

Acute gastroenteritis associated
with enterotoxigenic *Escherichia*
coli and norovirus in children in
Malawi

Thesis submitted in accordance with the
requirements of the University of Liverpool for

The Degree of Doctor of Medicine

by

Eamonn Trainor

BSc (Hons) MBChB (Hons) FRCPath

December 2018

Declaration

This thesis is the result of my own work. The material presented in it has not been presented, either wholly or in part, for any other degree or qualification.

Funding

I was granted a career development award by the Association of Clinical Pathologists (ACP) to support the research and programme of study leading to submission of this thesis.

Acknowledgements and thanks

I would like to acknowledge and thank the following individuals and staff, who offered technical and practical assistance:

- The laboratory staff at the Department of Clinical Infection, Microbiology and Immunology at the Institute of Infection and Global Health, University of Liverpool, particularly Mrs W Dove for her technical support and advice.
- Dr D Allen at Public Health England (PHE) for his support with the work presented in Chapter 4.

I would like to express my grateful thanks to:

My supervisors, Professor Nigel Cunliffe and Professor Miren Iturriza-Gomara, who provided constant support, encouragement, advice and enthusiasm throughout the 6 years which this thesis has been prepared.

Abstract

Introduction and Methods: Enterotoxigenic *Escherichia coli* (ETEC) are an important cause of diarrhoeal disease among children in low-income countries. The burden of ETEC infection has not been previously described in children in Malawi. Archived faecal samples collected from hospitalised children < 5 years of age with diarrhoea in Blantyre, Malawi over a 10-year period (1997-2007); and a comparison group without diarrhoea over a 2-year period (1997-1999) prior to rotavirus vaccine introduction were examined by conventional PCR for ETEC heat-labile (LT) and heat-stable (STh and STp) enterotoxins. As well as ETEC, noroviruses are endemic in human populations and recognised as a leading cause of acute gastroenteritis worldwide, with GII.4 norovirus strains most frequently identified. The molecular epidemiology of GII.4 norovirus strains among sporadic or community cases of gastroenteritis is poorly described, especially in resource poor settings. Current surveillance programmes are biased towards strains identified from outbreaks in high-income settings. Norovirus strains detected during studies of sporadic diarrhoeal disease in the UK and Malawi between 1993 and 2009 were characterised to better understand the distribution of GII.4 norovirus strains in the wider population.

Results and Conclusions: ETEC was detected in 10.6% of children with and 7.3% of children without diarrhoea. The most prevalent circulating toxin types were STh (6.6%), followed by LT (2.1%) and STp (0.9%). ETEC infection was most prevalent in infants aged 6-11 months and co-infection with rotavirus was common. The burden of ETEC infection in young children in Malawi is significant and following the introduction of rotavirus vaccine should become a focus of diarrhoea prevention efforts. GII.4 norovirus strains of global epidemic importance have circulated in the community up to 18-years prior to their subsequent recognition as pandemic strains associated with increases in outbreaks. More inclusive surveillance programmes that comprise strains associated with sporadic cases may allow for earlier detection of emerging strains with pandemic potential and inform vaccine formulations as they become available.

Table of Contents

Declaration.....	ii
Funding.....	ii
Acknowledgements and Thanks.....	iii
Abstract.....	iv
Table of Contents.....	v
1 Introduction	
1.1 Global burden of childhood diarrhoeal disease.....	1
1.2 Aetiology of infectious childhood diarrhoea.....	2
1.3 Norovirus.....	3
1.3.1 Basic structure.....	4
1.3.2 Genome and molecular structure.....	5
1.3.3 Virus classification and nomenclature.....	7
1.3.4 Norovirus strain diversity.....	8
1.3.5 Norovirus replication.....	9
1.3.6 Pathogenesis.....	11
1.3.7 Epidemiology and transmission.....	12
1.3.8 Immunity and host susceptibility.....	15
1.3.9 Clinical features.....	16
1.3.10 Diagnosis.....	17
1.3.11 Treatment and prevention of childhood diarrhoeal disease.....	19
1.3.12 Treatment of Norovirus gastroenteritis.....	21
1.3.13 Prevention of Norovirus gastroenteritis.....	22
1.4 Enterotoxigenic <i>Escherichia coli</i>	23

1.4.1 Classification and nomenclature of <i>Escherichia coli</i>	23
1.4.2 Diarrhoeagenic <i>Escherichia coli</i>	24
1.4.3 Pathogenesis.....	24
1.4.4 Epidemiology and transmission.....	27
1.4.5 Immunity.....	29
1.4.6 Clinical features.....	30
1.4.7 Diagnosis.....	30
1.4.8 Treatment and prevention.....	31
1.5 Aims and Objectives.....	35
2 Materials and Methods.....	37
2.1 Specimens.....	37
2.2 Subject enrolment.....	40
2.3 Specimen storage.....	41
2.4 Preparation of faecal suspensions.....	41
2.5 Identification of enteric pathogens.....	42
2.5.1 Enteric viruses.....	42
2.5.1.1 Rotavirus.....	42
2.5.1.2 Norovirus.....	42
2.5.2 Enteric bacteria.....	48
2.5.2.1 Campylobacter.....	48
2.5.2.2 Enterotoxigenic <i>Escherichia coli</i>	48
2.6 Statistical analysis.....	53
2.7 Ethics.....	53
3 Enterotoxigenic <i>Escherichia coli</i> in Hospitalised Children with and without diarrhoea in Blantyre, Malawi.....	54

3.1 Introduction.....	54
3.2 Aims.....	56
3.3 Study design.....	56
3.4 Results.....	56
3.4.1 Baseline characteristics of the study population.....	56
3.4.2 Prevalence.....	57
3.4.3 Age distribution.....	58
3.4.4 Seasonality.....	58
3.4.5 Toxigenic profile of circulating ETEC strains.....	58
3.4.6 Mixed infections.....	59
3.5 Discussion.....	64
4 Molecular epidemiology of GII.4 Norovirus strains in Malawi and the UK...74	
4.1 Introduction.....	74
4.2 Aims.....	85
4.3 Materials and methods.....	86
4.3.1 Specimens.....	86
4.3.2 Nucleic acid extraction, reverse transcription and norovirus testing.....	87
4.3.3 Amplification of the GII.4 Norovirus P2 domain for strain characterisation.....	88
4.4 Results.....	88
4.4.1 Molecular epidemiology of GII.4 strains detected in the UK and Malawi.....	89
4.4.2 Analysis of amino acid variation in the hypervariable domain of the major capsid protein.....	93

4.5 Discussion.....	96
5 General discussion and recommendations.....	100
5.1 Summary of main findings of research.....	100
5.2 Strengths and limitations.....	102
5.3 Further work and recommendations.....	104
References.....	106
Appendices.....	128
Appendix 1: Publications arising from thesis data.....	128
Appendix 2: Related publications.....	129

1 INTRODUCTION

1.1 Global burden of childhood diarrhoeal disease

Morbidity and mortality from childhood diarrhoeal disease presents an important global health challenge, especially in low and middle-income countries. Excluding deaths that occur in the neonatal period (0-28 days), after acute respiratory infections, diarrhoeal disease is the second most common cause of mortality in children <5 years of age worldwide (1). In 1990 it was estimated that childhood mortality (<5 years) from diarrhoeal disease was approximately 1.6 million, falling to 0.5 million in 2017 (2). Although significant progress has been made, diarrhoeal disease burden remains high despite improvements in childhood nutrition, the use of oral rehydration solutions and the introduction of rotavirus vaccine (3) in some of the poorest resource settings.

