or legal guardians of all subjects. The study was approved by the National Institute for Medical Research, Tanzania and the Institutional Review Board of the University of Virginia.

**Anthropometry.** Child lengths were measured on enrollment and monthly using a recumbent length measuring board and recorded in 0.1-cm increments. Length-for-age z scores (LAZs) were calculated using World Health Organization child growth standards.12 Three-month changes in LAZs (ΔLAZs) were calculated using the difference between the most recent LAZ before the episode of diarrhea as the baseline.

**Stool diagnostics.** DNA extraction was performed using the QIAmp DNA Stool Mini Kit (Qiagen, Valencia, CA) using...
a modified protocol, including bead beating to lyse organs.
isms. DNA was stored at −20°C until use. RNA was
extracted using the Quickgene RNA Tissue Kit II on the
Fujifilm Quickgene-810 (Fujifilm, Tokyo, Japan) and
stored at −80°C until testing. Extrinsic controls for DNA (107 phocine
herpesvirus) and RNA (107 copies MS2 bacteriophage) were
included in 60 specimens (30 cases and 30 matched controls) to
assess for bias in extraction or amplification efficiency between
diarrheal and non-diarrheal samples. TaqMan Array Cards
were used to amplify nucleic acid for 19 enteropathogens: rota-
virus, norovirus GII, adenovirus, astrovirus, sapovirus, enterotoxigenic Escherichia coli (ETEC), enteropathogenic E. coli (EPEC), enteroaggregative E. coli (EAEC), Shiga-toxigenic E. coli (STEC), Shigella/enteroinvasive E. coli (EIEC), Salmonella, Campylobacter jejuni/C. coli, Vibrio cholerae, Clostri-
dium difficile, Cryptosporidium, Giardia lamblia, Entamoeba histolytica, Ascaris lumbricoides, and Trichuris trichiura. Multi-
ple targets were tested for the diarrheagenic E. coli subgroups,
which were defined as follows: ETEC, ST and/or LT; EPEC, eae and bfpA; EAEC, aatA or aaiC; STEC, stx1 or stx2. Prior
testing of clinical specimens with this platform yielded a sensi-
tivity and specificity of 85% and 77%, respectively, compared
with conventional methods (including microscopy, culture, and
immunoassay) and a sensitivity and specificity of 98% and
96%, respectively, compared with molecular methods. Detection
below a quantification cycle (Cq) of 35 was used as a uniform
cutoff for test positivity based on a limit of detection calculated from
analytical specimens of 34.8 ± 1.2 (mean ± SD). Cq value was also used as an estimate of pathogen quantity for positive samples, where a 1-unit decrease in Cq corre-
sponds to a twofold increase in target copy number.

Statistical analysis. The proportions of cases and controls
positive for each pathogen were compared using the McNemar
test. The related samples Wilcoxon signed rank test was used to
compare continuous measures between diarrheal and surveil-
lance stools. Generalized estimating equations were used to fit
a logistic regression model to analyze the association between
stool enteropathogen quantity and both diarrhea and vomiting as
well as a linear regression model to analyze the association
between stool enteropathogen quantity and LAZ. An independ-
dent working correlation matrix was assumed. Age, sex, and
season were included in all regression models, and the pre-
diarrheal baseline LAZ was included in the growth model. The
backward elimination method was used to identify statistically
significant variables. For the association of enteropathogens
with diarrhea, a significance level of 0.05 was used. For all other
analyses, a significance level of 0.005 was used to adjust for
multiple comparisons. All odds ratios (ORs) represent the
increase in odds per log10 increase in pathogen quantity. The
population attributable fraction (AF) for each pathogen was
calculated as per the calculations in the work by Bruzzi and
others. In all analyses, negative samples were assigned the
cutoff Cq value. All statistical analyses were performed using

RESULTS
Child age at the time of diarrhea was 38–372 days. The
duration of diarrhea was 3.7 ± 2.3 days (mean ± SD), with a
maximum number of loose stools of 5.1 ± 1.2 per 24 hours.
Visible blood was present in seven (9.9%) diarrheal stools.
Vomiting accompanied diarrhea in 23 (32.4%) episodes. With
qualitative detection using the Cq cutoff of 35, there was no
significant difference in the number of pathogens found in
cases (median = 3, range = 0–7, mean = 3.1) and the preced-
ing controls (median = 3, range = 0–8, mean = 3.0, P = 0.73).
The proportions of case and control stools that were positive
were not significantly different for any of the tested patho-
gens (Figure 1).

