THE LANCET

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Liu J, Platts-Mills JA, Juma J, et al. Use of quantitative molecular diagnostic methods to identify causes of diarrhoea in children: a reanalysis of the GEMS case-control study. *Lancet* 2016; **388:** 1291–301.

Table S1. Real time PCR assays on TaqMan Array Card used for GEMS re-analysis.

	Pathogen	Gene	Reference
Viruses	Adenovirus 40/41	Fiber gene	[3]
	Adenovirus	Hexon	[4, 5]
	Astrovirus	Capsid	[4, 5]
	Norovirus GI	ORF1-2	[3]
	Norovirus GII	ORF1-2	[4, 5]
	Rotavirus	*NSP3*	[4, 5]
	Sapovirus	*RdRp*	[4, 5]
Bacteria	EAEC*	*aaiC*	[4, 5]
	EAEC*	*aatA*	[4, 5]
	EPEC*	*eae*	[4, 5]
	EPEC*	*bfpA*	[4, 5]
	ETEC*	*LT*	[4, 5]
	ETEC*	*STh* and *STp*†	[4, 5]
	STEC*	*stx1*	[4]
	STEC*	*stx2*	[4]
	Aeromonas	Aerolysin	[3]
	Bacteroides fragilis	*EGBF*	[3]
	Campylobacter jejuni and *C.coli*	*cadF*	[4, 5]
	Campylobacter	*cpn60*	[3]
	Clostridium difficile	*tcdA* and *tcdB*	[3, 4]
	Helicobacter pylori	*ureC*	[3]
	Salmonella	*ttr*	[3]
	Shigella/EIEC	*ipaH*	[4, 5]
	Vibrio cholerae	*hlyA*	[3]
Fungi	*Enterocytozoon bieneusi*	*ITS*	[3]
	Encephalitozoon intestinalis	SSU rRNA	[3]
Protozoa	*Cryptosporidium*	18S rRNA	[4, 5]
	Cyclospora cayetanensis‡	18S rRNA	[3]
	Cystoisospora belli‡	18S rRNA	[3]
	Entamoeba histolytica	18S rRNA	[4, 5]
	Giardia	18S rRNA	[4, 5]
Helminth	*Ancylostoma duodenale*	*ITS2*	[3]
	Ascaris lumbricoides	*ITS1*	[3]
	Necator americanus	*ITS2*	[3]
	Strongyloides stercoralis	Dispersed repetitive sequence	[3]
	Trichuris trichiura	18S rRNA	[4]
Controls	MS2	*MS2g1*	[4, 5]
	PhHV	*gB*	[4, 5]

* Each *E. coli* pathotype is defined as the original GEMS microbiology: EAEC (*aaiC*, or *aatA*, or both), typical EPEC (*bfpA* with or without *eae*), atypical EPEC (*eae* without either *bfpA*, *stx1*, or *stx2*), ST-ETEC (STh with or without LT), LT-ETEC (LT only), STEC (*eae* with *stx1*, *stx2*, or both, and without *bfpA*).
† ST-ETEC was defined as *STh* only ETEC to be consistent with the original GEMS, although *STp* was also tested on TAC.
‡ The *Cyclospora* and *Cystoisospora* assays target similar sequence, thus when both were positive we confirmed results by sequencing.

Validity of qPCR results. Validity accounted for

1. experimental flags: the results are excluded from analysis when assigned BADROX in combination with NOISE or SPIKE by the real time PCR instruments.
2. external controls: the negative results of a sample are valid only when its external controls are positive, Cq ≤ 35; otherwise, invalid. RNA and DNA targets are based on MS2 and PhHV, respectively.
3. extraction blanks: the positive results are valid only when the corresponding extraction blank is negative for the relevant targets, Cq > 35; otherwise, invalid.
4. each case and control pair: a pair is included in the analysis only when both the case and its paired control yielded valid results for the complete list of interrogated targets.

Conversion of Cq values to copy numbers. For derivation of pathogen target copy numbers from Cq, we generated standard curves at each laboratory using combined positive control materials of known copy number (constructed plasmids for DNA targets and in vitro transcripts for RNA targets). The combined extraction and amplification efficiencies for RNA and DNA targets were calculated as external control MS2 and PhHV copy numbers in the sample divided by input MS2 and PhHV copy numbers, respectively (defined as 95% percentile of MS2 or PhHV signal in all the clinical samples by site to avoid the variation in reagent concentrations among sites). Target copy numbers were then calculated from Cq values and adjusted for the extraction and amplification efficiencies of the specimen from the external controls.

Table S2. Assays used to confirm ipaH detection for *Shigella*/EIEC and identify *S. flexneri* and *S. sonnei*. qPCR was performed on nucleic acid from a subset of stools on 384-well plates along with the cognate *ipaH* assay on a random subset of 450 case and control pairs of age group 2 from Mali, Mozambique, and India (150 pairs each). PCR conditions identical to those of TaqMan Array Card was applied using AgPath One Step RT-PCR reagents, with a reverse transcription step at 45°C for 20 min, and an initial denaturation at 95°C for 10min, followed by 40 cycles of 95°C 15 sec, 60°C 1 min.

Species	Target Gene	Oligonucleotide sequences (F: forward, R: reverse, P: probe)	Reference
Shigella spp. and EIEC	*ial*	F: GTGAGGTTTTATTGTCTTTTTGTCATG R: CAATCTAAAAAAAAACGCCGTTCA P: TTTGATGGACATTTCAAG	6
	ShET2	F: ATGTGCCTGCTATTATTTAT R: CATAATAATAAGCGGTCAGC P: CCCGATATTCGTCCTC	7
	virA	F: TCACATCACGTCTTCCTCTGT R: AGCCTGTTTTCTATGTTTTCGGA P: CACATATATGCCCAAATTAT	8
	virG	F: TCAGAAAGGTAATTGGCATGGA R: AGAACCGCGCCCAAAGA P: AGGGCGGAATATT	9
Shigella spp. and a fraction of EIEC strains	*ipaH3*	F: CGCGGCACCGGAAAA R: CCAGCCGTACCTGTAAGAAATCA P: TATCTCGGTTTCCTCTGAGTG	10
Shigella flexneri	Putative periplasmic protein*	F: TGGGTGCATCCTGACCTGT R: GACAAACAATAACGAGCTACCGAT P: ACCACGGAATAATCCCGCAG	10
	O-antigen#	F: CTCCTATCCGTGATTATAGTGCA R: GCACACACAACTCACTGTATTT P: TCCTTCTCACGATTAAAATC	10
	Type 3 restriction enzyme#	F: CTTTCAACGCACGAATATCAAC R: GAACCTGATCCAGACGGAGA P: TTCTTCAGAACCGGGTTTTG	10
Shigella sonnei	Putative methylase	F: TGCCGCTAAAATCCTTCTGT R: GCGTACGACGAAAGGAAAAA P: GAAGTTATTGATTCCGCCC	This study

*This assay detects most *S. flexneri* serotypes except for serotype 6.
#The combination of these two assays identifies *S.flexneri* serotype 6 when both are positive (Cq≤35).