Diarrhoea leads to dehydration and electrolyte disturbances, which are often fatal in paediatric populations with poor nutritional status and impaired immune responses (including HIV), as often seen in low and middle-income countries (4). In addition repeated episodes of diarrhoea lead to malnutrition and growth stunting that impacts upon childhood developmental milestones and predisposes to other illnesses like pneumonia (5).

1.2 Aetiology of infectious childhood diarrhoea

The aetiological agents responsible for childhood diarrhoeal disease are varied and include bacterial (e.g. *Campylobacter* and *Shigella*), viral (e.g. rotavirus and norovirus) and parasitic/ protozoan (e.g. *Cryptosporidium* and *Giardia*) enteric pathogens. Re-analyses of faecal samples from two large diarrhoea aetiology studies using sensitive molecular diagnostic panels have provided some of the most comprehensive data to date on pathogen specific disease burdens in low and middle income settings.

The Global Enteric Multicenter Study (GEMS) (6) was a large prospective case control study examining the aetiology of moderate to severe diarrhoeal disease (MSD) in children < 5 years in seven study sites in Africa and Asia (Manhica, Mozambique; Bassei, The Gambia; Nyanza Province, Kenya; Karachi, Pakistan; Bamako, Mali; Mirzapur, Bangladesh; Kolkata, India; Manhica, Mozambique). Liu J *et al* (7) re-examined 5304 faecal sample pairs (11,400 specimens, from 5,700 cases and controls) using a quantitative molecular diagnostic approach with a TaqMan™ array card for 32 enteropathogens. The authors calculated population attributable fractions to determine which enteric pathogens were associated with diarrhoeal disease in light of, or to adjust for the high prevalence of asymptomatic pathogen carriage in controls. This approach involved the use of a multiple conditional logistic regression model that adjusted for the presence of other pathogens. The six most attributable enteropathogens were *Shigella* spp./ enteroinvasive *E coli* (EIEC); rotavirus; adenovirus 40/41; heat stable enterotoxin producing

E coli (ST-EPEC); *Cryptosporidium* spp. and *Campylobacter* spp, together accounting for 77.8% of all attributable diarrhoea. Compared with the original study which used a combination of diagnostic techniques including culture, EIA and PCR, more pathogens were attributed as causes of diarrhoea in the re-analysis (89.3% vs. 51.5%), and the incidence was greater using qPCR than the original methods, particularly for the following pathogens: adenovirus 40/41; *Shigella* spp./EPEC; *Campylobacter* spp. and ST-EPEC. *Shigella* spp./EPEC, rotavirus and ST-EPEC were consistently high across all age groups and regions. Norovirus genogroup II showed associations with diarrhoea, but the prevalence was low. The original GEMS study examined for the heat-labile (LT) and human heat-stable (STh) enterotoxins of EPEC, however the re-analysis also examined for the porcine heat stable enterotoxin (STp); this may explain the increased burden noted in the re-analysis. The additional inclusion of primers for STp increased the overall burden of EPEC by 15%. There was essentially no diarrhoeal burden attributed to LT enterotoxin alone. In 249 (43.9%) of 505 samples in which ST-EPEC (STh) was detected in diarrhoeal quantities, LT was also detected in similarly high quantities suggesting a role for both the toxins causing diarrhoea when present together.

The Etiology, Risk Factors, and Interactions of Enteric Infections and Malnutrition and the Consequences for Child Health and Development (MAL-ED) cohort study (8) was a large multi-site study at eight locations (Dhaka, Bangladesh; Vellore, India; Bhaktapur, Nepal; Naushero Feroze, Pakistan; Venda, South Africa; Haydom, Tanzania; Fortaleza, Brazil; and Loreto, Peru)

in children ≤ 2 years. Platts-Mills *et al* (9) re-examined 6625 diarrhoeal and 30,968 non-diarrhoeal surveillance stools from 1715 children using TaqMan™ array cards for 29 enteropathogens. As with the GEMS re-analysis the authors calculated population attributable fractions. The ten most attributable enteropathogens in order of importance were *Shigella*, sapovirus, rotavirus, adenovirus 40/41, enterotoxigenic *Escherichia coli*, norovirus, astrovirus, *Campylobacter jejuni* or *C coli*, *Cryptosporidium*, and typical enteropathogenic *E coli*. Again more pathogens were attributed as causes of diarrhoea in the re-analysis compared to the original study (64.9% vs. 32.8%). Four of the top five aetiologies identified in the GEMS and MAL-ED re-analyses were shared, *shigella*, rotavirus, adenovirus 40/41, and enterotoxigenic *E coli*.

