Published in final edited form as:

Lancet. 2016 September 24; 388(10051): 1291–1301. doi:10.1016/S0140-6736(16)31529-X.

Use of quantitative molecular diagnostic methods to identify causes of diarrhoea in children: a reanalysis of the GEMS case-control study

Jie Liu, PhD[*],
Division of Infectious Diseases and International Health, University of Virginia, Charlottesville, VA, USA

James A Platts-Mills, MD[*],
Division of Infectious Diseases and International Health, University of Virginia, Charlottesville, VA, USA

Jane Juma, MS[†],
Center for Global Health Research, Kenya Medical Research Institute, Nairobi, Kenya

Furqan Kabir, MS[†],
Department of Paediatrics and Child Health, Aga Khan University, Karachi, Pakistan

Joseph Nkeze, MS[†],
Center for Vaccine Development and Institute of Global Health, University of Maryland School of Medicine, Baltimore, MD, USA; Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA

Catherine Okoi, BS[†],
Medical Research Council Unit, Banjul, The Gambia

Darwin J Operario, PhD[†],
Division of Infectious Diseases and International Health, University of Virginia, Charlottesville, VA, USA

Jashim Uddin, MS[†],
International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR, B), Dhaka, Bangladesh

Shahnawaz Ahmed, MBBS,

Correspondence to: Prof Eric R Houpt, Division of Infectious Diseases and International Health, University of Virginia, Charlottesville, VA 22908, USA, erh6k@virginia.edu.
[*]Contributed equally
[†]Contributed equally

Contributors: JL and JAP-M led data acquisition, data analysis, and writing of the report. JJ, FK, JN, COk, DJO, and JU led data acquisition at the site laboratories and did the site data analyses. SA, SMB, JG, SJ, BK, COc, JBO, MO, COn, AK, FA, SQ, MT, and DT contributed to data acquisition and site data analyses. WCB, TLM, and YW contributed to data analysis. JHR and SES contributed to data acquisition. PLA, MA, RFB, ASGF, BF, RH, AH, MJH, FQ, NTI, IM, SP, TR, DS, SOS, DS, BT, SMT, and AZ contributed to data acquisition and interpretation. JPN, KLK, and MML designed the original protocol and contributed to data interpretation. ERH conceived the project and contributed to data acquisition, analysis, and interpretation and writing of the report.

Declaration of interests: We declare no competing interests.

International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR, B), Dhaka, Bangladesh

Prof Pedro L Alonso, MD,
Centro de Investigação em Saúde da Manhiça, Maputo, Mozambique; Barcelona Centre for International Health Research (CRESIB, Hospital Clinic-Universitat de Barcelona), Barcelona, Spain

Prof Martin Antonio, PhD,
Medical Research Council Unit, Banjul, The Gambia

Stephen M Becker, PhD,
Division of Infectious Diseases and International Health, University of Virginia, Charlottesville, VA, USA

Prof William C Blackwelder, PhD,
Center for Vaccine Development and Institute of Global Health, University of Maryland School of Medicine, Baltimore, MD, USA; Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA

Prof Robert F Breiman, MD,
Global Disease Detection Division, Kenya Office of the US Centers for Disease Control and Prevention, Nairobi, Kenya

Abu S G Faruque, MD,
International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR, B), Dhaka, Bangladesh

Barry Fields, PhD,
Global Disease Detection Division, Kenya Office of the US Centers for Disease Control and Prevention, Nairobi, Kenya

Jean Gratz, MS,
Division of Infectious Diseases and International Health, University of Virginia, Charlottesville, VA, USA

Rashidul Haque, PhD,
International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR, B), Dhaka, Bangladesh

Anowar Hossain, MD,
International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR, B), Dhaka, Bangladesh

M Jahangir Hossain, MBBS,
Medical Research Council Unit, Banjul, The Gambia

Sheikh Jarju, DVM,
Medical Research Council Unit, Banjul, The Gambia

Farah Qamar, MBBS,
Department of Paediatrics and Child Health, Aga Khan University, Karachi, Pakistan

Najeeha Talat Iqbal, PhD,
Department of Paediatrics and Child Health, Aga Khan University, Karachi, Pakistan

Brenda Kwambana, PhD,
Medical Research Council Unit, Banjul, The Gambia

Inacio Mandomando, PhD,
Centro de Investigação em Saúde da Manhiça, Maputo, Mozambique

Timothy L McMurry, PhD,
Public Health Sciences, University of Virginia, Charlottesville, VA, USA

Caroline Ochieng, MS,
Center for Global Health Research, Kenya Medical Research Institute, Nairobi, Kenya

John B Ochieng, PhD,
Center for Global Health Research, Kenya Medical Research Institute, Nairobi, Kenya

Melvin Ochieng, BS,
Center for Global Health Research, Kenya Medical Research Institute, Nairobi, Kenya

Clayton Onyango, PhD,
Global Disease Detection Division, Kenya Office of the US Centers for Disease Control and Prevention, Nairobi, Kenya

Sandra Panchalingam, PhD,
Center for Vaccine Development and Institute of Global Health, University of Maryland School of Medicine, Baltimore, MD, USA

Adil Kalam, MS,
Department of Paediatrics and Child Health, Aga Khan University, Karachi, Pakistan

Fatima Aziz, MS,
Department of Paediatrics and Child Health, Aga Khan University, Karachi, Pakistan

Shahida Qureshi, MS,
Department of Paediatrics and Child Health, Aga Khan University, Karachi, Pakistan

Thandavarayan Ramamurthy, PhD,
National Institute of Cholera and Enteric Diseases, Kolkata, India

James H Roberts,
Public Health Sciences, University of Virginia, Charlottesville, VA, USA

Debasish Saha, PhD,
Medical Research Council Unit, Banjul, The Gambia

Prof Samba O Sow, MS,
Centre pour le Développement des Vaccins, Bamako, Mali

Suzanne E Stroup, MS,
Division of Infectious Diseases and International Health, University of Virginia, Charlottesville, VA, USA

Dipika Sur, MD,

National Institute of Cholera and Enteric Diseases, Kolkata, India

Boubou Tamboura, PhD,
Centre pour le Développement des Vaccins, Bamako, Mali

Mami Taniuchi, PhD,
Division of Infectious Diseases and International Health, University of Virginia, Charlottesville, VA, USA

Sharon M Tennant, PhD,
Center for Vaccine Development and Institute of Global Health, University of Maryland School of Medicine, Baltimore, MD, USA; Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA

Deanna Toema, BS,
Center for Vaccine Development and Institute of Global Health, University of Maryland School of Medicine, Baltimore, MD, USA; Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA

Yukun Wu, PhD,
Center for Vaccine Development and Institute of Global Health, University of Maryland School of Medicine, Baltimore, MD, USA; Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA

Prof Anita Zaidi, MBBS,
Department of Paediatrics and Child Health, Aga Khan University, Karachi, Pakistan

Prof James P Nataro, MD,
Department of Pediatrics, University of Virginia, Charlottesville, VA, USA

Prof Karen L Kotloff, MD,
Center for Vaccine Development and Institute of Global Health, University of Maryland School of Medicine, Baltimore, MD, USA; Department of Pediatrics, University of Maryland School of Medicine, Baltimore, MD, USA; Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA

Prof Myron M Levine, MD, and
Center for Vaccine Development and Institute of Global Health, University of Maryland School of Medicine, Baltimore, MD, USA; Department of Pediatrics, University of Maryland School of Medicine, Baltimore, MD, USA; Department of Medicine, University of Maryland School of Medicine, Baltimore, MD, USA

Prof Eric R Houpt, MD
Division of Infectious Diseases and International Health, University of Virginia, Charlottesville, VA, USA

Summary

Background—Diarrhoea is the second leading cause of mortality in children worldwide, but establishing the cause can be complicated by diverse diagnostic approaches and varying test

characteristics. We used quantitative molecular diagnostic methods to reassess causes of diarrhoea in the Global Enteric Multicenter Study (GEMS).

Methods—GEMS was a study of moderate to severe diarrhoea in children younger than 5 years in Africa and Asia. We used quantitative real-time PCR (qPCR) to test for 32 enteropathogens in stool samples from cases and matched asymptomatic controls from GEMS, and compared pathogen-specific attributable incidences with those found with the original GEMS microbiological methods, including culture, EIA, and reverse-transcriptase PCR. We calculated revised pathogen-specific burdens of disease and assessed causes in individual children.

Findings—We analysed 5304 sample pairs. For most pathogens, incidence was greater with qPCR than with the original methods, particularly for adenovirus 40/41 (around five times), *Shigella* spp or enteroinvasive *Escherichia coli* (EIEC) and *Campylobactor jejuni* or *C coli* (around two times), and heat-stable enterotoxin-producing *E coli* ([ST-ETEC] around 1·5 times). The six most attributable pathogens became, in descending order, *Shigella* spp, rotavirus, adenovirus 40/41, ST-ETEC, *Cryptosporidium* spp, and *Campylobacter* spp. Pathogen-attributable diarrhoeal burden was 89·3% (95% CI 83·2–96·0) at the population level, compared with 51·5% (48·0–55·0) in the original GEMS analysis. The top six pathogens accounted for 77·8% (74·6–80·9) of all attributable diarrhoea. With use of model-derived quantitative cutoffs to assess individual diarrhoeal cases, 2254 (42·5%) of 5304 cases had one diarrhoea-associated pathogen detected and 2063 (38·9%) had two or more, with *Shigella* spp and rotavirus being the pathogens most strongly associated with diarrhoea in children with mixed infections.