Table S3. Quantitative PCR linearity across the laboratories. Linearity was tested via standard curves for each pathogen using combined positive controls at each site, R^2 was averaged from all laboratories.

Target Name	R^2	CV, %
Adenovirus 40/41	0·996	0·4
Aeromonas	0·992	0·9
Astrovirus	0·996	0·8
Campylobacter jejuni and *C. coli*	0·998	0·2
Cryptosporidium	0·996	0·5
Cyclospora cayetanensis	0·994	1·0
Cystoisospora belli	0·989	2·0
Entamoeba histolytica	0·986	1·9
EAEC aaiC	0·998	0·1
EAEC aatA	0·996	0·5
EPEC bfpA	0·998	0·2
EPEC eae	0·998	0·2
ETEC LT	0·999	0·1
ETEC STh	0·998	0·2
ETEC STp	0·990	1·7
Helicobacter pylori	0·998	0·2
MS2	0·998	0·2
Norovirus GII	0·991	1·7
PhHV	0·998	0·2
Rotavirus	0·998	0·1
Salmonella	0·997	0·3
Sapovirus	0·996	0·5
Shigella/EIEC	0·997	0·6
STEC stx1	0·995	0·4
STEC stx2	0·996	0·4
Vibrio cholerae	0·999	0·2

Figure S1. Sensitivity and specificity of the original microbiologic workup compared with qPCR. This is the overall comparison on 5304 pairs. To harmonize the comparison of qPCR with the original diagnostic results, we combined culture results for *Campylobacter jejuni* and *Campylobacter coli,* and for *Salmonella* Typhi and non-typhoidal *Salmonella.* In the upper panel, coloured bars show the sensitivity/specificity of original comparator methods versus qPCR results at the analytical cutoff (Cq 35) while bars with dashed line show sensitivity when diarrhoea-associated Cq cutoffs were applied (improved modestly from 26±14% to 41±16%). In the lower panel, for all pathogens tested, comparator test positive samples had significantly higher quantity than those of comparator test negative samples.

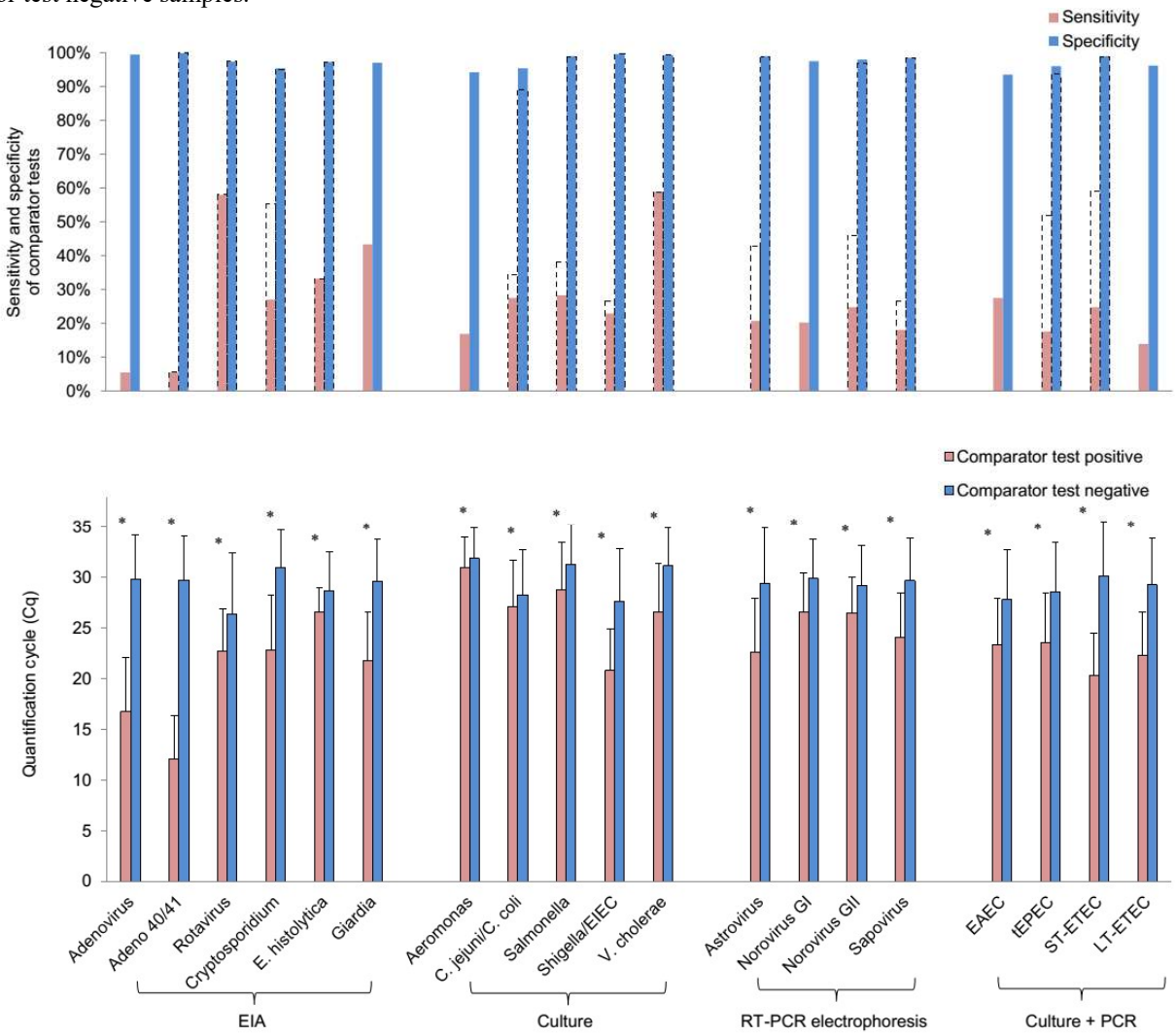
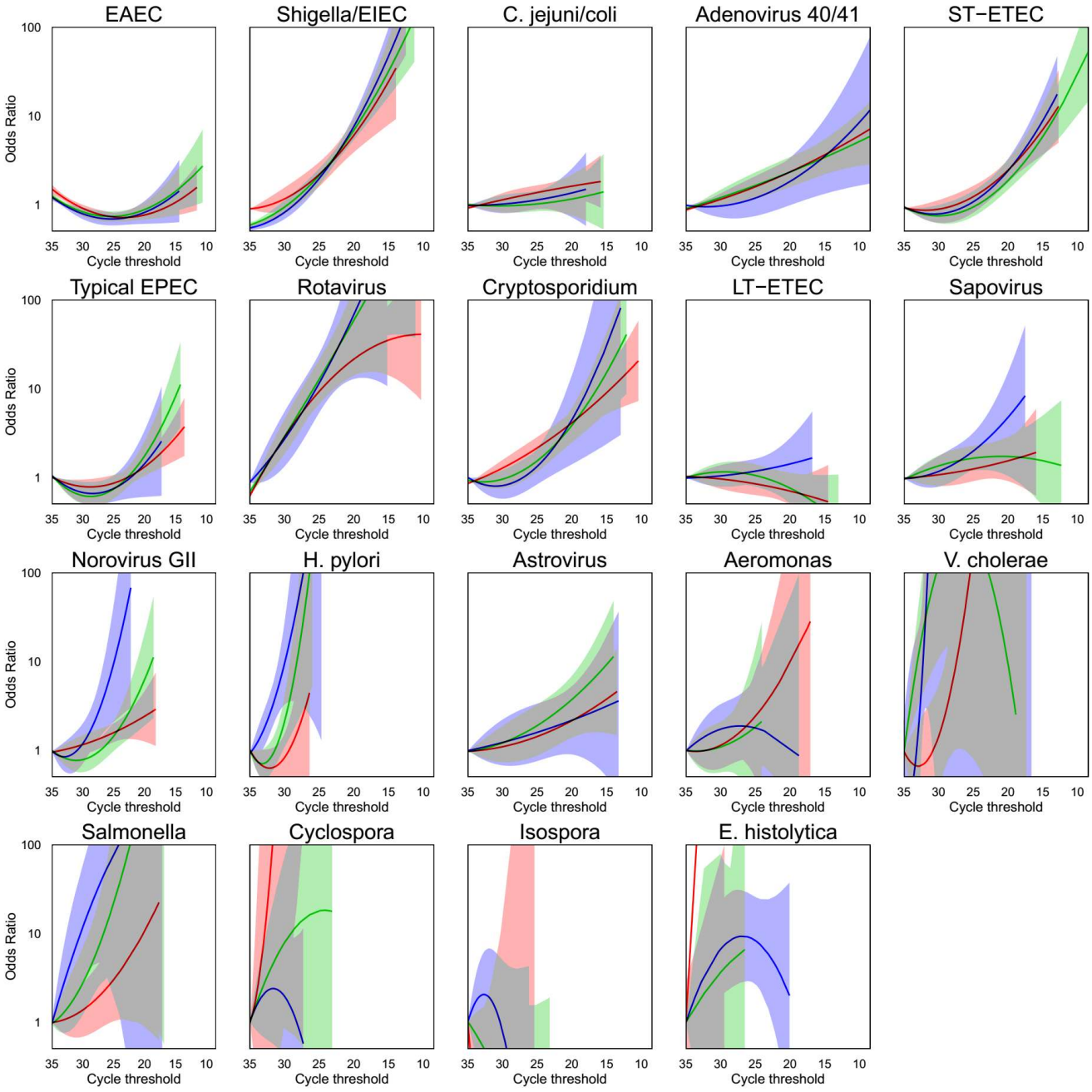