![Figure 1](Image)
We then pursued a quantitative analysis of enteropathogen burden in stool samples. We observed no systematic bias in extraction or amplification efficiency between cases and controls for either the DNA (median Cq for cases = 27.0, interquartile range [IQR] = 25.1–27.6; median Cq for controls = 27.1, IQR = 24.7–27.8; P = 0.57) or RNA extrinsic controls (median Cq for cases = 28.0, IQR = 26.9–29.3; median Cq for controls = 27.9, IQR = 27.1–29.3; P = 0.94); therefore, we proceeded with a comparison of the unadjusted Cq values. Rotavirus and Shigella/EIEC were detected at higher quantities in diarrheal stools (Figure 2). Differences for astrovirus, ETEC, and Cryptosporidium were not significant (P = 0.37, P = 0.41, and P = 0.31, respectively). Pathogens for which quantity was statistically significantly associated with diarrhea in the logistic regression model are shown in Table 1. The pathogen-specific AF for diarrhea in this cohort is also shown. The association between pathogen quantities detected in each individual pair of samples for the pathogens found to be associated with diarrhea is shown in Figure 3. Using the same quantitative approach, Shigella/EIEC burden was significantly associated with visible blood in diarrheal stools (OR = 2.84 per log_{10} increase in pathogen quantity, P = 0.001). Rotavirus (OR = 2.98, P = 0.001) was significantly associated with vomiting accompanying diarrhea, with a non-significant trend for astrovirus (OR = 1.76, P = 0.007).

Baseline and follow-up LAZs approximately 3 months (range = 85–95 days) after collection of the control samples were available for 66 case-control pairs, and the mean ∆LAZ (SD) was −0.20 (0.65). We looked for associations between the quantity of specific pathogens in either diarrheal cases or pre-diarrheal controls and subsequent linear growth. No specific pathogen quantity in diarrheal stools was significantly associated with poor growth. However, asymptomatic detection of ETEC (0.15 SD decline in LAZ per log_{10} increase in quantity, P < 0.001) and C. jejuni/C. coli (0.11 SD, P = 0.003) in pre-diarrheal stools was significantly associated with poor linear growth.

**DISCUSSION**

The major finding of this study is that an analysis of pathogen nucleic acid quantity in stool samples using quantitative PCR offers better resolution for describing the relationship between pathogen infection and disease. Importantly, it suggests that the improved sensitivity of nucleic acid amplification testing may obscure meaningful disease associations unless infectious burden is used to distinguish high-quantity clinically significant detection from low-quantity detection of unclear significance, which is not new to clinical infectious diseases; knowledge of the quantity of pathogen present in

---

**Figure 2.** Box plots of quantification cycle results in case and control samples positive for each pathogen. Only pathogens with more than two detections in cases are shown. Whiskers represent minimum and maximum values. * The quantities of rotavirus (P = 0.03) and Shigella (P = 0.05) were significantly higher in diarrheal stools.

**Table 1**

<table>
<thead>
<tr>
<th>Pathogens</th>
<th>P value</th>
<th>OR (95% CI)</th>
<th>Prevalence in cases (n)</th>
<th>AF (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rotavirus</td>
<td>&lt; 0.001</td>
<td>2.70 (1.60–4.55)</td>
<td>21</td>
<td>14.1 (10.5–15.8)</td>
</tr>
<tr>
<td>Shigella/EIEC</td>
<td>0.04</td>
<td>1.47 (1.02–2.13)</td>
<td>19</td>
<td>12.5 (1.1–16.4)</td>
</tr>
<tr>
<td>Astrovirus</td>
<td>0.003</td>
<td>1.49 (1.10–2.02)</td>
<td>12</td>
<td>8.1 (3.3–9.9)</td>
</tr>
</tbody>
</table>