Interpretation—A quantitative molecular diagnostic approach improved population-level and case-level characterisation of the causes of diarrhoea and indicated a high burden of disease associated with six pathogens, for which targeted treatment should be prioritised.

Funding—Bill & Melinda Gates Foundation.

Introduction

Diarrhoea remains the second leading cause of death in children younger than 5 years worldwide, being associated with around 500 000 deaths per year, and ranks sixth in global disability-adjusted life-year burden.[1–3] Public health interventions rely on estimates of pathogen-specific burden for prioritisation. Previous estimates of the infectious causes of diarrhoea have been derived from studies that used varying approaches for pathogen detection,[4] including different media, stool concentration techniques, and detection methods, from microscopy to antigen detection to PCR. Each method has its own sensitivity, which could substantially affect burden estimates. Furthermore, asymptomatic carriage of some enteropathogens is common in children living in developing countries, for which few studies have attempted to control.[5–7] The quantity of nucleic acid is thought to distinguish clinical disease from asymptomatic shedding.[8–14]

We have shown previously that quantitative real-time PCR (qPCR) approaches could detect bacterial, viral, and parasitic pathogens across multiple laboratories with high sensitivity and good reproducibility and quantification.[15] We therefore did this reanalysis of specimens from the Global Enteric Multicenter Study (GEMS), which was a large case-control study of moderate to severe diarrhoea in children younger than 5 years in Africa and Asia,[6] to inform

estimates of the global burden of disease.[3] We aimed to reassess estimates of pathogen-specific diarrhoea causes and incidence at the population level, and to provide a case-level understanding of mixed diarrhoeal infections.

Methods

Samples

GEMS samples were obtained from regions in seven countries in Asia (Bangladesh, India, and Pakistan) and Africa (The Gambia, Kenya, Mali, and Mozambique).[6] These countries were chosen to represent a range of child health indicators and urban and rural settings, and because they have sufficient research infrastructures.[6,16] Each country provided a population census via a demographic surveillance system to enable estimation of population-level incidence of diarrhoea. Inclusion criteria for cases were three or more loose stools within 24 h and onset of the episode within the previous 7 days, after at least 7 days free from diarrhoea, and with at least one of sunken eyes, loss of skin turgor, intravenous hydration, dysentery, or hospital admission. Three age strata were assessed: infants (0–11 months), toddlers (12–23 months), and children (24–59 months). At least one control without diarrhoea in the previous 7 days, matched for age (within 2 months for patients aged 0–23 months and within 4 months for patients aged 24–59 months), sex, and residence (same or nearby village or neighbourhood), was enrolled within 14 days of each index case.

The original study enrolled 9439 children with diarrhoea and 13 129 children without diarrhoea. The microbiological diagnostic methods used in GEMS had included culture for bacteria, EIA for rotavirus, adenovirus, and protozoa, and multiplex reverse-transcriptase PCR (RT-PCR) with gel electrophoresis for norovirus, sapovirus, and astrovirus.[6] For this study, we retested up to 300 randomly selected cases and the first available matched controls from each age stratum and study site. Stool specimens had been collected between December, 2007, and March, 2011, and stored at –80°C until testing. Ethics approval for this study was obtained from the University of Maryland, MD, USA, all field sites, and the University of Virginia, VA, USA.

Procedures

We assessed all enteropathogens associated with diarrhoea in the original GEMS as well as those detected by molecular methods that had a possible association with diarrhoea in univariate analysis. We formulated a custom TaqMan Array Card (Thermo Fisher, Carlsblad, CA, USA) that compartmentalised probe-based qPCR assays for 32 enteropathogens. This platform has shown high reproducibility within and across sites.[15] All the assays have been described previously and have been extensively validated (appendix).[15,17,18] Nucleic acid was extracted with the QIAamp Fast DNA Stool mini kit (Qiagen, Hilden, Germany) with pretreatment steps that included bead beating.[18] We added two external controls, bacteriophage MS2 and phocine herpesvirus, to monitor efficiency of nucleic-acid extraction and amplification. We included one extraction blank per batch and one no-template amplification control per three cards to exclude laboratory contamination (appendix).

Case and control samples were tested simultaneously. Quantification cycles (Cqs) are the PCR cycle values at which fluorescence from amplification exceeds the background, which acts as an inverse metric of quantity of nucleic acid. All detections with a Cq greater than 35 were deemed negative.[15] Valid results required proper functioning of controls and excluded data flagged by the PCR software ViiA 7 (version 1.2.4, appendix). Pathogen target copy numbers were calculated (appendix). In a subset of 150 case-control pairs in the 12–23 month age group from each of India, Mali, and Mozambique, confirmatory qPCR assays were done to verify detection of *Shigella* spp or enteroinvasive *Escherichia coli* (EIEC) and identify *Shigella flexneri* and *Shigella sonnei* (appendix).

Statistical analysis

We retained the conceptual framework and analytical approach used in the original study.[6,19] Specifically, we used population attributable fractions to adjust for the high prevalence of asymptomatic pathogen carriage. For all analyses, we used the Cq values as inverse measures of pathogen quantity, where a one-unit increase corresponded to a two times decrease in quantity. We then estimated pathogen-specific burdens of diarrhoea by calculating attributable fractions, which incorporated the prevalence of the pathogen in cases and the strength of association between the pathogen quantity and case or control status.[20] Association strength for each pathogen was estimated by calculation of odds ratios (ORs) and 95% CIs in a multivariable conditional logistic regression model, where the outcome was case or control status and predictors were pathogen quantity, the quantity of other pathogens, a random slope for each study site, and an interaction between pathogen and age stratum. We hypothesised that the relation between pathogen quantity and case or control status could be non-linear and, therefore, included a quadratic term for pathogen quantity, which maintained or improved model fit for all pathogens, as assessed by the Akaike information criterion. We also considered power and spline models but rejected them on the basis of model fit. Attributable fractions were calculated by summing pathogen attributions across each of *j* cases with the following equation:

$$\sum_{i=1}^{j} AF_i$$

where $AF_i = 1/j \times (1 - 1/OR_i)$. 95% CIs were estimated by bootstrapping with 1000 iterations. Attributable incidence rates were calculated from the attributable fractions.[6] We extrapolated attributable fractions to pathogen-specific population attributable incidence, using the incidence of moderate to severe diarrhoea from demographic surveillance. To identify model-derived Cq cutoff values for detection of diarrhoea-associated pathogens that were independent of age, we fitted models that excluded the interaction between pathogen and age group but were otherwise identical to those used for the attributable fraction calculations. We then defined a pathogen as being diarrhoea associated if it was detected at a quantity for which the lower 95% CI of the OR exceeded 1. We used Wilcoxon's rank sum test to examine whether the quantity of *Shigella* spp or EIEC was different when detected in dysenteric cases. All analyses were done in R version 3.2.2.

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

We tested 11 400 specimens from 5700 cases and matched controls in five laboratories that showed good assay reproducibility (table 1). Identical positive-control constructs were provided to all laboratories, with consistent results for assay linearity (appendix). Valid results were obtained from 5304 (93·1%) matched case-control pairs (table 1). When compared with qPCR, the original diagnostic methods showed low and variable sensitivity but high specificity (98% [SD 2%], appendix).[15] Samples positive by the original diagnostic methods had higher quantities with molecular testing methods than did originally negative samples, across all 16 pathogens detected in the original study.

Shigella spp or EIEC, adenovirus 40/41, rotavirus, *Cryptosporidium* spp, and heat-stable enterotoxin-producing *E coli* (ST-ETEC) had high prevalence and strong quantity-dependent associations with diarrhoea (figure 1). *Campylobacter jejuni* or *C coli*, sapovirus, and typical enteropathogenic *E coli* (tEPEC) were also highly prevalent, but even at the highest quantities were only moderately associated with diarrhoea. Norovirus GII, *Helicobacter pylori*, astrovirus, *Vibrio cholerae*, *Salmonella* spp, *Cyclospora cayetanensis*, and *Entamoeba histolytica* also showed associations with diarrhoea but prevalence was low (figure 1). Entero aggregative *E coli* and *Aeromonas* spp were associated with diarrhoea only in specific study sites or age strata (enteroaggregative *E coli* was associated with diarrhoea only in the 12–13 months age group in Mozambique; *Aeromonas* spp was associated with diarrhoea only in the 0–11 months age group in Pakistan and in the 24–59 months age group in Pakistan, Mozambique, and Bangladesh). *Cystoisospora belli* was associated only with dysentery. All other enteropathogens tested were not associated with diarrhoea at any quantity or site or in any age group (appendix). The quantitative relations were robust across age groups (appendix).