Figure S2. Relationship between pathogen quantity and diarrhoea by age group. Pathogens are ordered from top to bottom and left to right according to their prevalence in diarrhoeal cases. The x axis shows pathogen quantity (quantitative PCR Cq; quantity increases from left to right for each pathogen). The lines represent odds ratios for diarrhoea derived from the multivariable conditional logistic regression model as described in the methods for age 0-11 months (red), 12-23 months (green), and 24-59 months (blue) and the surrounding bands denote the 95% CI. Grey reveals overlap. EAEC=enteroaggregative E. coli, EIEC=enteroinvasive E. coli, EPEC=enteropathogenic E. coli, LT-ETEC=LT-producing enterotoxigenic E. coli, ST-ETEC=STh-producing enterotoxigenic E. coli, and STEC=Shiga toxin producing E. coli.



Attributable fraction comparisons between methodologies, Cq and copy number, and ETEC targets. For these estimates, since we were not making an inference about the population, we used Monte Carlo simulation to produce uncertainty in the AF estimates. Specifically, for all cases at the appropriate combination of site (Figures S3 and S5) or site and age (Figure S2), we generated 1000 estimates of the coefficients from the pathogen-specific regression models using draws from a joint normal distribution based on the mean coefficients and the covariance matrix from the conditional logistic regression model for each pathogen. Each draw was used to calculate an AF, and the 95% confidence intervals were derived from the 2·5% and 97·5% quantiles from this distribution.

Figure S3. Overall adjusted pathogen-specific attributable fractions (AFs) with qPCR using raw Cq values versus Cq-derived copy numbers. There was no statistically significant difference between AF estimates (Wilcoxon signed-rank test, P = 0·94).