ORs represent the change in odds per log_{10} increase in pathogen quantity. CI = confidence interval.
clinical samples has been used in other contexts to distinguish infection from contamination or asymptomatic carriage.\textsuperscript{17,18} Furthermore, pathogen burden in stool is currently used implicitly: stool culture relies on the ability of a pathogen to outcompete normal flora, and quantitative culture methods confirm that a pathogen is more likely to do so when present at a higher burden.\textsuperscript{19} Cryptosporidium oocyst excretion quantity correlates with diarrheal symptoms.\textsuperscript{20} A correlation between enzyme-linked immunosorbent assay (ELISA) optical density and Cq value has been shown for norovirus and \textit{Giardia}.\textsuperscript{21,22} Thus, we frequently use quantitative or semi-quantitative tests but interpret them dichotomously.

An increasing number of studies have analyzed the relationship between stool pathogen burden and manifestations of disease using molecular tests. Kang and others\textsuperscript{8} showed a statistically significant correlation between rotavirus PCR Cq and disease severity in children with acute gastroenteritis.\textsuperscript{8} Barletta and others\textsuperscript{9} compared the quantity of the intimin gene of EPEC isolated from healthy and diarrheal stools from Peru and observed that gene quantity was higher in diarrheal samples. In the present study, we show more broadly the value of an analysis of quantitative enteropathogen detection for epidemiologic studies of diarrhea. Rotavirus and \textit{Shigella} are established causes of diarrheal disease in children worldwide, and their importance in this population is confirmed here. To our knowledge, astrovirus has not been associated with diarrhea in a controlled study of diarrheal etiology in Africa.\textsuperscript{23} A single case, representing a prevalence of 0.4\%, was previously reported in a hospital-based study of childhood diarrhea in Dar es Salaam, Tanzania.\textsuperscript{24} Here, we found astrovirus to be responsible for 8.1\% of the diarrheal episodes captured in this cohort. Interestingly, all of the high-burden astrovirus detections in diarrheal stools were in the first 6 months of life. A particularly high prevalence of astrovirus infection early in life has been previously reported in Egypt.\textsuperscript{25}

The ORs reported here for association with diarrhea are per log\textsubscript{10} increase above the Cq cutoff for positivity. For example, the median quantity of \textit{Shigella}/EIEC detected in positive cases had a Cq of 25.2, which was 2.95 logs below the cutoff and thus, corresponds to an OR of 5.0. In turn, the AFs reported are integrated across individual estimates of the strength of association based on the quantity of pathogen detected in each diarrheal episode. These data can also show the sensitivity of the AF estimate to the Cq cutoff used if one were to dichotomize the test results. Although such comparisons were not significant in this study, using \textit{Shigella}/EIEC as an example, a cutoff of 33 yields a point estimate of the AF of 16.4\%, whereas a cutoff of 32 yields an AF of 13.4\% and a cutoff of 34 yields an AF of 8.2\%. This result shows the power of the diagnostic method to impact epidemiologic studies of diarrhea etiology. A quantitative analysis permits the use of more sensitive tests without substantially diluting the association between detection and disease.

The link between both ETEC and \textit{Campylobacter} infection and malnutrition has been previously described. In Bangladesh, ETEC diarrhea was associated with a significant negative effect on weight gain in children less than 5 years old.\textsuperscript{26} Meanwhile, baseline malnutrition has been associated with an increased incidence and severity of ETEC infection.\textsuperscript{27} Lee et al. have described an association between \textit{Campylobacter} infection and reduced early-childhood weight gain and linear growth in Peru.\textsuperscript{28} Additional longitudinal

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3.png}
\caption{Quantification cycle results for each case and asymptomatic matched control collected before the episode ($N = 71$ pairs) for the pathogens found to be associated with diarrhea in this study. (A) Astrovirus. (B) Rotavirus. (C) \textit{Shigella}/EIEC.}
\end{figure}
testing using archived specimens from the MAL-ED Network is planned to validate and better understand this relationship.