Compared with the original microbiological methods, more pathogens were attributed as causes of diarrhoea, with more consistent attribution across sites. When assessed by age strata and study site, the pathogens with the highest attributable fractions for childhood diarrhoea were *Shigella* spp or EIEC, rotavirus, adenovirus 40/41, ST-ETEC, and *Cryptosporidium* spp (table 2). *Shigella* spp or EIEC, rotavirus, and ST-ETEC were consistently high across all age groups and regions. *Cryptosporidium* spp and adenovirus 40/41 were most evident in children up to age 23 months and the prevalence of *H pylori* was notable in children aged 24 months and older. *C jejuni* or *C coli* were particularly important causes in infants at the Bangladesh, India, Mali, and Mozambique sites. Estimates were robust irrespective of whether raw Cqs or those adjusted to standard curves and external controls to derive a pathogen target copy number were used (appendix).

For most pathogens, the qPCR-derived attributable incidence surpassed the original estimate, including for adenovirus 40/41 (by around five times), *Shigella* spp or EIEC and *C jejuni* or

C coli (around two times), and ST-ETEC (around 1·5 times; figure 2). With few exceptions, the increase in attribution for these pathogens was consistent across study sites (appendix). For rotavirus and *Cryptosporidium* spp, the attributable incidence remained similar to the original estimates (obtained with EIA). In infants, six pathogens had the greatest attributable incidence: in descending order, rotavirus (7·9 episodes per 100 child-years, 95% CI 6·2–9·7), adenovirus 40/41 (3·9, 2·4–5·4), *Cryptosporidium* spp (3·3, 2·2–4·4), *C jejuni* or *C coli* (2·4, 1·3–3·4), *Shigella* spp or EIEC (2·0, 1·4–2·6), and ST-ETEC (2·0, 1·4–2·6). In toddlers, the hierarchy was *Shigella* or EIEC (7·0, 5·0–9·0), rotavirus (4·1, 3·0–5·1), adenovirus 40/41 (1·8, 0·9–2·6), ST-ETEC (1·7, 1·1–2·3), and *Cryptosporidium* spp (1·5, 0·8–2·2). In older children, *Shigella* or EIEC was the dominant pathogen (2·3, 1·2–3·4). By qPCR, 89·3% (95% CI 83·2–96·0) of diarrhoeal episodes at the population level were attributable to a pathogen, as compared with 51·5% (48·0–55·0) by the original workup. *Shigella* spp, rotavirus, adenovirus 40/41, heat-stable enterotoxin-producing *E coli, Cryptosporidium* spp, and *Campylobacter* spp together accounted for 77·8% (74·6–80·9) of all attributable diarrhoea.

To further understand the increase in burden attributable to *Shigella* spp or EIEC, we examined the qPCR and culture results by site. Attribution was higher at all study sites with qPCR than with the previous diagnostic methods, but to a lesser degree in Bangladesh than in the others, which was due mainly to high quantities of *Shigella* spp and high sensitivity of the culture at that site (appendix).

Among analysed diarrhoeal episodes, 4077 (76·9%) of 5304 were described as being watery, and gross blood was reported as being seen in 1227 (23·1%). As expected, *Shigella* spp or EIEC was the dominant pathogen associated with dysentery (attributable fraction 63·8%, 95% CI 61·6–68·0), but was also associated with the second-highest burden of watery diarrhoea (12·9%, 11·0–16·4, figure 3), and, overall, 527·7 (40·3%) of 1310·3 attributable cases of diarrhoea attributable to *Shigella* spp or EIEC were non-dysenteric. Generally, quantities of *Shigella* spp or EIEC in dysentery cases were higher than those in cases of watery diarrhoea (mean Cq 22·0 [4·6] *vs* 26·2 [5·8], p<0·0001; appendix). Confirmation of *Shigella* spp or EIEC detection in 450 case-control pairs showed similar attributable fractions with different PCR targets (appendix). Additionally, qPCR for gene regions specific to *S flexneri* and *S sonnei* showed these species account for more than 70% of all *Shigella* spp or EIEC attributable fractions. Adenovirus 40/41 and *C jejuni* or *C coli* were also associated with a substantial burden of dysentery (attributable fractions 10·4%, 95% CI 4·5–16·7 and 7·8%, 5·5–12·9, respectively; these estimates control for the presence of other pathogens, such as *Shigella* spp).

The original GEMS microbiology interrogated for ST-ETEC producing the STh subtype (STh-ETEC) and heat-labile enterotoxin-producing *E coli* (LT-ETEC), but not ST-ETEC producing the STp subtype, which might have led to underestimation of the burden for ST-ETEC. qPCR showed that most of the burden was attributable to the STh subtype and was mainly evident at specific study sites (appendix). Inclusion of STp subtype in the definition of ST-ETEC increased the overall burden estimates for this pathogen by 15%. There was essentially no burden attributable to LT-ETEC, and the presence of heat-labile enterotoxin did not modify the strength of the association between ST-ETEC producing the STh subtype

and case or control status (with heat-labile enterotoxin, OR 0·92, 95% CI 0·74–1·13). Therefore, to be consistent with the original study, we categorised ST-ETEC producing the STh subtype with or without heat-labile enterotoxin. In 249 (43·9%) of 505 samples in which ST-ETEC producing STh subtype was detected in diarrhoea-associated quantities, heat-labile enterotoxin was detected in similarly high quantities, which suggests that these infections might be due to both toxins and the remainder were due to ST-ETEC producing STh subtype only.

To examine the causes of diarrhoea in individual children, we identified model-derived quantitative cutoffs for diarrhoea association (figure 1, appendix). Diarrhoea-associated quantities varied greatly by pathogen, ranging from $7{\cdot}3 \times 10^3$ copies per g of stool for *V cholerae* to $2{\cdot}7 \times 10^9$ copies per g for tEPEC. For the subset of 15 pathogens for which cutoff values could be identified, at least one was detected at a diarrhoea-associated quantity in 4317 (81·4%) of 5304 cases (one in 2254 [42·5%] and more than one in 2063 [38·9%]; figure 4). Combining all study sites and age strata, *Shigella* spp or EIEC was the most frequent cause, followed by adenovirus 40/41, rotavirus, *Cryptosporidium* spp, and ST-ETEC (figure 4). In 1517 (79·6%) of 1905 cases in which *Shigella* spp or EIEC and 984 (84·2%) of 1168 in which rotavirus was present in diarrhoea-associated quantities, these were the primary pathogens (ie, the only diarrhoea-associated pathogen or at a quantity that delivered the highest diarrhoea association). The predominance of *Cryptosporidium* spp and ST-ETEC was slightly less (313 [58·0%] of 540 and 332 [65·6%] of 506 cases, respectively). The opposite was seen for adenovirus 40/41, which in most cases (1070 [71·4%] of 1499), even at diarrhoea-associated quantities, occurred with pathogens more strongly associated with diarrhoea (appendix). In general, we saw no preferential co-occurrence between particular diarrhoea-associated pathogens, although *Shigella* spp or EIEC was frequently identified as a diarrhoea-associated co-infection when *H pylori* was detected as the primary diarrhoea-associated pathogen (appendix). In many samples, quantities of *C jejuni* or *C coli*, tEPEC, *H pylori*, *Cryptosporidium* spp, ST-ETEC, and norovirus GII were low and the pathogens were not associated with diarrhoea.

Discussion

In this reanalysis of the causes of diarrhoea in GEMS, we have shown with a quantitative molecular diagnostic approach that the estimated burdens of *Shigella* spp, adenovirus 40/41, ST-ETEC, and *C jejuni* or *C coli* were underestimated with the original diagnostic methods. By contrast, estimates for rotavirus and *Cryptosporidium* spp, for which detection was based on EIA diagnostics, changed little. These six pathogen groups were responsible for 77·8% of attributable diarrhoea in this study and, therefore, targeted interventions could have larger public health benefits than previously projected.[4]

The increases in estimated burden were a function of improved sensitivity with molecular diagnostics and the higher resolution provided by pathogen quantification. Previous studies of the causes of diarrhoea have generally used non-quantitative diagnostics, which yield dichotomous results at detection limits that might not be clinically relevant. Such results become potentially problematic for the study of causes of diarrhoea in children in resource-limited settings because the rate of enteropathogen carriage shortly after birth is high.[7,21]

We identified a high burden of diarrhoea attributable to *Shigella* spp or EIEC, which we suspect is due to the fastidious nature of this pathogen in culture, particularly when quantities are low and in areas where antibiotics are frequently used. Underdetection of *Shigella* spp by culture compared with PCR has been seen in various settings worldwide.[8,22–25] The *ipaH* target has been widely used for detection of *Shigella* spp, and its presence has only been described in *Shigella* spp and EIEC, both of which genera are highly similar at the DNA level.[26] Nevertheless, EIEC has not been a major diarrhoea-associated pathogen in similar settings.[7] The detection of additional gene regions clarified that *Shigella* spp explained most of the burden estimate. *Shigella* spp were particularly associated with dysentery, which was a case definition inclusion criterion in GEMS and was common at the Bangladesh site, but were also notably associated with watery diarrhoea across all study sites, which raises questions about the treatment guidelines for infection with *Shigella* spp, which emphasise the presence of blood in stools as an indication.[27] The higher quantities of *Shigella* spp in samples from children who had dysentery than in those who had watery diarrhoea without a diagnosis of dysentery supports previous findings[26] and suggests that watery diarrhoea might indicate a less advanced stage of this disease. Development of a vaccine against *Shigella* spp is underway, and these findings and the high detection rate in samples in this study suggest that it should be prioritised. Vaccination of children early in life could be beneficial.