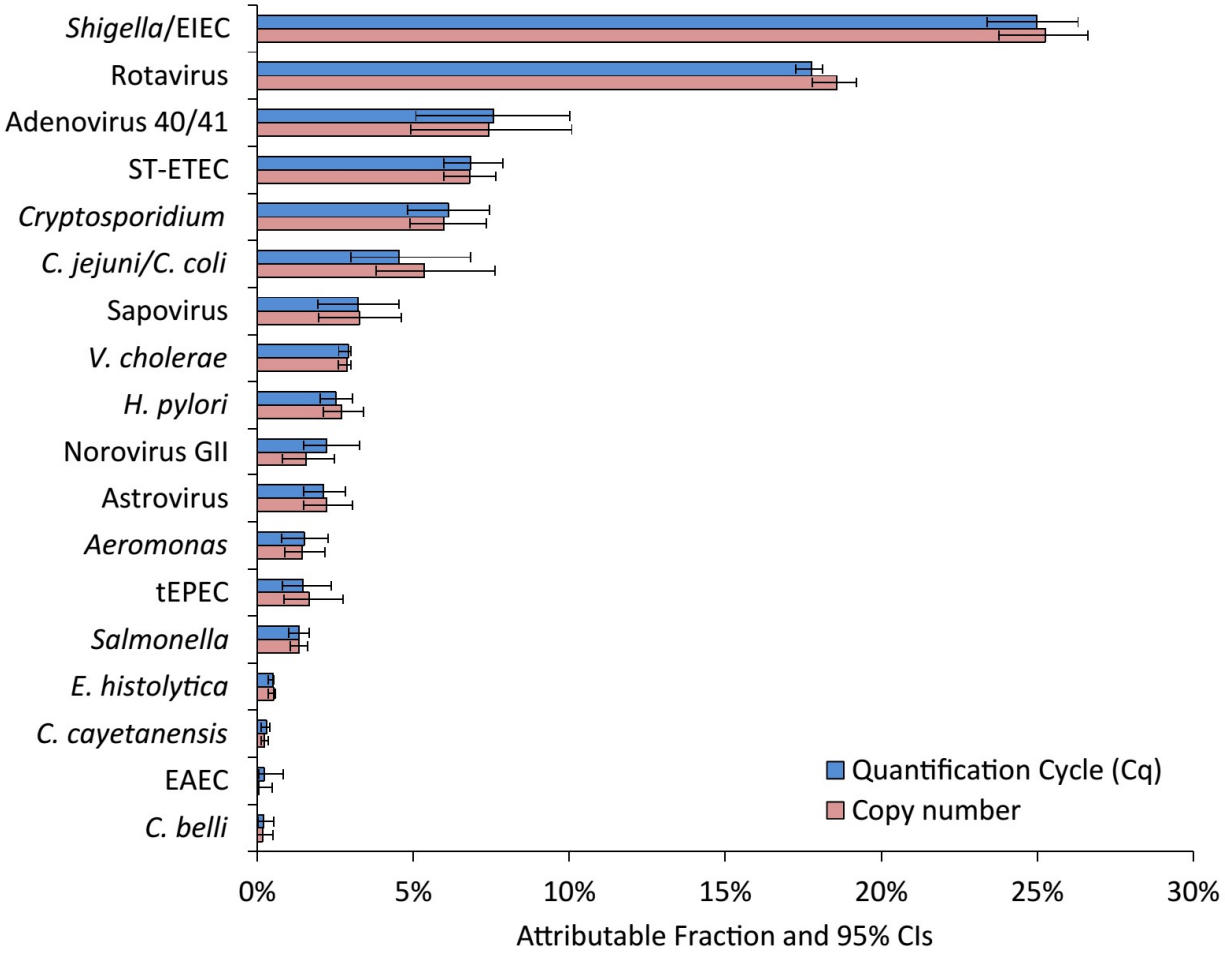


Figure S4. Pathogen-specific adjusted attributable fractions estimated by original GEMS study diagnostics and qPCR at each site.

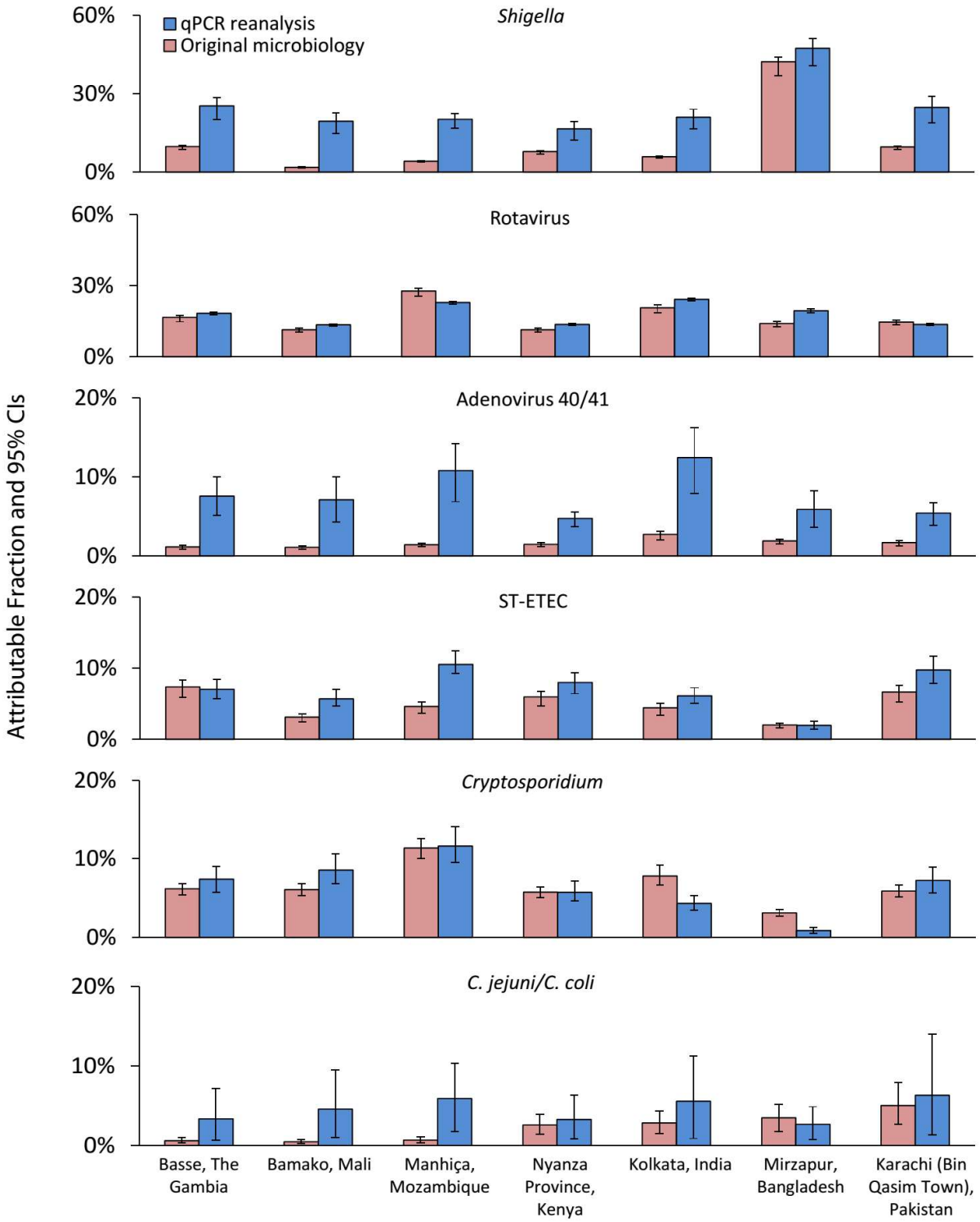


Figure S5. *Shigella*/EIEC quantity and sensitivity of culture by site. The top panel shows boxplots of Cq values in *Shigella*-positive stools by qPCR. The bottom panel shows the sensitivity of *Shigella* culture across ranges of Cq values. *Shigella* Cq values were lower in Bangladesh than any other site (Mann-Whitney test, P < 0·001).