1.3 NOROVIRUS

1.3.1 Basic structure

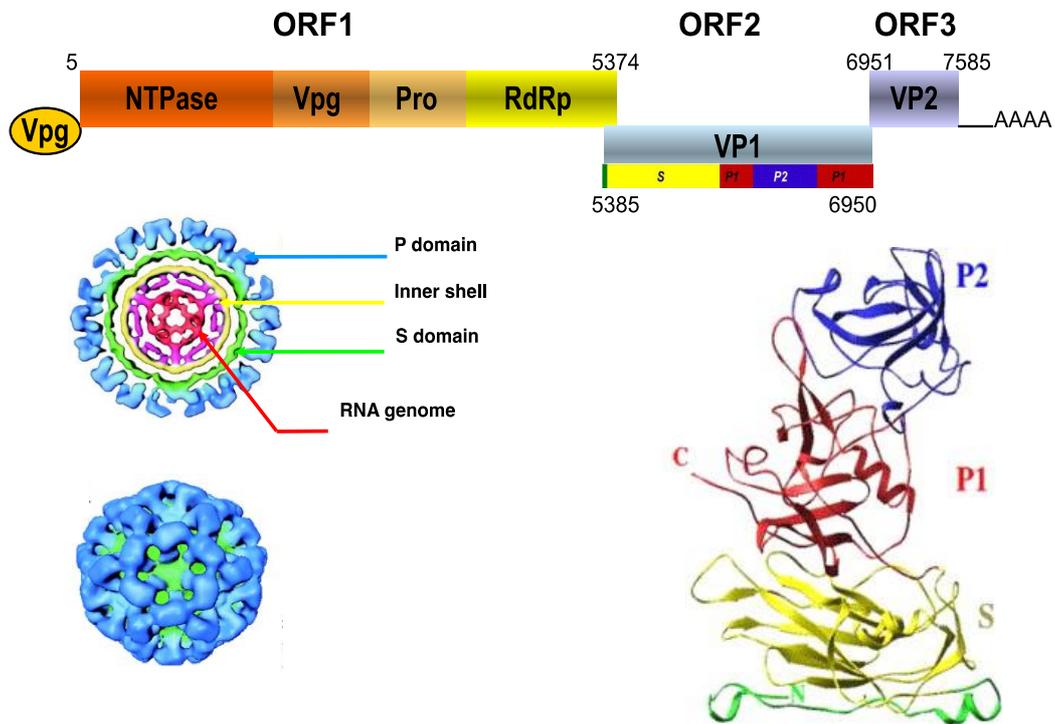
Norwalk virus, the prototype norovirus was first identified in 1972 by immune electron microscopy using inactivated convalescent sera from experimentally infected volunteers. These non-enveloped norovirus virions measuring 27 to 32 nm in diameter were observed in diarrhoeal stool obtained from volunteers administered bacterial free stool filtrates obtained from an outbreak of infectious non-bacterial gastroenteritis at a primary school in Norwalk, Ohio, USA (10). Further elucidation of the 3-dimensional structure of the virus followed cloning of the virus and expression of Norwalk virus capsids/virus like

particles in the baculovirus system (11). Using PAGE (12) and electron cryomicroscopy (13) norovirus virions were demonstrated to contain a single primary structural protein (VP1) organised into a icosahedral capsid symmetry (180 copies of the VP1 protein arranged in 90 dimers) with distinct cup-shaped surface depressions. These characteristic structural features are common to the Caliciviridae, the name of which is derived from the Latin word *calix*, meaning 'cup' or 'goblet'.

1.3.2 Genome and molecular structure

The norovirus genome is composed of 7654 nucleotides of plus sense, non-segmented, single stranded RNA (excluding a polyadenylated tail), which contains three open reading frames (ORFs) (14) [Figure 1]. ORF1 encodes a non-structural polyprotein that is cleaved by viral protease into six proteins involved in virus replication, including the RNA-dependent RNA-polymerase (RdRp). ORF2 encodes the major (VP1) capsid protein, which is composed of an S (shell) domain and a P (protruding) domain. The P domain is further divided into two subdomains (P1 and P2), the P2 sub-domain represents the most exposed surface of the viral capsid, and is believed to play an important role in antigenic and immune recognition functions, consequently most genomic sequence diversity is seen at this site. ORF3 encodes the minor (VP2) capsid protein, which has a proposed role in enhanced production and stability of the VP1 capsid protein (15).

Figure 1. Basic structure of the norovirus virion and genomic structure of the RNA genome (Reproduced with kind permission of M Iturriza-Gomara 2014).



1.3.3 Virus classification and nomenclature

Noroviruses are a genus within the Caliciviridae, along with four other genera, vesivirus, lagovirus, nebovirus and sapovirus (16, 17). The vesiviruses cause vesicular disease in porcine, equine, feline and murine mammals and upper respiratory tract infection in felines; the lagoviruses cause haemorrhagic disease in lagomorphs; neboviruses cause endemic diarrhoea in bovine calves, while the sapoviruses cause acute gastroenteritis in humans.

Noroviruses are classified using genetic analysis since, until very recently, inability to culture the virus has prohibited classification based on neutralization with anti-sera into distinct serotypes. There has however been recent success in laboratory cultivation of multiple human norovirus strains by a research team at the Baylor College of Medicine, Houston, Texas, USA using a bile dependent stem cell derived, non-transformed human intestinal enteroid monolayer culture (18).

Noroviruses are divided into seven distinct genogroups (GI-GVII), with genogroups GI, II and IV causing human infection; porcine (GII), bovine (GIII), feline (IV), murine (GV) and canine (GVI and VII) strains have also been described (19). Genogroups are further divided into genotypes and variants (subtypes) based on genomic sequence diversity. For example the prototype norovirus strain, Norwalk virus, belongs to genogroup I and genotype I, and is thus designated GI.I. Genogroup classification is based upon the VP1 protein divergence, and genotype classification upon either the RNA dependent RNA polymerase sequence or the capsid sequence. The propensity of noroviruses

to recombine, most commonly at the junction of ORF1/ORF2 necessitates that current norovirus nomenclature comprise both polymerase and capsid genotypes, as neither alone can fully describe the genotypic features of observed field strains (20, 21). Dual genotyping can be conducted in practice by submitting sequence information to an automated web-based tool: RIVM (the Dutch National Institute for Public Health and the Environment)-NoroNet (<http://www.rivm.nl/mpf/norovirus/typingtool>) (22). To date, 14 GI and 29 GII polymerase genotypes and 9 GI and 22 GII capsid genotypes have been described. Determination of variants (subtypes) requires sequencing of highly variable regions of the norovirus genome e.g. segment 5 (encompassing part of the P2 domain and the P1.2 subdomain) and the hypervariable P2 region (23, 24). The majority of newly emerging outbreak variants belong to GII.4 noroviruses. These variants are named using the geographical location where the strain was first isolated and the year in which they were detected e.g. GII.4 Sydney 2012.

1.3.4 Norovirus strain diversity

The great diversity of norovirus strains that cause human infection is driven by immune selection pressure; this is also observed in other RNA viruses e.g. influenza and Human Immunodeficiency Virus (HIV) (25). The most exposed portion of the viral capsid (P domain), which is also the presumptive binding site in initial infection (26), evolves in the face of host herd immunity. This genetic diversity occurs because the viral RNA genome is replicated by virus encoded RNA-dependent RNA polymerase, which has no intrinsic proof reading activity and leads to a relatively high mutation rate during genomic

translation and increased diversity resulting from accumulation of point mutations. This has been estimated at between 10^{-4} and 10^{-6} mutations per base pair per generation compared to 1.1×10^{-8} in the human genome (27, 28). This high diversity of viral quasi species, which refers to a population of genetically related viruses that are closely distributed around a consensus sequence (29) is also seen in other RNA viruses e.g. influenza and HIV (30).

1.3.5 Norovirus replication

The inability to culture human noroviruses until very recently has hindered characterisation of the viral life cycle. Consequently, current understanding of the norovirus life cycle has been gained by utilising several different approaches, including: cultivable surrogates (feline calicivirus and murine norovirus); a Norwalk virus replicon system and other biochemical and cell based approaches (31). An overview of the norovirus replication cycle is presented. For a more detailed description and reference library see (32).