The importance of rotavirus and *Cryptosporidium* spp was reaffirmed, with burden estimates being similar to those in the original study, whereas the importance of ST-ETEC and *Campylobacter* spp was increased. The association of adenovirus 40/41 with diarrhoea has been noted previously,[28] including by use of EIA in the original GEMS study,[6] but molecular diagnostics revealed a substantially higher prevalence and retained diarrhoea association. Our array card for qPCR did detect panadenovirus of any serotype, but this was not associated with diarrhoea after controlling for the presence of adenovirus 40/41. The association of *H pylori* with diarrhoea was unexpected. This pathogen has been reported in children with diarrhoea, most often those with chronic diarrhoea or hypochlorhydria,[29,30] but other studies have identified no risk or suggested protection.[31,32] We found frequent co-infection with *H pylori* and *Shigella* spp, which has been reported previously[33] and suggests that infection with *H pylori* could be a risk factor for infection with *Shigella* spp. Further study is needed to understand the role of *H pylori* in diarrhoea.

By identifying model-derived quantitative cutoff values for detection of diarrhoea-associated pathogens, we were able to describe the causes of diarrhoea in individual children, which is crucial for clinical diagnosis or case definitions. Quantitative cutoffs have previously been described for tEPEC, norovirus GII, and rotavirus,[9–14] mostly in comparisons of quantities in cases versus controls, but those studies involved fewer samples and did not adjust for other pathogens or provide the resolution afforded by the present analytical methods. We were able to identify cutoff values for 15 pathogens. This approach worked well for most pathogens but less so for others. For example, *Campylobacter* spp had weak but significant associations with diarrhoea that yielded poorly discriminatory quantitative cutoff values despite a substantial burden in the overall population due to the high prevalence of this group of pathogens. The cutoffs associated with diarrhoea in this study varied by pathogen. The *ipaH* Cq for detection of *Shigella* spp that maximally separated cases and controls was

previously estimated from a subset of GEMS sites of approximately 10^7 copies per g of stool.[8] We found a slightly lower value of 2×10^6 copies per g. The quantitative relations we identified were reasonably generalisable across age groups for the pathogens with the greatest attributions and, therefore, these cutoff values should be useful for future studies when control stools are not available.

More than 80% of samples revealed pathogens at diarrhoea-associated quantities, which is a much greater diagnostic yield than has been found with routine microbiological methods.[4,34,35] Among those containing a diarrhoea-associated pathogen, roughly half had more than one. This extent of mixed infections, even after filtering out probable asymptomatic infections, was substantial. Some pathogens, such as *Shigella* spp and rotavirus, were most frequently detected as sole pathogens in diarrhoea-associated quantities, meaning that misattribution of diarrhoea to these pathogens would be uncommon. Many other pathogens, such as *Cryptosporidium* spp, ST-ETEC, norovirus GII, adenovirus 40/41, and *Campylobacter* spp, were not detected in diarrhoea-associated quantities or present in quantities associated with diarrhoea but with other pathogens that showed stronger associations (figure 4). This finding underscores the difficulty in attributing diarrhoea to such pathogens without quantification or considering other pathogens. It will also challenge vaccine trials, because efficacy would be underestimated if a non-quantitative PCR result were used that might over-attribute diarrhoea to the pathogen of interest. For such trials, more stringent quantitative cutoffs could be applied, for example using the quantity of the highly diarrhoea-associated ST-ETEC to improve the specificity of the case definition for ETEC-associated diarrhoea (appendix). Other criteria, such as clinical severity scores or inflammatory markers, could be assessed to see whether they increase the specificity of detection.

Achieving optimum clinical management in the era of molecular diagnostic panels for enteropathogens is a work in progress. Commonly available commercial panels are qualitative PCR-based tests that do not distinguish between low-level enteropathogen detection of unclear relevance and infections that are more clearly clinically important.[36] For example, detection of the *Clostridium difficile* toxin gene by PCR in the setting of a negative toxin EIA is not clinically relevant.[37] We believe that the quantitative cutoff concept we have assessed will be helpful, although it places additional requirements on commercial assay developers, and the exact cutoffs used might depend on the clinical setting. Management guidelines for acute diarrhoea recommend supportive care with rehydration and zinc supplementation without antibiotics unless signs are indicative of dysentery or cholera is suspected,[27] but could be adapted as pathogen-specific diagnosis improves and hopefully becomes more widely available. For a high-mortality entity, such as childhood diarrhoea, an overly stringent cutoff could lead to underdiagnosis of treatable causes. Thus further assessment of quantitative cutoff values for clinical care is needed, along with studies that assess therapeutic response, for instance whether treatment of diarrhoea associated with *Shigella* spp in children in low-resource settings that is negative to culture but positive on PCR improves outcomes. The necessity of treating multiple pathogens during mixed infection is also unknown. Finally, the existing quantitative PCR assays for enteropathogens do not assess antimicrobial resistance and, therefore, for now these emerging molecular technologies should ideally be used in conjunction with conventional culture methods.

This study had some limitations. The quantitative approach, while incrementally useful, will inherently function less well for pathogens that are shed with high frequency, in high quantities, and for an extended duration in the absence of diarrhoea. Longitudinal studies might further refine understanding for such pathogens. Theoretically, the sum of the attributable fractions for individual pathogens can be greater than 100%, although this possibility is the same for estimates derived from molecular and non-molecular diagnostic approaches, and is implicit in any clinical entity with co-infections. Although GEMS was carefully designed to study the causes of diarrhoea and the study sites were broadly representative of countries with moderate to high mortality in children younger than 5 years, some heterogeneity in the hierarchy of pathogens was noted between sites. Generalisability, therefore, should be audited by local epidemiology. Finally, vaccine development often relies on speciation and subtyping of infections,[38] and our molecular assays did not provide such information.

This reanalysis of the causes of diarrhoea in young children by use of molecular diagnostic methods substantially increased attribution at the population level and clinical diagnostic yields. We found a high burden of diarrhoea associated with six main pathogens: *Shigella* spp, rotavirus, adenovirus 40/41, ST-ETEC, *Cryptosporidium* spp, and *Campylobacter* spp. Mixed infection with diarrhoea-associated pathogens was common and has implications for vaccine efficacy trials, responses to treatments, and the clinical diagnosis of pathogen-specific diarrhoea.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the families who participated, the project field staff, and administration staff at all sites for assistance. We thank Eileen Johnson, Haochen Li, and Jixian Zhang, University of Virginia, Charlottesville, VA, USA, for testing specimens. This work was supported by the Bill and Melinda Gates Foundation (OPP1019093).

References

1. GBD 2013 DALYs and HALE Collaborators. Global, regional, and national disability-adjusted life years (DALYs) for 306 diseases and injuries and healthy life expectancy (HALE) for 188 countries, 1990–2013: quantifying the epidemiological transition. Lancet. 2015; 386:2145–91. [PubMed: 26321261]
2. Liu L, Johnson HL, Cousens S, et al. Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. Lancet. 2012; 379:2151–61. [PubMed: 22579125]
3. GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet. 2015; 385:117–71. [PubMed: 25530442]
4. Lanata CF, Fischer-Walker CL, Olascoaga AC, et al. Global causes of diarrheal disease mortality in children <5 years of age: a systematic review. PLoS One. 2013; 8:e72788. [PubMed: 24023773]
5. Huilan S, Zhen LG, Mathan MM, et al. Etiology of acute diarrhoea among children in developing countries: a multicentre study in five countries. Bull World Health Organ. 1991; 69:549–55. [PubMed: 1659953]