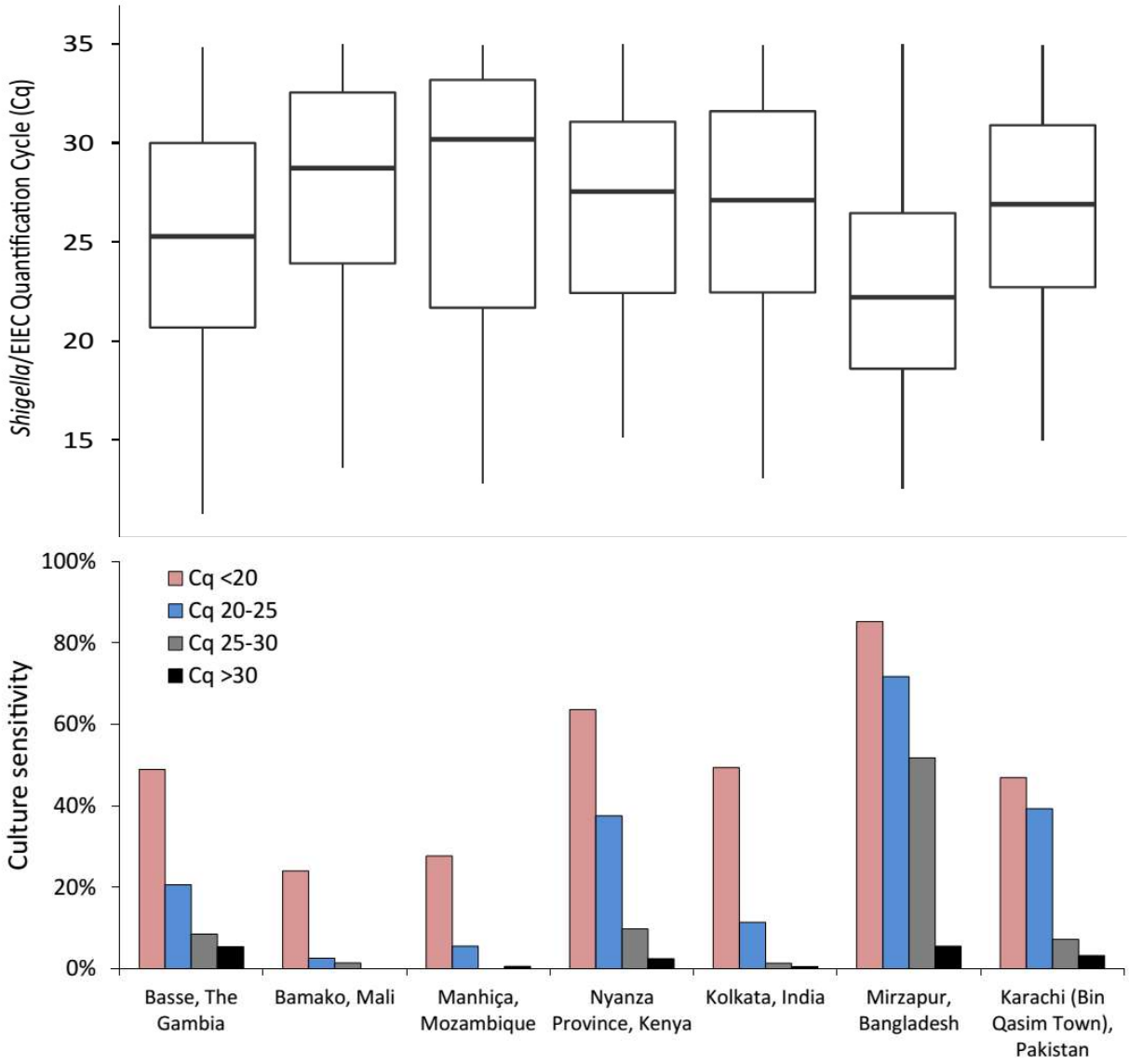


Figure S6. *Shigella*/EIEC quantity in watery and dysenteric cases. Box-and-whiskers plots are shown using Tukey's method, with points representing outliers. For cases in which *Shigella*/EIEC was detected by qPCR, the Cq values are shown by country for watery (blue) and dysenteric (red) episodes. Overall Cq 22·0 ± 4·6 in dysenteric cases vs. 26·2 ± 5·8 in watery cases, Wilcoxon rank sum test P < 0·0001.

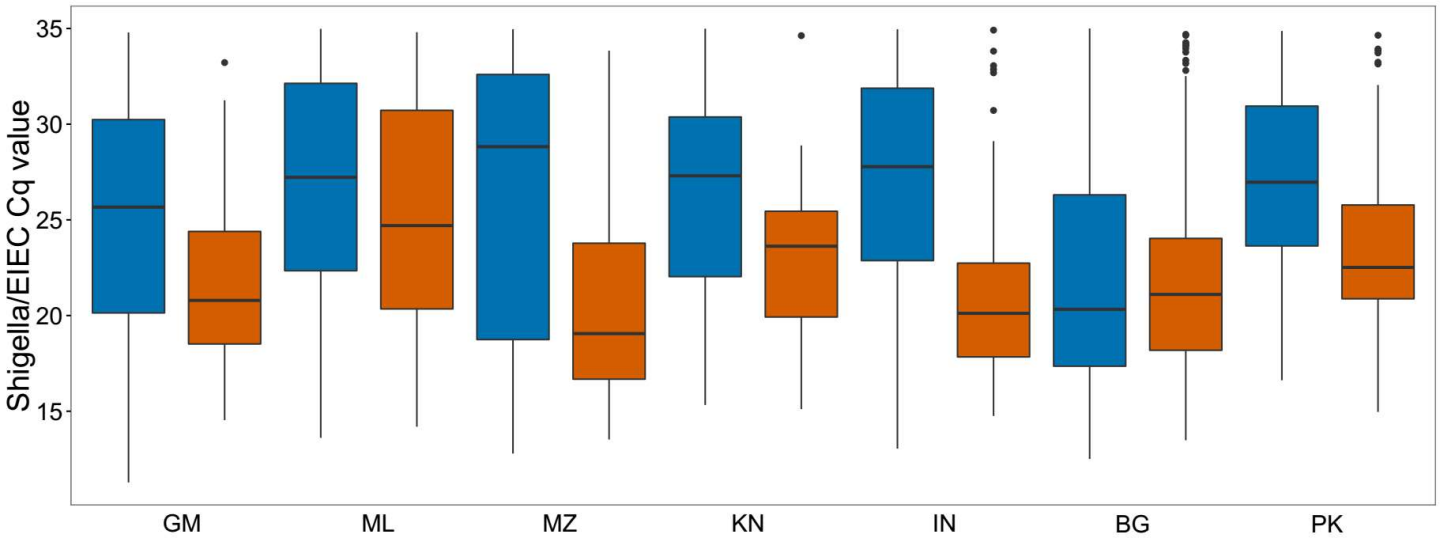


Figure S7. *Shigella*/EIEC attributable fractions derived from alternative virulence factors and *Shigella* species-specific qPCR assays. The assays, listed in Table S4, were performed on 384-well plates along with the cognate *ipaH* assay on a random subset of 450 case and control pairs of age group 2 from Mali, Mozambique, and India (150 pairs each).