Norovirus binding and cellular entry

Noroviruses are known to interact with carbohydrate moieties on cell surface membranes. In the case of human norovirus, the presumptive initial binding site in establishing infection is believed to be histo-blood group antigens (HBGAs), a family of complex glycans expressed on the surface of red blood cells, respiratory and gastrointestinal epithelia, as well as in saliva and breast milk (33). Cellular entry and uncoating proceed through currently undefined pathways, and as yet unidentified cellular receptor(s) are also thought

required to mediate cellular entry. Further elucidation of these mechanisms will likely follow from the recent successful cultivation of human norovirus *in vitro* (18).

Viral protein translation

Following capsid-uncoating, VPg linked viral genomic RNA acts as mRNA, and is translated into protein by the host cellular translational apparatus. Translation is believed to be via interaction with components of the eukaryotic eIF4F translation initiation factor complex. The ORF1 polyprotein undergoes post-translational cleavage by viral protease NS6, and results in release of viral NS proteins. Translation of the major (VP1) and minor (VP2) viral capsid proteins occurs primarily from the subgenomic RNA. Use of a subgenomic RNA is thought to produce the large quantities of primary structural protein required for virus assembly, as more subgenomic than genomic RNA has been observed in infected cells.

Genome replication

Replication complex formation occurs in close association with host derived cytoplasmic membrane complexes, and in murine norovirus, interaction with NS1/2 and NS4. Genome replication occurs via a negative-strand intermediate, and the viral RdRp, using both de novo and VPg-dependent mechanisms of RNA synthesis, generates genomic and subgenomic RNA.

Virion assembly and exit

The mechanisms of viral assembly, encapsidation and exit of noroviruses are largely unknown. The major capsid protein VP1 self assembles into VLPs, which may be sufficient to drive capsid assembly, however the minor capsid protein VP2, although not required for VLP assembly, is essential for production of infectious virions. VP2 may be involved in encapsidation via interaction with the viral RNA. No direct studies have examined exit and release of the assembled norovirus virion; however the induction of cellular apoptosis has been proposed as a possible exit mechanism for other caliciviruses.

1.3.6 Pathogenesis

Lack of a cell culture system (until very recently) or suitable animal model has meant our limited understanding of the pathogenesis of norovirus gastroenteritis comes primarily from data obtained during volunteer challenge studies (34, 35). Biopsy specimens taken from the proximal jejunum of ill volunteers have demonstrated histopathological changes associated with acute infection including; an intact mucosa, broadening and blunting of the intestinal villi, crypt-cell hyperplasia, cytoplasmic vacuolization, and infiltration of polymorphonuclear and mononuclear cells into the lamina propria. Histologic changes were not observed in gastric fundus, antrum or colonic mucosa, nor in biopsy specimens obtained during the convalescent phase of the illness and from those who did not display clinical symptoms. Small intestine brush border enzymatic activity (alkaline phosphatase, sucrase and

trehalase) is reduced, resulting in mild steatorrhea and transient carbohydrate malabsorption.

Abnormal gastric motor function, observed as marked delays in gastric emptying in infected volunteers may explain the nausea and vomiting associated with norovirus gastroenteritis (36). More recently analysis of biopsies from immunocompromised patients with chronic gastroenteritis has identified the small intestinal enterocyte as the primary site of infection (37).

1.3.7 Epidemiology and transmission

Norovirus gastroenteritis occurs in all age groups and a wide variety of settings, each associated with a broad spectrum of clinical outcomes. It has been estimated that norovirus is associated with 18% of all diarrhoeal disease worldwide (38). In high-income settings, norovirus disease incidence is approximately five times higher in children <5 years (21%), compared to the whole population (4.5% per year) (39). Limited data from developing countries points towards an even higher norovirus disease burden in children from these settings, in Peru for example, 50 cases per 100 person-years in the first two years of life (40).

Transmission of noroviruses is via the faecal-oral and vomit-oral routes; this may occur by direct person-to-person contact, contact with environmental fomites or ingestion of contaminated food and/or water. Norovirus is exceptionally contagious, with an ID₅₀ estimated at between 18.2 and 2800

genomic equivalents (41).

The epidemiological features of norovirus gastroenteritis can be divided into three distinct settings; (i) healthcare associated, (ii) epidemic foodborne and (iii) sporadic disease.

Healthcare –associated infection

Healthcare-associated infection typically occurs in semi-closed settings including: nursing/residential homes, hospital wards and day-care centres. Outbreaks typically start by introduction from a single individual followed by ongoing environmental transmission; herculean infection control efforts are often needed to control these outbreaks. Elderly and comorbid (e.g. chronic renal disease) individuals are disproportionately affected, and experience a more prolonged clinical course, typically 4–6 days, with an associated mortality not often described in other settings (42, 43).

Epidemic foodborne infection

Epidemic food borne gastroenteritis generally affects large numbers of otherwise healthy individuals (often adults) over a short time period. Foodborne outbreaks, which often receive significant media attention, again occur in semi-closed settings that favor rapid transmission, including: cruise ships, restaurants and catered events. The source of these outbreaks is typically from infectious food handlers and infected foodstuffs; faecally

contaminated oysters are often implicated. Individuals typically experience symptoms for 1-3 days (44).

Sporadic infection

Noroviruses are increasingly recognised as playing a major role in sporadic gastrointestinal disease. The highest rates of norovirus gastroenteritis are seen in young children, where norovirus is recognized as the second most common cause of acute viral gastroenteritis after rotavirus, and often requires hospitalization (45). In adults, acute sporadic gastroenteritis caused by norovirus is increasingly recognised; a large community-based study estimated that norovirus accounts for three out of seventeen million cases (18%) of sporadic gastroenteritis occurring in the United Kingdom each year, making norovirus the single most commonly detected pathogen (46).

1.3.8 Immunity and host susceptibility

That norovirus infection occurs in all age groups and populations may be explained by two mechanisms, (i) weak immunity following infection or (ii) great diversity of strains. Early challenge studies in previously healthy volunteers have suggested short-lived immunity by demonstrating partial immunity at virus rechallenge at 6-14 weeks, but not at 2-3 years (47).

However these studies did use very high infecting doses, not likely encountered in natural infection, so longer lived immunity at lower infective doses is possible.

Observations from challenge studies and outbreaks noted that a proportion of those exposed did not develop clinical illness (47, 48). However this could not be explained by previous infection, as antibody levels in volunteers did not support this observation. This suggested the possibility of some other mechanism of natural protection. Ultimately, this was explained by genetically determined host-susceptibility factors related to binding of norovirus VLPs to HBGAs (49) . There are now known to be three major HBGA families involved in norovirus binding (i) ABO, (ii) Lewis and (iii) Secretor. Human host susceptibility is determined by the secretor status of an individual, which is itself controlled by the fucosyltransferase 2 (FUT2) gene (50). Around 20% of Europeans are non-secretors, and will demonstrate resistance to infection with the prototype norovirus Norwalk GI.I, and to a variable extent other strains including common GI.4 virus (51). The great strain diversity seen in noroviruses mean non-secretors may not be resistant to all strains however.