6. Kotloff KL, Nataro JP, Blackwelder WC, et al. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. Lancet. 2013; 382:209–22. [PubMed: 23680352]
7. Platts-Mills JA, Babji S, Bodhidatta L, et al. Pathogen-specific burdens of community diarrhoea in developing countries: a multisite birth cohort study (MAL-ED). Lancet Glob Health. 2015; 3:e564–75. [PubMed: 26202075]
8. Lindsay B, Ochieng JB, Ikumapayi UN, et al. Quantitative PCR for detection of *Shigella* improves ascertainment of *Shigell*a burden in children with moderate-to-severe diarrhoea in low-income countries. J Clin Microbiol. 2013; 51:1740–46. [PubMed: 23536399]
9. Phillips G, Tam CC, Conti S, et al. Community incidence of norovirus-associated infectious intestinal disease in England: improved estimates using viral load for norovirus diagnosis. Am J Epidemiol. 2010; 171:1014–22. [PubMed: 20360244]
10. Barletta F, Ochoa TJ, Mercado E, et al. Quantitative real-time polymerase chain reaction for enteropathogenic *Escherichia coli*: a tool for investigation of asymptomatic versus symptomatic infections. Clin Infect Dis. 2011; 53:1223–29. [PubMed: 22028433]
11. Mukhopadhya I, Sarkar R, Menon VK, et al. Rotavirus shedding in symptomatic and asymptomatic children using reverse transcription-quantitative PCR. J Med Virol. 2013; 85:1661–68. [PubMed: 23775335]
12. Kaplon J, Fremy C, Pillet S, et al. Diagnostic accuracy of seven commercial assays for rapid detection of group A rotavirus antigens. J Clin Microbiol. 2015; 53:3670–73. [PubMed: 26378280]
13. Dung TT, Phat VV, Nga TV, et al. The validation and utility of a quantitative one-step multiplex RT real-time PCR targeting rotavirus A and norovirus. J Virol Methods. 2013; 187:138–43. [PubMed: 23046990]
14. Barreira DM, Ferreira MS, Fumian TM, et al. Viral load and genotypes of noroviruses in symptomatic and asymptomatic children in Southeastern Brazil. J Clin Virol. 2010; 47:60–64. [PubMed: 20004146]
15. Liu J, Kabir F, Manneh J, et al. Development and assessment of molecular diagnostic tests for 15 enteropathogens causing childhood diarrhoea: a multicentre study. Lancet Infect Dis. 2014; 14:716–24. [PubMed: 25022434]
16. Kotloff KL, Blackwelder WC, Nasrin D, et al. The Global Enteric Multicenter Study (GEMS) of diarrheal disease in infants and young children in developing countries: epidemiologic and clinical methods of the case/control study. Clin Infect Dis. 2012; 55(suppl 4):S232–45. [PubMed: 23169936]
17. Liu J, Gratz J, Amour C, et al. A laboratory-developed TaqMan Array Card for simultaneous detection of 19 enteropathogens. J Clin Microbiol. 2013; 51:472–80. [PubMed: 23175269]
18. Liu J, Gratz J, Amour C, et al. Optimization of quantitative PCR methods for enteropathogen detection. PLoS One. 2016; 11:e0158199. [PubMed: 27336160]
19. Blackwelder WC, Biswas K, Wu Y, et al. Statistical methods in the Global Enteric Multicenter Study (GEMS). Clin Infect Dis. 2012; 55(suppl 4):S246–53. [PubMed: 23169937]
20. Bruzzi P, Green SB, Byar DP, Brinton LA, Schairer C. Estimating the population attributable risk for multiple risk factors using case-control data. Am J Epidemiol. 1985; 122:904–14. [PubMed: 4050778]
21. Taniuchi M, Sobuz SU, Begum S, et al. Etiology of diarrhoea in Bangladeshi infants in the first year of life analyzed using molecular methods. J Infect Dis. 2013; 208:1794–802. [PubMed: 24041797]
22. von Seidlein L, Kim DR, Ali M, et al. A multicentre study of *Shigella* diarrhoea in six Asian countries: disease burden, clinical manifestations, and microbiology. PLoS Med. 2006; 3:e353. [PubMed: 16968124]
23. Buss SN, Leber A, Chapin K, et al. Multucenter evaluation of the BioFire Filmarray Gastrointestinal panel for etiologic diagnosis of infectious gastroenteritis. J Clin Microbiol. 2015; 53:915–25. [PubMed: 25588652]

24. Duong VT, Phat VV, Tuyen HT, et al. Evaluation of Luminex xTAG gastrointestinal pathogen panel assay for detection of multiple diarrheal pathogens in fecal samples in Vietnam. J Clin Microbiol. 2016; 54:1094–100. [PubMed: 26865681]
25. Harrington SM, Buchan BW, Doern C, et al. Multicenter evaluation of the BD Max Enteric Bacterial Panel PCR assay for rapid detection of *Salmonella* spp., *Shigella* spp., *Campylobacter* spp., and Shiga Toxin 1 and 2 gens. J Clin Microbiol. 2015; 53:1639–47. [PubMed: 25740779]
26. Vu DT, Sethabutr O, Von Seidlein L, et al. Detection of Shigella by a PCR assay targeting the *ipaH* gene suggests increased prevalence of shigellosis in Nha Trang, Vietnam. J Clin Microbiol. 2004; 42:2031–35. [PubMed: 15131166]
27. WHO. The treatment of diarrhoea: a manual for physicians and other senior health workers. Geneva: World Health Organization; 2005. p. 44
28. Kotloff KL, Losonsky GA, Morris JG Jr, Wasserman SS, Singh-Naz N, Levine MM. Enteric adenovirus infection and childhood diarrhoea: an epidemiologic study in three clinical settings. Pediatrics. 1989; 84:219–25. [PubMed: 2546121]
29. Clemens J, Albert MJ, Rao M, et al. Impact of infection by Helicobacter pylori on the risk and severity of endemic cholera. J Infect Dis. 1995; 171:1653–56. [PubMed: 7769312]
30. Sullivan PB, Thomas JE, Wight DG, et al. *Helicobacter pylori* in Gambian children with chronic diarrhoea and malnutrition. Arch Dis Child. 1990; 65:189–91. [PubMed: 2317065]
31. Isenbarger DW, Bodhidatta L, Hoge CW, et al. Prospective study of the incidence of diarrheal disease and *Helicobacter pylori* infection among children in an orphanage in Thailand. Am J Trop Med Hyg. 1998; 59:796–800. [PubMed: 9840601]
32. Graham DY, Opekun AR, Osato MS, et al. Challenge model for *Helicobacter pylori* infection in human volunteers. Gut. 2004; 53:1235–43. [PubMed: 15306577]
33. Shmuely H, Samra Z, Ashkenazi S, et al. Association of *Helicobacter pylori* infection with Shigella gastroenteritis in young children. Am J Gastroenterol. 2004; 99:2041–45. [PubMed: 15447770]
34. Guerrant RL, Van Gilder T, Steiner TS, et al. Practice guidelines for the management of infectious diarrhoea. Clin Infect Dis. 2001; 32:331–51. [PubMed: 11170940]
35. Koplan JP, Fineberg HV, Ferraro MJ, Rosenberg ML. Value of stool cultures. Lancet. 1980; 2:413–16. [PubMed: 6105529]
36. Spina A, Kerr KG, Cormican M, et al. Spectrum of enteropathogens detected by the FilmArray GI Panel in a multicentre study of community-acquired gastroenteritis. Clin Microbiol Infect. 2015; 21:719–28. [PubMed: 25908431]
37. Polage CR, Gyorke CE, Kennedy MA, et al. Overdiagnosis of *Clostridium difficile* infection in the molecular test era. JAMA Intern Med. 2015; 175:1792–1801. [PubMed: 26348734]
38. Livio S, Strockbine NA, Panchalingam S, et al. *Shigella* isolates from the global enteric multicenter study inform vaccine development. Clin Infect Dis. 2014; 59:933–41. [PubMed: 24958238]

Research in context

Evidence before this study

We reviewed relevant studies on the cause of diarrhoea, identified by searching PubMed with the terms “PCR” and “diarrhoea/diarrhea”. We did not restrict our search by language, date, or age group. We identified 4258 reports, of which only 216 described case-control studies (to control for asymptomatic carriage) and 17 that assessed pathogen quantity. Quantitative case-control analyses on rotavirus, norovirus GII, diarrhoeagenic *Escherichia coli*, and *Shigella* spp were done in six studies, but either involved only one pathogen or a panel of enteropathogens at only one study site.

Added value of this study

This study of diarrhoea aetiology used quantitative molecular diagnostics for 32 enteropathogens in multiple study sites, with cases and controls. Through this approach we were able to reveal each pathogen's relative importance at the population level and in individual children.

Implications of all the available evidence

Use of quantitative real-time PCR led to revised estimates of global diarrhoeal burden. The most attributable pathogens were, in descending order, *Shigella* spp, rotavirus, adenovirus 40/41, heat-stable enterotoxin-producing *E coli*, *Cryptosporidium* spp, and *Campylobacter* spp, which together accounted for 78% of all attributable diarrhoea. Interventions targeting these pathogens should be prioritised.