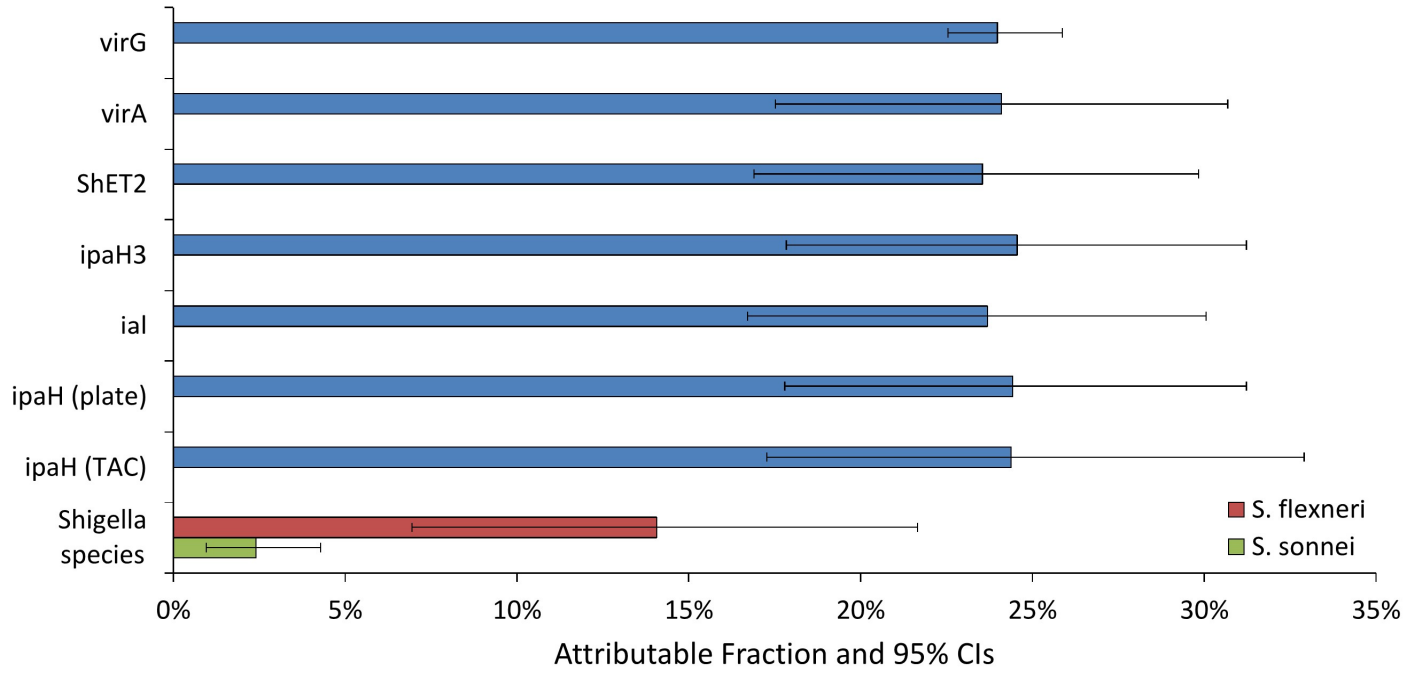
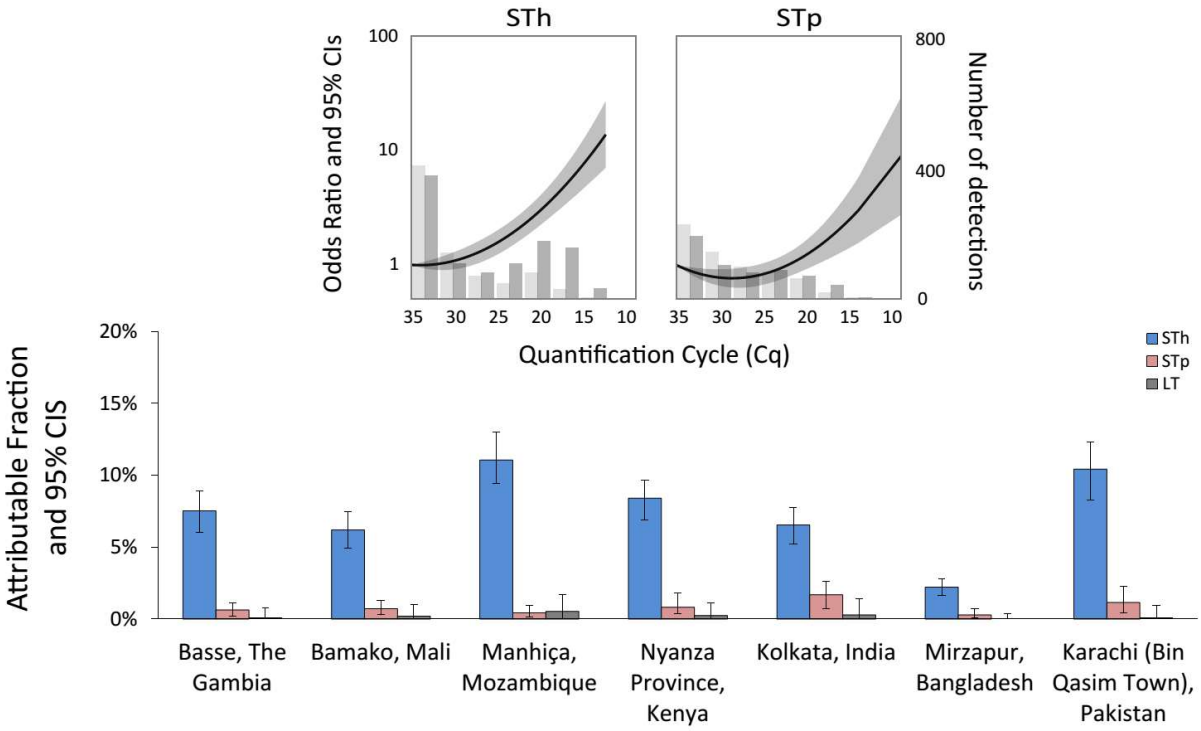


Figure S8. Adjusted attributable fractions of diarrhoea by ETEC target.