Our study has several important limitations. Despite the additional statistical power attained by treating the exposures (i.e., pathogen quantities) as continuous variables, this study was underpowered to detect all pathogens associated with diarrhea. As such, this study is a preliminary study, showing the use of such an approach, and analysis of additional spec-

imens from the larger MAL-ED study will be forthcoming when the study is complete. Of the pathogens tested in this study, the prevalence of adenovirus, Campylobacter, ETEC, norovirus, and EAEC was higher than the prevalence for astrovirus, rotavirus, and Shigella/EIEC. This finding sug-

gests that any undetected association with diarrhea for the former five pathogens is less strong than the association detected here for the latter three pathogens. AF estimates are only provided for pathogens significantly associated with diarrhea, which may exclude pathogens with a substantial burden of diarrhea despite being less pathogenic. However, this work stands as proof of principle for the use of pathogen quantity quantification in this setting. To improve the capture of diar-

rehal episodes, collection of samples was allowed up to 48 hours after cessation of diarrhea. If pathogen nucleic acid quantity declines rapidly after cessation of symp-

toms, it could decrease our ability to detect an association of pathogen quantity with diarrhea. Very little data with serial quantitative measurements during and after diarrhea are available to address this possibility, although a norovirus challenge study found that nucleic acid quantity did not decline for several days after cessation of symptoms and often peaked after resolution of diarrhea. Similarly, it is possible that quantification of nucleic acid may not be com-

parable between liquid and solid stools because of dilution of either target nucleic acid or inhibitors. Our extrinsic controls showed no evidence of quantification bias by stool type, and furthermore, we would expect that these dilutional effects would be small relative to differences of more than 1 million-

fold in pathogen quantity between samples.

We expect that quantitative approaches will become more common as a way to take full advantage of the excellent analytic sensitivity and specificity of molecular diagnostics, improving our understanding of the association between path-

ogens and not only diarrhea but also malnutrition, gut function, and vaccine immunogenicity. They also offer the opportunity to better describe subtle interactions between pathogens (e.g., whether the presence of one pathogen changes the relation-

ship between the burden of a second pathogen and disease). The complex dynamic environment of the gut is particularly well-suited for this quantitative approach.

Received July 30, 2013. Accepted for publication September 14, 2013.

Acknowledgments: This project was undertaken in collaboration with the Etiology, Risk Factors and Interactions of Enteric Infections and Malnutrition and the Consequences for Child Health and Development Project (MAL-ED) Network. The MAL-ED Network is carried out as a collaborative project supported by the Bill & Melinda Gates Foundation, the Foundation for the National Institutes of Health, and the National Institutes of Health/Fogarty International Center. The authors thank the staff and participants of the MAL-ED Network Project for their important contributions.

Financial support: This work was supported by National Institute of Allergy and Infectious Diseases, National Institutes of Health Grant U01 AI075396 and Bill & Melinda Gates Foundation Global Health Grant OPP1019093 (to E.R.H.).

Authors’ addresses: James A. Platts-Mills, Jean Gratz, Jie Liu, and Eric R. Houpt, Division of Infectious Diseases and International Health, Department of Medicine, University of Virginia, Charlottesville, VA. E-mails: jpl5@virginia.edu, jean.gratz@gmail.com, jli3yj@virginia.edu, and erhbk@virginia.edu. Esto Mduma, Erling Svensen, and Caroline Amour, Haydom Lutheran Hospital, Haydom, Tanzania, E-mails: estomduma@haydom.co.tz, erling.svensen@chuhb.no, and lynecamour@gmail.com. Athanasia Maro, Queen Saidi, Ndeallilia Swai, Happiness Kumburu, and Gibson Kibiki, Kilimanjaro Clinical Research Institute, Moshi, Tanzania, E-mails: athanasia.marom@yahoo.com, queensaidi@gmail.com, lilu2005tz@yahoo.co.uk, hk9@virginia.edu, and g.kibiki@krci.ac.tz. Benjamin J. J. McCormick, Fogarty International Center, National Institutes of Health, Bethesda, MD, E-mail: ben.mcormick@gmail.com.

REFERENCES

tinal disease using viral load. BMC Infect Dis 9: 63.