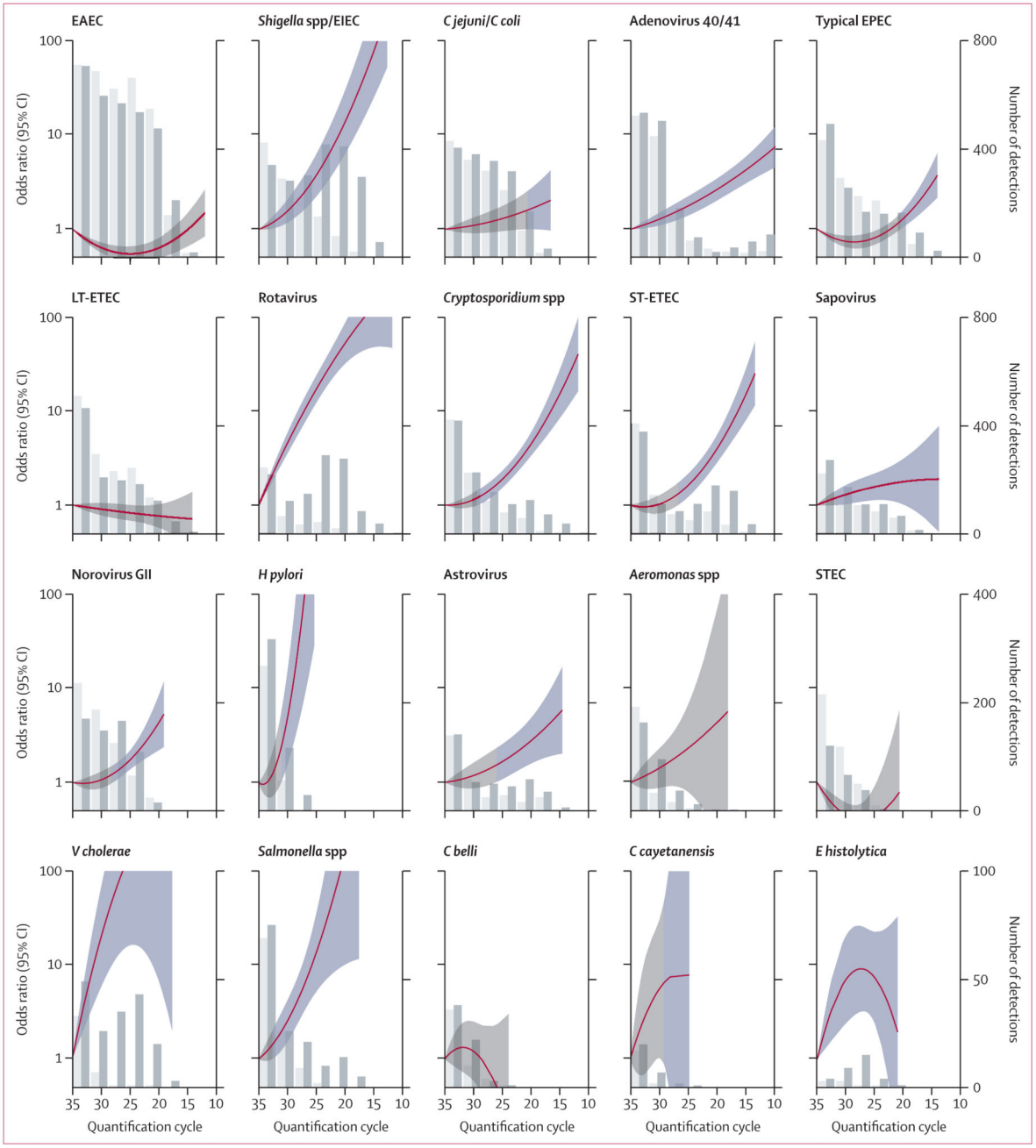

Figure 1. Relation between pathogen quantity and diarrhoea

Pathogens are ordered from top to bottom and left to right by prevalence in cases. Data are numbers of dectections (vertical bars) in cases (dark grey) and controls (light grey) with odds ratios (red lines) and 95% CIs (bands). The model-derived cutoffs used for identification of diarrhoea-associated pathogens in individuals (overlaid in blue) are defined as all detections above the point at which the 95% CI no longer includes 1. EAEC=enteroaggregative *E coli*. EIEC=enteroinvasive *E coli*. EPEC=enteropathogenic *E coli*. LT-ETEC=heat-labile enterotoxin-producing *E coli*. ST-ETEC=STh-producing enterotoxigenic *E coli*. STEC=Shiga toxin producing *E coli*.

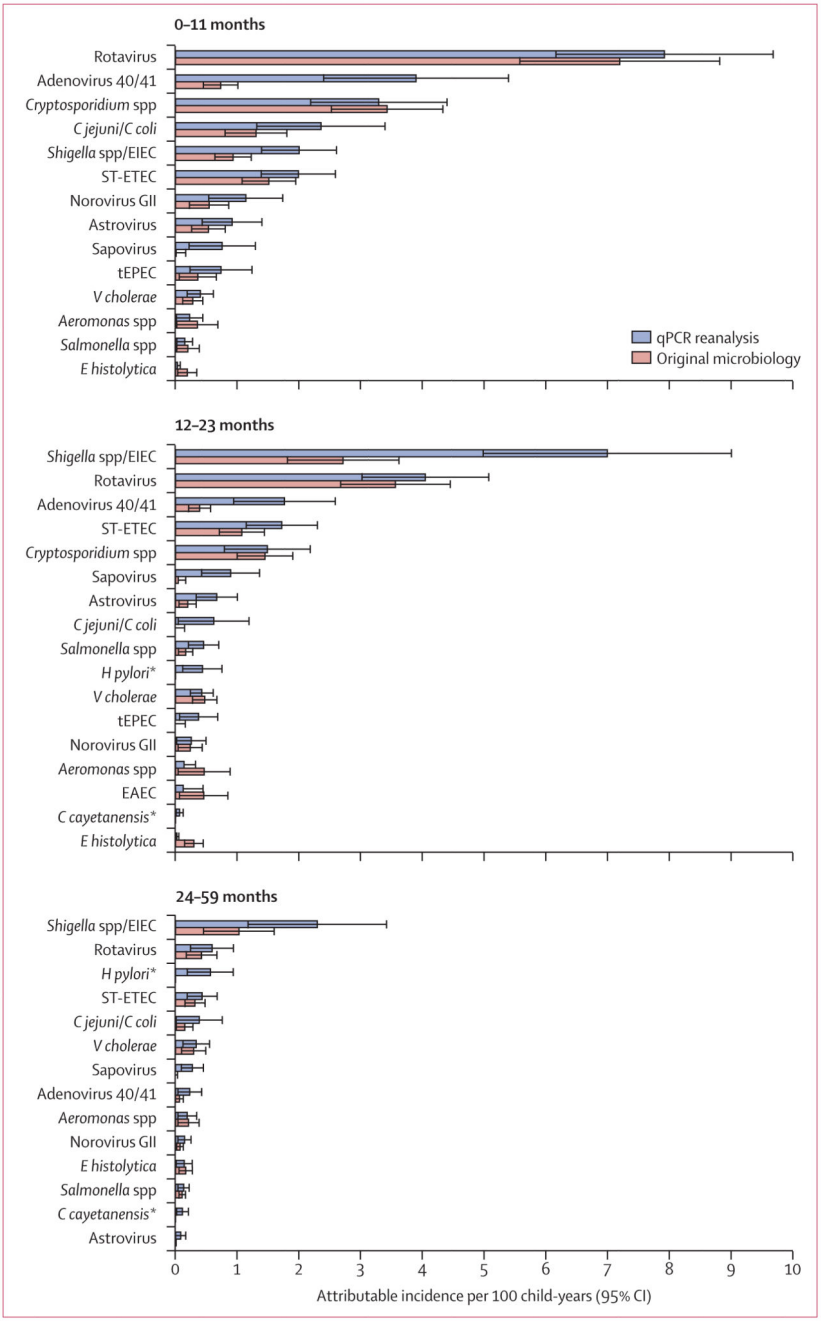

Figure 2. Attributable incidence of pathogen-specific moderate to severe diarrhoea per 100 child-years, by age stratum, across study sites, in this and the original GEMS

For each age stratum, any pathogens significantly associated with diarrhoea by one or both diagnostic approaches are shown. GEMS=Global Enteric Multicenter Study. EAEC=enteroaggregative *E coli*. EIEC=enteroinvasive *E coli*. tEPEC=typical enteropathogenic *E coli*. ST-ETEC=STh-producing enterotoxigenic *E coli*. qPCR=quantitative realtime PCR. *Indicates the microbiology in the original GEMS did not test for *H pylori* or *C cayetanensis*.

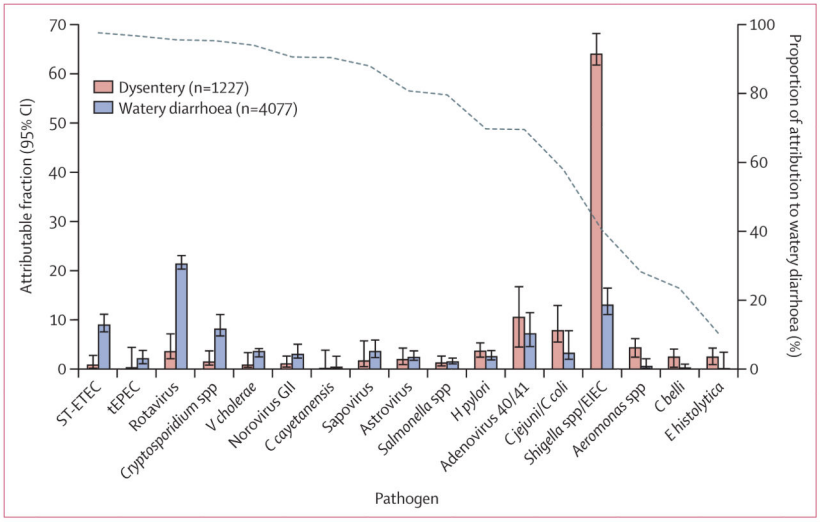

Figure 3. Relative attribution of watery diarrhoea and dysentery to each pathogen

Data are overall adjusted attributable fractions (vertical bars) with 95% CIs. Pathogens are ordered by the proportion of total attributable cases that were watery diarrhoea (dotted line). All pathogens significantly associated with either dysentery or watery diarrhoea are shown. ST-ETEC=STh-producing enterotoxigenic *E coli*. tEPEC=typical enteropathogenic *E coli*. EIEC=enteroinvasive *E coli*.