Calculation of alternative cutoffs. First, we calculated an alternative model-derived cutoff, where diarrhoea-associated quantities (as described in the Methods in the main text) were further restricted to those quantities above which the point estimate of the odds ratio was greater than 2. As an alternative to model-derived cutoffs, we also calculated a cutoff designed to maximize discrimination between cases and controls: using all cases and controls in which each pathogen was detected, we determined the maximum Youden Index (YI) from a receiver operating characteristic (ROC) curve with case-control status as the outcome and pathogen Cq value as the independent variable, where $YI = Sensitivit \ \ + Specificity - 1$.

Table S4. Quantitative cutoffs for each pathogen. Diarrhoea-associated pathogen quantities were defined as all quantities above the point at which the 95% confidence interval of the odds ratio exceeded 1. We defined "highly diarrhoea-associated" quantities as all quantities above the point at which the point estimate of the odds ratio exceeded 2. The ROC cutoff is the pathogen quantity that maximally discriminates case-control status (Youden Index = Sensitivity + Specificity -1).

Pathogen	Diarrhoea-associated quantity		Highly diarrhoea-associated quantity	ROC Cutoff
	Cq	Copy number*	Cq	Cq (Youden Index)
Vibrio cholerae	34·9	$7{\cdot}3 \times 10^{3}$	33·8	29·3 (0·55)
Rotavirus	35·0	$1{\cdot}5 \times 10^{5}$	32·6	26·9 (0·48)
Entamoeba histolytica	34·8	$4{\cdot}6 \times 10^{6}$	32·8	30·1 (0·43)
Cyclospora cayetanensis	29·6	$3{\cdot}7 \times 10^{5}$	29·6	34·0 (0·40)
Shigella/EIEC	33·1	$2{\cdot}1 \times 10^{6}$	27·9	26·1 (0·36)
Salmonella	32·4	$2{\cdot}2 \times 10^{5}$	30·7	29·7 (0·29)
ST-ETEC (STh)	26·2	$2{\cdot}0 \times 10^{7}$	22·8	25·4 (0·25)
Helicobacter pylori	30·8	$3{\cdot}6 \times 10^{5}$	30·8	33·0 (0·18)
Astrovirus	25·5	$4{\cdot}8 \times 10^{8}$	22·2	28·1 (0·18)
Cryptosporidium	29·1	$1{\cdot}8 \times 10^{6}$	24·0	27·5 (0·17)
Norovirus GII	27·6	$1{\cdot}1 \times 10^{8}$	23·4	28·8 (0·15)
Adenovirus 40/41	35·0	$2{\cdot}5 \times 10^{5}$	22·7	30·2 (0·08)
C. jejuni/C. coli	19·7	$2{\cdot}5 \times 10^{7}$	15·4	25·8 (0·08)
Typical EPEC (*bfpA*)	19·5	$2{\cdot}7 \times 10^{9}$	16·0	19·9 (0·07)
Sapovirus	31·6	$9{\cdot}4 \times 10^{5}$	NA	34·1 (0·02)

*per gram of stool

Figure S9. Detection of co-infections in diarrhoeal cases with highly diarrhoea-associated quantities. Each diarrhoea case with one of the 14 pathogens was further categorized as to whether there was only one pathogen detected at a highly diarrhoea-associated quantity (pink), multiple pathogens detected at highly diarrhoea-associated quantities (if so the primary pathogen was the one at a quantity with the highest odds ratio (red), and all other highly diarrhoea-associated quantities were considered secondary (blue), or if detections were at non-highly diarrhoea associated quantities (grey). The Cq cutoff used to identify highly diarrhoea-associated detections is shown in parentheses after each pathogen