1.3.9 Clinical features

Clinical observations from volunteer studies and outbreaks have demonstrated that following exposure to norovirus, individuals may experience an acute onset vomiting or diarrhoeal illness, and in approximately one-third an asymptomatic infection. In healthy adult volunteers, following an incubation period of 10-51 hours (mean 24 hours), illness typically lasts 24-48 hours (52). However, in hospitalised patients and infants <1 year, clinical symptoms may be more prolonged (i.e. 4-6 days) and more severe (53, 54). The major clinical features of symptomatic infection include nausea, vomiting, non-bloody watery diarrhoea, abdominal cramps and fever as well as other

constitutional symptoms including headache and myalgia (55). In infants <1 year diarrhoea is the predominant clinical feature, with vomiting typically seen in older infants and adults (54). Although norovirus is typically considered to cause a self-limiting mild gastroenteritis in otherwise healthy adults, there have been reports of fatalities in vulnerable elderly patients (42). Additionally in young infants in resource poor settings, rapid fluid and electrolyte loss from diarrhoea can often be fatal (56).

1.3.10 Diagnosis

Since norovirus was first described as an aetiological agent of infectious gastroenteritis, various clinical and laboratory tests have been developed to aid clinical diagnosis and epidemiological studies.

Clinical

Clinical and epidemiological features suggestive of norovirus gastroenteritis have been used to develop clinical diagnostic tools. The Kaplan criteria (57) have been validated for the detection of foodborne community norovirus outbreaks, and include mean illness duration 12-60 hours; mean incubation period 24-48 hours and vomiting in >50% of individuals. The utility of the Kaplan criteria for healthcare associated and sporadic disease is less clear.

Laboratory

Norovirus (the Norwalk particle) was first discovered in 1972 using immune EM (10). EM, although initially used in diagnosis, has been replaced in recent decades by immunological and molecular diagnostic modalities, which offer increased sensitivity and require less specialised training and expensive equipment. Immunological tests employ norovirus specific antibodies to detect norovirus antigens in clinical samples. Various commercially available ELISA and immunochromatographic platforms exist for detection of genogroup I and II noroviruses (58); however the great antigenic diversity seen within noroviruses restricts the diagnostic sensitivity of these tests and their routine use in clinical settings is limited.

Molecular methodologies offer the best sensitivity of currently available diagnostic tests. Conventional reverse-transcriptase PCR has the advantage that PCR amplicons can be sequenced for typing, however turnaround times are typically over five hours and this approach tends to be utilised more in research or reference laboratory settings (58). Real time reverse transcriptase PCR offers high specificity and sensitivity, as well as more efficient throughput and turnaround times; for these reasons it is the most commonly used test for genogroup I and II norovirus detection in diagnostic laboratories in developed settings (58). More recently, several diagnostic platforms have come onto the market that are able to detect multiple enteric pathogens (including viral, bacterial and parasitic) simultaneously (59). These include the xTAG GPP (Luminex Corporation, Toronto, Canada) and EntericBio (Serosep Limited,

Annacotty Business Park, Annacotty, Co. Limerick, Ireland) platforms (60, 61). Although attractive these platforms may also present diagnostic challenges; especially where multiple enteric pathogens are detected from a single clinical specimen and it is not clear whether this represents true infection, colonisation or shedding from a previous infection.

Diarrhoeal stool is the gold standard diagnostic sample for norovirus gastroenteritis, however other clinical samples with lower diagnostic sensitivities have also been used including rectal swabs, vomitus and mouthwash samples collected after emesis (62, 63).

1.3.11 Treatment and prevention of childhood diarrhoeal disease

The World Health Organisation (WHO) has published a document on the treatment of diarrhoea that provides a comprehensive overview on the principals and practice of treating infectious diarrhoea, especially in young children (56). A brief overview of the most essential principals will be discussed here, prior to a more specific discussion around treatment and prevention of norovirus gastroenteritis.

Dehydration

A clinical history and examination of the child with diarrhoea is required to assess the degree of dehydration and select an appropriate treatment plan. In non-severe dehydration this should include oral rehydration therapy, either at home or in a health-care facility. Fluids should contain salt to address electrolyte loss e.g. oral rehydration solution or salted rice water. In patients

with severe dehydration rapid intravenous (IV) rehydration is the preferred treatment, ideally in a hospital setting using Ringer's Lactate Solution or normal saline; nasogastric rehydration can be considered where IV therapy is not available. Measurement of serum electrolytes in children with dehydrating diarrhoea is not necessary in most instances, and may lead to inappropriate treatment.

Malnutrition

Children who die from dehydrating diarrhoea in developing settings are frequently already malnourished, often severely so. Continuing with the usual infant diet during the diarrhoeal episode and increasing this afterwards can prevent further nutritional damage. Food should never be withheld and should be as nutrient rich as possible. Breastfeeding mothers should be encouraged to continue to do so.

Zinc

Zinc deficiency is common among children in developing settings, and this is clinically significant as zinc has an important role in normal cellular growth and immune function. Numerous randomised controlled studies suggest that zinc supplementation significantly reduces the severity of illness and duration of diarrhoea in children. The WHO recommends that zinc (10-20mg/day) be given for 10-14 days to all children with diarrhoea. However a recent Cochran

review in 2013 suggests that it may only be of benefit to children >6 months of age (64).

Antimicrobial and antimotility agents

The empirical use of antimicrobials is not advocated, as often it is not possible clinically to distinguish between episodes of diarrhoea that may benefit from treatment or not. The exception to this is (i) children with bloody diarrhoea (probable shigellosis), (ii) suspected cholera with severe dehydration and (iii) laboratory proven, and symptomatic infection with *Giardia duodenalis*. For bacterial infections, choice of agent should be informed by local antimicrobial sensitivity profiles to those pathogens. Antimotility agents should be avoided, especially in children <5 years; they do not prevent dehydration or improve nutritional status, and can be associated with serious side effects such as paralytic ileus.

1.3.12 Treatment of norovirus gastroenteritis

Norovirus gastroenteritis is usually a mild and self-limiting disease, requiring no specific treatment. When treatment is required it is typically supportive with oral rehydration therapy and antipyretics. In more severe cases, intravenous fluids and electrolyte replacement may be required, especially in elderly hospitalised patients and infants, who are at increased risk of rapid volume depletion and resulting electrolyte disturbance (55, 65).

No specific antiviral therapy has demonstrated clinical efficacy against norovirus, although ribavirin and interferons have been shown to inhibit viral replication *in vitro* (66). Immune globulin (intravenous or oral) has been suggested as a treatment, but evidence is insufficient to recommend routine clinical use.