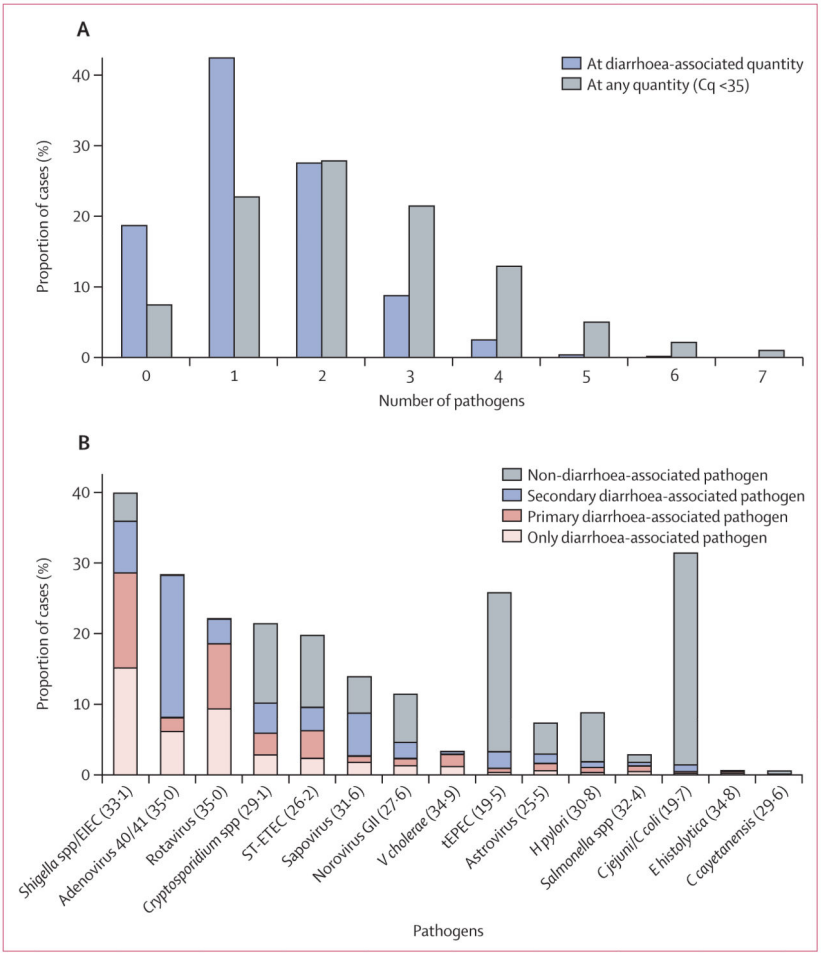

Figure 4. Detection of co-infections in diarrhoeal cases

(A) Numbers of pathogens at diarrhoea-associated quantities and any quantity in individual cases of diarrhoea. (B) Distribution of pathogens, alone and in co-infections, by quantity and association with diarrhoea. The quantification cycle cutoff used to identify diarrhoea-associated detections is shown in parentheses after each pathogen name. Cq=quantification cycle. EIEC=enteroinvasive *E coli*. ST-ETEC=STh-producing enterotoxigenic *Escherichia coli*. tEPEC=typical enterpathogenic *E coli*.

Table 1

Paired case and control samples tested by quantitative real-time PCR, by age stratum and study site

	Basse, The Gambia*	Bamako, Mali†	Manhiça, Mozambique†	Nyanza Province, Kenya*	Kolkata, India†	Mirzapur, Bangladesh*	Karachi (Bin Qasim Town), Pakistan*	Total
Age stratum‡								
0–11 months	300 (400)	300 (727)	267 (374)	303 (673)	300 (672)	300 (550)	300 (633)	2070
12–23 months	301 (455)	300 (682)	159 (195)	300 (410)	300 (588)	300 (476)	300 (399)	1960
24–59 months	165 (174)	283 (624)	95 (112)	301 (393)	300 (308)	300 (368)	226 (226)	1670
Valid pairs analysed§	685 (894%)	834 (94·5%)	484 (92·9%)	787 (87·1%)	849 (94·3%)	877 (974%)	788 (954%)	5304 (93·1%)
Reproducibility (%)¶								
MS2	6·0%	5·2%	6·0%	6·8%	5·9%	6·2%	5·8%	67%
PhHV	5·2%	4·9%	4·1%	5·5%	5·0%	54%	6·1%	5·5%

For each combination of age and site, we aimed to test 300 cases and first available matched controls; for sites with fewer than 300 cases enrolled, all available cases and first available matched controls were tested. PhHV=phocine herpes virus.

*Extracted and tested on site

†Extracted at University of Maryland, MD, USA, and tested at University of Virginia, VA, USA.

‡Numbers in parentheses are the number of cases in the original Global Enteric Multicenter Study.

§Percentage of case-control pairs that were included in the analysis and did not exhibit extraction blank failures (139 pairs excluded, ranging from 0 to 7·0% by site), amplification failures (220 pairs excluded, ranging from 2·0% to 5·3% by site), or quantitative real-time PCR software flags (77 pairs excluded, ranging from 0·2% to 2·7% by site).

¶Measured by the coefficient variation of MS2 or PhHV Cqs on the clinical specimens.

Table 2

Adjusted attributable fractions for pathogens causing moderate to severe diarrhoea assessed with quantitative molecular diagnostic methods

	Basse, The Gambia	Bamako, Mali	Manhiça, Mozambique	Nyanza Province, Kenya	Kolkata, India	Mirzapur, Bangladesh	Karachi (Bin Qasim Town), Pakistan
0–11 months							
Number of case-control	258	285	247	267	275	293	284
Rotavirus	23·0% (18·8–29·1)	22·7% (18·5–28·3)	34·5% (29·4–41·1)	22·4% (17·6–27·9)	30·7% (26·1–37·2)	23·4% (18·9–28·7)	27·0% (22·7–33·5)
Adenovirus 40/41	9·6% (4·9–15·9)	10·8% (6·8–20·9)	13·3% (8·1–24·1)	8·5% (4·7–13·3)	22·4% (7·5–33·8)	8·3% (47–13·9)	7·2% (0·9–10·6)
Cryptosporidium spp	11·5% (4·6–16)	17·2% (13·1–26·4)	14·7% (11·3–24·1)	9·9% (4·9–15·5)	5·4% (2·8–9·0)	1·2% (0·1–2·7)	10·4% (5·1–16·1)
Shigella spp or EIEC	7·5% (3·9–12·9)	4·0% (1·6–7·2)	4·0% (1·6–7·0)	6·1% (2·8–11·1)	3·9% (1·5–7·5)	15·8% (12–20·4)	13·8% (97–197)
C jejuni or *C coli*	*	8·1% (3·3–14·3)	11·2% (3·0–20·0)	6·1% (0·1–137)	9·3% (2·4–16·8)	12·3% (7·6–19·8)	7·0% (0–17·1)
ST-ETEC	5·5% (1·8–9·7)	5·0% (2·5–8·4)	9·3% (5·8–14·3)	8·8% (5·4–13·6)	4·6% (2·4–8·3)	2·0% (0·5–3·9)	12·2% (8·9–17·6)
Norovirus GII	4·4% (1·8–11·1)	4·2% (0·4–9)	3·8% (1·7–9·2)	3·1% (0·8–7·8)	4·2% (1·3–9·5)	1·8% (0·2–5·3)	2·7% (0–6·2)
tEPEC	2·7% (0–5·7)	3·4% (0·9–8·7)	3·4% (1·5–10·5)	2·2% (0·5–9·9)	1·8% (0·1–5·1)	0·3% (0–2·6)	2·1% (0–5·1)
Sapovirus	1·0% (0–5·4)	1·6% (0–6·4)	0·5% (0–2·6)	2·4% (0–7·2)	37% (1·4–8·7)	1·1% (0–3·4)	4·4% (0·2–11·1)
Astrovirus	0·7% (0–4·4)	2·8% (0–6·4)	0·2% (0–1·2)	1·8% (0·2–4·5)	5·5% (2·6–10·2)	1·4% (0–3·9)	3·3% (0·1–7·6)
V cholerae	*	0·3% (0–1·1)	0·4% (0–1·2)	0·7% (0–1·9)	1·6% (0·3–3·1)	0·6% (0–1·7)	5·6% (2·3–8·8)
H pylori	0·6% (0–2·8)	0·8% (0–4·3)	1·1% (0–2·8)	0·5% (0–1·5)	0·9% (0–3·6)	0·2% (0–1·4)	*
Aeromonas spp	*	0·2% (0–1·3)	0·8% (0–2·5)	*	0·9% (0–4·5)	1·4% (0–4·2)	2·9% (0·1–8·2)
Salmonella spp	0·7% (0–3)	0·7% (0–2·1)	1·3% (0–2·7)	0·7% (0–3·1)	0·1% (0–0·8)	0·4% (0–3·3)	0·2% (0–0·9)
C cayetanensis	*	0·3% (0–1·1)	*	0·3% (0–1·3)	*	0·2% (0–1·0)	0·3% (0–1·1)
C belli	*	0·1% (0–0·6)	0·7% (0–2·4)	*	0·3% (0–1·2)	*	*
EAEC	*	*	0·1% (0–8·1)	*	*	*	*
E histolytica	*	0·4% (0·3–1·2)	*	*	*	*	*
12–23 months							
Number of case-control	273	274	151	264	289	291	293
Shigella spp or EIEC	32·7% (27·0–42·0)	27·1% (20·5–35·9)	34·8% (27·5–44·8)	15·4% (9·8–23·4)	19·2% (14·1–26·9)	64·9% (60·4–71·5)	35·0% (30·0–43·3)
Rotavirus	17·1% (13·0–21·9)	11·4% (7·9–15·2)	14·8% (10·0–20·7)	14·5% (10·7–19·0)	28·9% (24·4–347)	29·8% (25·4–357)	9·4% (6·5–13·1)
ST-ETEC	6·3% (2·2–10·8)	6·6% (3·2–10·9)	14·0% (8·7–22·6)	9·5% (6·0–15·2)	7·8% (4·8–12·8)	2·2% (0·4–4·4)	12·1% (8·2–18·5)
Adenovirus 40/41	8·4% (2·8–15·9)	9·0% (4·0–19·4)	11·3% (4·8–23·4)	4·1% (1·9–7·2)	10·3% (3·6–16·8)	4·1% (1·8–8·5)	5·8% (0·1–9·3)