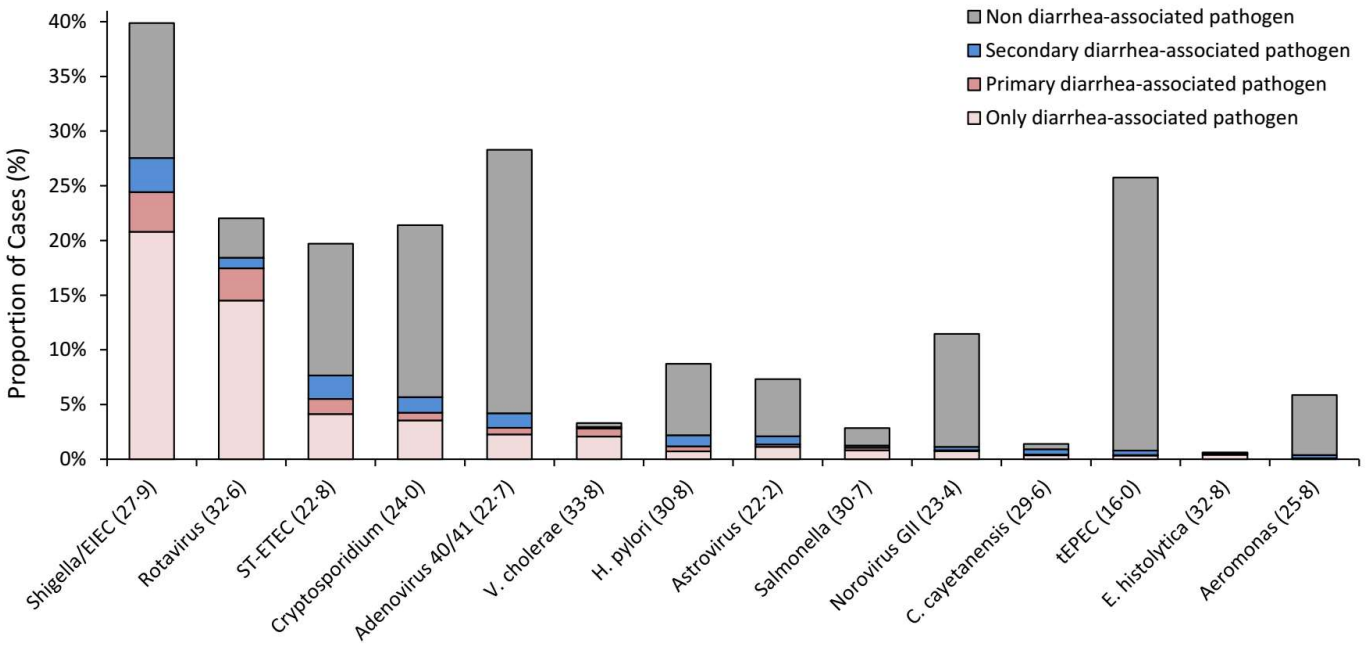
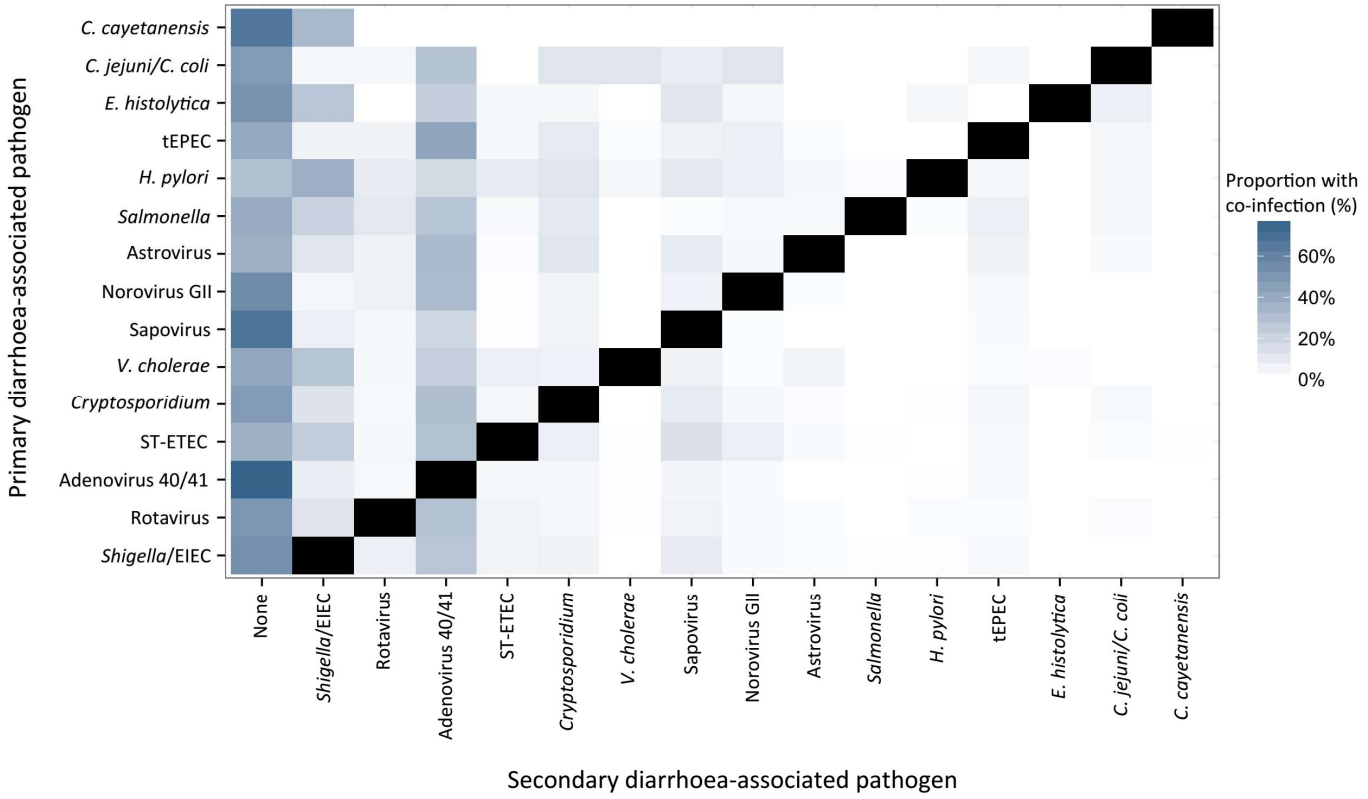


Figure S10. Co-infections in diarrhoeal cases. Each box represents the proportion of cases in which the primary diarrhoea-associated pathogen (Y-axis; defined as the pathogen with the highest quantity-specific odds ratio of those pathogens present in each stool at diarrhoea-associated quantities) is present for which an additional, secondary, diarrhoea associated-pathogen is present (X-axis; defined as any other pathogen present at a diarrhoea-associated quantity).


Primary diarrhoea-associated pathogen
C. cayetanensis
C. jejuni/C. coli
E. histolytica
tEPEC
H. pylori
Salmonella
Astrovirus
Norovirus GII
Sapovirus
V. cholerae
Cryptosporidium
ST-ETEC
Adenovirus 40/41
Rotavirus
Shigella/EIEC
None
Shigella/EIEC
Rotavirus
Adenovirus 40/41
ST-ETEC
Cryptosporidium
V. cholerae
Sapovirus
Norovirus GII
Astrovirus
Salmonella
H. pylori
tEPEC
E. histolytica
C. jejuni/C. coli
C. cayetanensis
Secondary diarrhoea-associated pathogen
Proportion with co-infection (%)
60%
40%
20%
0%

References

1. 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(9888): 209-22.
2. 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: S232-45.
3. Liu J, Gratz J, Amour C, et al. Optimization of quantitative PCR methods for enteropathogen detection. PLOS One 2016; 11(6): e0158199.
4. Liu J, Gratz J, Amour C, et al. A laboratory-developed TaqMan Array Card for simultaneous detection of 19 enteropathogens. Journal of clinical microbiology 2013; 51(2): 472-80.
5. Liu J, Kabir F, Manneh J, et al. Development and assessment of molecular diagnostic tests for 15 enteropathogens causing childhood diarrhoea: a multicentre study. The Lancet Infectious diseases 2014; 14(8): 716-24.
6. Frankel G, Giron JA, Valmassoi J, et al. Multi-gene amplification: simultaneous detection of three virulence genes in diarrhoeal stool. Mol Micro 1989; 3(12): 1729-34.
7. Vargas M, Gascon J, De Anta M, et al. Prevalence of Shigella Enterotoxins 1 and 2 among Shigella Strains Isolated from Patients with Traveler's Diarrhea. J Clin Micro 1999; 37(11): 3608-11.
8. Villalobo E, Torres A. PCR for detection of Shigella spp. in mayonnaise. Appl Environ Micro 1998; 64(4): 1242-5.
9. Alipour M, Talebjannat M, Nabiuni M. Polymerase chain reaction method for the rapid detection of virulent Shigella spp. International J Mol Clin Micro 2012; 2: 134-7.
10. Sahl JW, Morris CR, Emberger J, et al. Defining the phylogenomics of Shigella species: a pthway to diagnostics. J Clin Micro 2015; 53(3): 951-60.