1.3.13 Prevention of norovirus gastroenteritis

There is no licensed vaccine currently available to protect from norovirus infection. Inability to culture (until very recently) the virus has limited traditional whole-cell vaccine approaches. A number of norovirus vaccines, with the major capsid protein as their target are currently in development (67). Briefly, these vaccines have used recombinant adenoviruses, virus like particles and P particles and examined multiple delivery routes including intranasal, oral and intramuscular. These include five vaccines in preclinical trials and two vaccines that have reached clinical trials in human subjects. An oral monovalent vaccine (Vaxart, Inc) against G1.1 has recently completed a Phase IB clinical trial. The vaccine was immunogenic and no safety concerns were identified. An intramuscular bivalent vaccine (Takeda Vaccines, Inc) against GI.1 and GII.4 has recently completed a Phase IIB clinical trial.

Prevention and control of norovirus outbreaks is extremely challenging. Public Health England (PHE), UK and the Centers for Disease Control and Prevention (CDC), USA have produced practical guidelines for outbreak management (68, 69). Written for a public health audience, these guidelines

are applicable to outbreaks occurring in various settings including healthcare e.g. hospitals and nursing homes, and non-healthcare e.g. cruise ships and schools. Essentially there are three principals to consider following identification of a suspected or confirmed norovirus outbreak in any setting; hand hygiene, exclusion & isolation, and environmental disinfection. Hand washing with soap and water is perhaps the single most important intervention to interrupt the chain of transmission (faecal –oral), and is the only one with a significant evidence base. The efficacy of alcohol hand rubs remains uncertain and is not recommended during outbreaks of diarrhoeal and/or vomiting illness. Exclusion and isolation are best illustrated in the healthcare setting, where rapid cohorting/isolation of symptomatic patients and exclusion of symptomatic healthcare workers is key to preventing further spread by limiting environmental contamination. This is especially important in semi-enclosed settings, such as multi-bedded wards common in the UK National Health Service (NHS). Finally environmental disinfection is necessary to reduce and ultimately remove infectious virions from fomites that act as a nidus for transmission. Key recommendations for healthcare settings include; increased frequency of cleaning, prompt cleaning of faeces or vomitus and environmental disinfection with sodium hypochlorite (PHE guidance recommend a concentration of 0.1%, or 1000ppm available chlorine). Particular attention should be paid to frequently touched surfaces e.g. bedside tables, door handles, toilet handles and taps.

1.4 ENTEROTOXIGENIC *ESCHERICHIA COLI* (ETEC)

1.4.1 Classification and nomenclature of *Escherichia coli*

Escherichia coli (*E coli*) is the type species of the genus *Escherichia*, which belongs to the *Enterobacteriaceae*. *E coli* are motile Gram-negative bacilli. The genus is named after Theodore Escherich, a German-Austrian paediatrician, who in 1885 described the organism *bacterium coli commune* (posthumously named *E coli*) following studies on the faecal flora of neonates (70). *E coli* encompass an incredibly genetically and phenotypically diverse population of bacterial species; which are commonly differentiated using the modified Kauffmann scheme on the basis of lipopolysaccharide (O), flagellar (H), and polysaccharide (K) antigens (71). The O antigen defines the serogroup, of which currently 170 are described, and the combination of O and H antigens defines the serotype, for example *E coli* O157:H7 (a serotype which serves as a marker for virulent enterohaemorrhagic *E coli* strains).

The human colonic flora is predominately anaerobic with *E coli* an important member of the remaining facultative anaerobic flora. In the immunocompetent host these non-pathogenic strains of *E coli* are harmless commensals and play an important role in intestinal physiology (72). However several pathogenic strains of *E coli* exist and depending on their host tropism may cause one of three clinical syndromes (i): urinary tract infection (ii): sepsis/meningitis and (iii): enteric/diarrhoeal disease.

1.4.2 Diarrhoeagenic *E coli*

Transmission of diarrhoeagenic *E coli* (DEC) is via the faecal oral route, through ingestion of contaminated food or water (73). Following colonisation of intestinal mucosa, various pathogenic mechanisms are responsible for causing disease (74). There are six distinct subtypes/strains of diarrhoeagenic *E coli* (i) Enterotoxigenic *E coli* (ETEC); (ii) Enteroaggregative *E coli* (EAEC); (iii) Enteroinvasive *E coli* (EIEC); (iv) Enteropathogenic *E coli* (EPEC); (v) Enterohaemorrhagic *E coli* and (vi) Diffusely Adherent *E coli* (75). ETEC are defined as *E coli* strains that can elaborate at least one member of two defined groups of enterotoxins, namely heat labile (LT) and heat stable (ST) toxins (72).

1.4.3 Pathogenesis

ETEC causes diarrhoea through the actions of several enterotoxins, including a heat-labile (LT) and two heat-stable enterotoxins: STh (human) and STp (porcine). Following colonisation of the small bowel mucosa, elaboration of these enterotoxins leads to a net secretory state (76). These enterotoxins are plasmid encoded and ETEC strains may express either or both of these toxins.

Heat-labile enterotoxin

The heat-labile toxin of ETEC is similar in structure and function to the cholera toxin (CT). This high molecular weight (85,000 DA) molecule is composed of a

pentameric B subunit that binds to cellular receptors on the surface of host enterocytes; and a single A subunit which is the enzymatically active toxic component (76). The precise mechanism of LT toxin delivery to host cellular receptors is unclear, however *in vitro* studies have demonstrated that much of secreted LT toxin remains associated with outer membrane vesicles which enter host cells via lipid raft dependent endocytosis (77).

Following initial binding of the LT-B subunit to GM1 ganglioside receptors centred in caveolae on human enterocytes, the LT toxin is internalised in vesicles by receptor-mediated endocytosis. Vesicles are transported through the Golgi and endoplasmic reticulum (ER) where the toxin is disassembled. The LT-A subunit is activated in the ER, following which irreversible activation of membrane bound adenylate cyclase takes place. This in turn stimulates production of cytoplasmic cAMP, accumulation of which activates cAMP-dependent protein kinase A (PKA), which induces phosphorylation of the cystic fibrosis transmembrane regulator (CFTR) causing a net secretion of Cl⁻ and HCO₃⁻ from the cell (78, 79). PKA also induces phosphorylation of the Na⁺/H⁺ exchanger, which inhibits Na⁺ re-absorption. These changes lead to increased fluid accumulation in the intestine and osmotically driven diarrhoea.

Heat-stable enterotoxin

Heat stable (ST) enterotoxins are small (18-19 aa) cysteine rich peptides. Two subtypes have been identified in human infection, and nomenclature taken from the animal where the toxin was first identified. STh (human) is exclusive

to human ETEC strains, and STp (porcine) is found in human, porcine and bovine strains (74). Large epidemiological studies have suggested that STh may be more pathogenic than STp (80). ST toxins are exported from ETEC to host cells through outer membrane TolC protein exporters (81).