	Basse, The Gambia	Bamako, Mali	Manhiça, Mozambique	Nyanza Province, Kenya	Kolkata, India	Mirzapur, Bangladesh	Karachi (Bin Qasim Town), Pakistan
Cryptosporidium spp	6·9% (1·0–12·0)	7·7% (4·0–16·7)	12·6% (8·3–247)	6·3% (2·1–11·7)	5·6% (2·9–9·9)	1·3% (0·2–3·0)	7·5% (3·0–13·7)
Sapovirus	2·1% (0–6·6)	2·9% (0–7·4)	2·2% (0–6·7)	4·2% (0·8–9·3)	6·7% (3·3–12·7)	2·5% (0·1–5·9)	6·3% (1·2–13·9)
Astrovirus	0·9% (0–2·8)	3·2% (0·8–6·8)	0·6% (0–2·4)	2·5% (0·8–4·8)	4·1% (2·0–7·3)	1·9% (0·3–4·0)	5·5% (2·0–10·2)
C jejuni or *C coli*	*	5·0% (0–14·0)	2·6% (0–10·7)	0·1% (0–6·8)	3·6% (0–15·2)	4·5% (1·6–9·6)	*
V cholerae	*	0·3% (0–1·1)	2·8% (0·6–5·8)	0·7% (0–1·8)	4·2% (2·1–7·0)	0·4% (0–1·3)	7·9% (5·2–12·0)
H pylori	3·1% (0·8–6·9)	2·3% (0·4–6·7)	*	2·3% (0·2–6·5)	1·3% (0·3–3·3)	1·2% (0·1–4)	0·5% (0–2·4)
tEPEC	2·6% (0–6·1)	1·9% (0·3–6·5)	3·1% (0·9–11·4)	1·4% (0·3–5·4)	1·1% (0–3·9)	0·4% (0–2·0)	1·3% (0–4·6)
Salmonella spp	3·9% (1·0–7·1)	2·5% (0·9–4·8)	2·8% (0·1–57)	2·3% (0·7–4·3)	0·2% (0–0·9)	1·2% (0·3–3·2)	0·4% (0–1·2)
Norovirus GII	1·2% (0·2–5)	0·5% (0–2·9)	0·9% (0–4·9)	0·9% (0–3·8)	2·4% (0·2–8·4)	1·1% (0·2–3·9)	0·9% (0–3·5)
Aeromonas spp	*	0·2% (0–1·7)	1·4% (0–6·5)	*	0·3% (0–4·7)	2·2% (0–6·0)	2·1% (0–7·2)
C cayetanensis	0·1% (0–0·9)	*	0·6% (0–3·1)	0·6% (0–1·8)	*	1·0% (0–2·3)	0·6% (0–2·0)
C belli	*	0·1% (0–1·7)	0·2% (0–3·9)	0·1% (0–1·5)	0·1% (0–1·9)	*	0·1% (0–0·8)
E histolytica	*	*	*	*	0·3% (0–1·0)	*	0·5% (0–1·7)
EAEC	*	0·9% (0–7·9)	4·0% (0·2–18·4)	*	*	0·1% (0–1·7)	*
24–59 months							
Number of case-control	154	275	86	256	285	293	211
Shigella spp or EIEC	25·9% (18·5–36·1)	20·1% (14·1–29·5)	41·9% (32·2–54·9)	17·8% (12·2–25·9)	30·0% (22·9–39·3)	83·2% (80·2–87·9)	29·5% (23·4–38·6)
Rotavirus	13·6% (9·0–19·9)	5·5% (3·0–8·3)	4·2% (0·6–8·9)	4·1% (1·9–6·8)	13·7% (10·2–18·4)	4·7% (2·7–7·2)	2·2% (0·5–4·3)
H pylori	6·7% (1·2–14·4)	8·0% (4·0–13·1)	4·2% (0·2–9·8)	5·1% (1·8–8·7)	11·0% (7·8–16·4)	2·0% (0·8–4·0)	2·7% (0·3–5·7)
ST-ETEC	4·9% (0·6–10·6)	4·0% (1·1–8·0)	8·9% (3·9–16·8)	8·3% (5·2–13·6)	7·4% (4·4–12·4)	1·2% (0–3·2)	8·0% (4·4–14·4)
V cholerae	0·6% (0–2·1)	0·5% (0–1·5)	10·6% (4·3–17·7)	1·5% (0·3–3·1)	9·6% (6·3–13·3)	2·8% (1·2–4·8)	16·1% (11·2–21·4)
C jejuni or *C coli*	*	9·0% (0·4–19·5)	2·2% (0–7·9)	0·6% (0–3·8)	5·4% (0–15·4)	2·0% (0·8–4·1)	0·6% (0–13·7)
Sapovirus	2·3% (0–5·4)	2·1% (0–4·8)	2·5% (0–6·2)	3·0% (0·8–6·3)	6·1% (3·1–10·8)	3·4% (0·4–6·2)	5·2% (0·9–10·3)
Adenovirus 40/41	2·7% (0–7·6)	1·5% (0·2–4·7)	2·9% (0·3–8·6)	1·6% (0·1–4·0)	5·1% (0·3–11·2)	5·3% (0·7–13·7)	2·6% (0–5·9)
Aeromonas spp	01% (0–0·6)	1·0% (0–2·9)	5·6% (1·1–12·5)	0·8% (0–4·2)	4·5% (0–9·2)	3·5% (1·7–6·8)	6·2% (2·2–12·3)
Norovirus GII	1·4% (0·2–3·7)	1·5% (0·1–3·7)	0·4% (0–2·1)	1·7% (0·4–4·0)	2·6% (0·6–5·8)	2·2% (0·6–4·8)	3·0% (0·1–6·0)
Salmonella spp	3·4% (1·1–6·9)	2·1% (0·7–4·2)	0·5% (0–2·2)	3·4% (1·6–5·9)	0·6% (0–1·6)	0·2% (0–1·0)	1·9% (0·4–4·0)
E histolytica	*	3·1% (1·1–5·7)	*	0·3% (0–1·2)	2·0% (0·6–3·7)	*	3·7% (1·2–7·1)
Cryptosporidium spp	0·7% (0–2·4)	1·0% (0–6·2)	2·1% (0–10·7)	0·6% (0–3·0)	2·0% (0–5·8)	0·1% (0–0·8)	2·3% (0–7·5)
Astrovirus	*	0·8% (0–3·0)	0·6% (0–3·3)	0·9% (0–2·7)	1·9% (0·4–4·5)	0·9% (0–3·5)	1·3% (0–4·4)

	Basse, The Gambia	Bamako, Mali	Manhiça, Mozambique	Nyanza Province, Kenya	Kolkata, India	Mirzapur, Bangladesh	Karachi (Bin Qasim Town), Pakistan
C belli	0·3% (0–3·5)	0·7% (0–4·2)	1·0% (0–6·8)	0·3% (0–2·3)	0·4% (0–3·0)	0·1% (0–0·8)	0·2% (0–1·2)
tEPEC	0·6% (0–2·3)	0·4% (0–2·6)	0·9% (0–5·6)	0·2% (0–2·4)	0·3% (0–2·0)	0·2% (0–1·0)	0·9% (0–3·0)
LT-ETEC	*	1·7% (0–11·3)	*	*	0·3% (0–5·5)	0·1% (0–3·7)	*
C cayetanensis	*	*	0·6% (0–6·9)	*	0·8% (0–3·2)	0·5% (0–2·0)	*
EAEC	*	0·6% (0–4·1)	0·4% (0–7·0)	*	*	*	*

Attributable fractions are shown as the percentage of cases for each age stratum and study site with 95% CIs, listed in descending order of overall value. All pathogens with attributable fractions >0 for each combination of age stratum and site are shown. Only attributable fraction CIs that exceed 0 are significant. EAEC=enteroaggregative *E coli.* EIEC=enteroinvasive *E coli.* tEPEC=typical enteropathogenic *E coli.* LT-ETEC=heat-labile enterotoxin-producing *E coli.* ST-ETEC=STh-producing enterotoxigenic *E coli.*

*Point estimate ≤0.