Heat stable toxins bind to the extracellular domain of guanylate cyclase C (GC-C) on intestinal brush border epithelium. Subsequent activation of the GC-C intracellular catalytic domain leads to intracellular accumulation of cGMP. High levels of cGMP cause phosphorylation of CFTR by two mechanisms (i): activation of cGMP-dependent protein kinase II and (ii): activation of PKA (74, 82). Secretory diarrhoea results from net Cl⁻ and HCO₃⁻ loss and inhibition of cells to re-absorb Na⁺.

Colonisation factors

Colonisation of host intestinal enterocytes is an essential virulence trait for ETEC, as it facilitates direct delivery of enterotoxins. Colonisation factors (CFs) are a heterogeneous group of preproteinaceous surface structures that facilitate attachment to host intestinal mucosa. CFs are mostly plasmid encoded and at least 25 have been described to date, of which seven (CFA/I and CS1-CS6) are generally more prevalent on characterised isolates (83, 84).

1.4.4 Epidemiology and transmission

Strains of ETEC are described as a cause of gastroenteritis in two epidemiological settings; firstly as a cause of weaning infant diarrhoea in developing countries, and secondly as a cause of travellers' diarrhoea among adults from industrialised countries visiting resource poor settings.

Infant diarrhoea

Infection with ETEC is described as a major cause of weaning diarrhoea among infants in developing settings. This clinical entity is typically observed in infants aged 6-24 months where weaning from maternal milk results in exposure to new environmental pathogens, coupled with deterioration of nutrition and loss of protective maternally transferred IgA antibodies (85). In 2015 the Institute for Health Metrics and Evaluation estimated that ETEC was globally responsible for 24,000 deaths in children aged <5 years (86). In one review of sporadic endemic infantile diarrhoea worldwide the percentage of cases, which were due to ETEC varied from 10-30% (75). In endemic settings, a very low incidence of symptomatic infection is seen in school age children and adults due to mucosal immunity acquired during previous infection. However there is evidence these individuals may experience asymptomatic infections where large numbers of organisms may be shed in faeces, thereby continuing to support the environmental reservoir (87).

Travellers' diarrhoea

Travellers' diarrhoea was first described as a clinical entity in 1963 by BH Kean (88) in people returning to the United States from Mexico; and ETEC was identified as a key aetiology just over a decade later in 1975 (89).

Travellers' diarrhoea is defined as three or more loose stools per day during a trip, usually in a traveller from a more to a less economically developed setting, and at least one other constitutional symptom e.g. fever, abdominal cramps, nausea or vomiting (90). During a two week journey, the incidence of travellers' diarrhoea ranges from 10-40%, and is dependent upon the destination and traveller characteristics (91). Lower budget and younger travellers, for example backpackers appear to be at higher risk, which may be due to a more adventurous nature and consumption of high-risk street food (92). The treatment and prevention of travellers' diarrhoea is discussed separately.

Transmission

Transmission of ETEC is transmitted via the faecal oral route, typically by ingestion of contaminated food and water. Although the infectious dose is relatively high, so too is the degree of environmental contamination in endemic settings. In endemic areas, ETEC infections display a distinct seasonality, with peaking during wet and warm months, when environmental conditions are favourable for proliferation of bacterial pathogens in food and water (72).

1.4.5 Immunity

Clinical observations and epidemiological studies suggest that in developing settings, the highest burden of ETEC infection is in children less than three years of age, and that rates of infection decline with increasing age (93). This is in contrast to ETEC infection in travellers from developed countries to endemic settings where the majority of infections are seen in adults. These observations suggest that there is an element of natural immunity following infection. In developing settings where infection is endemic, this is more likely acquired during infancy. Further supporting this observation are data from vaccine safety and immunogenicity studies that show protection from ETEC infection following re-challenge with the same ETEC strain (94).

The immune response to ETEC has been studied in both natural and human volunteer infection. Initial infection may provide highly specific protection against the homologous strain or strains expressing similar or related CFs (95). Protective immunity likely results from production of mucosal antibodies in the small intestine that prevents bacterial binding and toxin activity on epithelial cells. Following ETEC infection there is an intestinal IgA secretory response and a systemic IgA and IgG antibody response against the major virulence factors; CFs, LT and O antigen of the infecting strain (96); the immune response to ST is poor as this small molecule is weakly immunogenic.

1.4.6 Clinical features

Clinical observations from volunteers and patients infected with ETEC in endemic settings have demonstrated gastroenteritis characterised by abrupt onset watery, non-bloody diarrhoea and abdominal cramping with a short incubation period of 14-50 hrs (87, 97). Fever and vomiting may occur in a minority of cases. A wide spectrum of diarrhoeal disease is seen, ranging from mild self-limiting to a severe purging, which is characteristic of infection with *Vibrio cholerae* (98), which would also be a differential diagnosis. In otherwise healthy immunocompetent individuals, diarrhoea is usually self-limiting; however infection can be fatal in weaning infants in resource poor settings, due to rapid dehydration and electrolyte disturbance. In developing settings this situation is exacerbated by limited access to healthcare, clean water and oral rehydration therapy. The use of antimicrobials in these settings is complicated by limited resources and where available from increasing rates of antimicrobial resistance (93).

1.4.7 Diagnosis

Identification of ETEC requires that it be differentiated from other non-pathogenic strains of *E coli* that constitute the normal human colonic flora. Detection of traditional serotypic markers does not correlate well enough with strains of diarrhoeic *E coli* to allow definitive identification, and detection of ETEC CFs has proved impractical given their great number and heterogeneity. Therefore detection of ETEC relies on detection of the

enterotoxins LT and/or ST. Initially, physiologic assays were used as gold standard diagnostic modalities, specifically the rabbit ileal loop model (99) for LT, and the infant mouse assay (100) for ST. However, as well as issues with animal welfare, these techniques proved cumbersome and were later replaced by simpler assays including tissue culture to detect LT, namely the Y1 adrenal cell and Chinese hamster ovary cell lines (101, 102) and radioimmunoassay for ST (103). Since the advent of molecular diagnostic techniques in the early 1980s, diagnosis has moved away from more traditional phenotypic techniques to molecular methodologies including DNA probes and PCR (104). An advantage of PCR is that pathogens can be detected directly from faecal material and reactions can be multiplexed to allow for simultaneous detection of multiple toxins, as well as other diarrhoeagenic E.coli.

1.4.8 Treatment and Prevention

Paediatric diarrhoea in developing settings

An overview of the management of dehydrating diarrhoea in infants in developing settings is presented in section 1.3.11.

Travellers' diarrhoea

Treatment of travellers' diarrhoea whether caused by ETEC or not is often unnecessary given the self-limiting nature of this clinical syndrome. Where symptoms are severe, treatment goals include: prevention of dehydration and

electrolyte disturbance; reducing severity and duration of symptoms and restoration of normal activity for the afflicted traveller. In general all that is required is increased fluid intake with sugary drinks and salty crackers (105), however commercially available oral rehydration sachets may be used provided there is a readily available supply of clean drinking water, or facilities to purify contaminated supplies. Loperamide may have a role in symptomatic relief and is the antimotility drug of choice. However it should be used with caution in cases where there is bloody diarrhoea or high fever, often associated with pathogens that invade the intestinal wall such as *Shigella* and *E coli* O157. Antibiotics have been demonstrated to shorten the duration of symptoms in moderate to severe travellers' diarrhoea by about a day and a half. Choice of agent is dependent on location; quinolones e.g. ciprofloxacin are usually an appropriate choice for most destinations, however with increasing resistance in *Campylobacter* spp. in South and South East Asia, a macrolide such as azithromycin may be a better choice (91).

Vaccines

From a public health perspective, vaccination offers the best opportunity to impact upon ETEC disease burden in developing settings; where in the short to medium term at least, there lacks the economic and political means to eradicate enteric disease through effective sanitation and infrastructure alone. Indeed, an ETEC vaccine is a top priority for various public health agencies including the WHO and United Nations Children's fund (106). There is no ETEC-specific vaccine in development that has yet demonstrated clinically significant efficacy. In theory, there should be some cross protection from

cholera vaccine due to the antigenic similarity of the cholera (CT) and ETEC LT enterotoxins, but following a Cochran review (107) there was insufficient evidence to support the use of the oral cholera vaccine Dukoral® in protecting travellers against ETEC diarrhoea.

Vaccine development has taken two approaches using whole cell and subunit model systems:

- (i): block ETEC attachment and colonisation to host receptor cells by inducing immunity against CFA adhesins, and
- (ii): prevent toxin-mediated diarrhoea by inducing immunity against ETEC LT and ST enterotoxins.

Whole cell vaccines

The first experimental ETEC vaccine, colicin E2-inactivated ETEC prototype strain H10407 (O78:H11, LT⁺ STa⁺ CFA/I⁺), successfully induced IgA antibodies against CFA/I and LT following oral challenge in healthy volunteers (108). Unfortunately protection was against the homologous strain only, but provided proof of concept. Later and more widely studied whole cell candidates including rCTB-CF and ACE527 have combined multiple killed or live attenuated strains (94, 109). However these have met with significant issues, including poor protection for very young children in endemic settings, live products associated with adverse effects in volunteers and inducing protection against LT toxin antigen only. Research is ongoing with whole cell

vaccines to try and address issues of protective efficacy and side effect profile.

Subunit vaccines

In concept at least, ETEC subunit vaccines look promising, as it may be possible to induce immune responses to specific, well defined and characterised immunogens. CFA adhesins and LT have been primary targets in development to date, this is despite recent data to suggest the LT antigen alone is unlikely to be effective, as it does not provide cross protection against STa, which most strains express (6). Major hurdles remain however; ETEC strains express at least 23 immunologically heterogeneous CFAs and unlike LT the small STa peptide (19 amino acid residues) is poorly immunogenic. Conjugation to improve immunogenicity and structural modifications to reduce toxicity have proved technically very challenging.

Current research is exploring several avenues (110), firstly to examine the role of ST-LT toxoid fusion polypeptides, which have shown some promise *in vitro*. Secondly, using CFA adhesion tip antigen vaccines, which have demonstrated cross protection against different CFAs; currently the US Naval medical research centre is undertaking a phase 2b human immunization and challenge study of such a vaccine.

1.5 Aims and objectives

Despite the introduction of rotavirus vaccination, diarrhoeal disease remains among the top causes of young child (<5 years) mortality in the world's least developed countries (1). Both norovirus and ETEC are recognised as leading viral and bacterial causes respectively, of non-bloody dehydrating childhood diarrhoea in low income, high mortality settings, particularly sub-Saharan Africa, and therefore warrant further investigation (4). The advent of sensitive nucleic acid based diagnostics presents an opportunity to examine these pathogens in more detail than previously possible, both to better describe their epidemiology and inform future vaccines as they become available. This study has provided an opportunity to examine both these important enteric pathogens in a less developed, high disease burden paediatric population in Africa, where other pathogens, notably rotavirus (111) and *Campylobacter* (112) have already been investigated.

Previously I examined a cohort of 1941 diarrhoeal stool samples collected from hospitalised children in Blantyre, Malawi (1997-2007) for norovirus using real-time PCR. I found a norovirus prevalence in this population of 11.3%, and interestingly it appears strains reported to be associated with large norovirus epidemics in 2002 in Europe and the US were circulating in Malawi as early as 1999. I plan to further examine and define epidemiologically important norovirus strains identified in this cohort. In addition I plan to further examine the same cohort of stool samples for ETEC using molecular methods, in order to more fully define the disease burden in this hospitalised population of

children. Other researchers at the Institute of Infection and Global Health have previously examined this cohort of stool samples for rotavirus and *Campylobacter*.

Aims:

1. To determine the prevalence of ETEC in a cohort of children hospitalised with diarrhoea in Blantyre, Malawi (1997-2007), and;
 - a) develop and optimise a molecular based (PCR) method for detection of ETEC in diarrhoeal stool samples and control strains.

2. To further define epidemiologically important norovirus strains in a cohort of children hospitalised with diarrhoea in Blantyre, Malawi (1997-2007) and children and adults in the UK (1993-2009) and explore their importance from a global disease perspective.

2 MATERIALS AND METHODS

2.1 Specimens

Faecal specimens or cDNA from them used to undertake the work presented in this thesis were collected from four previous studies (Table 1).

Table 1. Origin of faecal samples or cDNA used for research presented in this thesis.

Study name	Infectious intestinal disease study in England (IID1) (113)	Malawi studies (111, 114-116)	Structured surveillance of infectious intestinal disease in pre-school children in the community (Nappy Study) (117)	Second study of infectious intestinal disease in the community (IID2) (46)
Sampling years	1993-1996	1998-2002	2007-2008	2008-2009
Countries	England	Malawi	England	UK
Number of samples	87	1941	19	185
Enteric pathogens examined in this thesis	Norovirus	Rotavirus Norovirus Campylobacter ETEC	Norovirus	Norovirus

IID1

Faecal samples from the infectious intestinal disease study in England (IID1) were obtained from cases of gastroenteritis in all age groups in the community or attending general practice, and healthy controls in the community between

1993-1996. The aims of this study were to determine the incidence and aetiology of infectious intestinal disease in the community and presenting to general practice, and to compare this with cases captured by national laboratory surveillance.

Nappy study

Faecal samples from the structured surveillance of infectious intestinal disease in pre-school children in the community (Nappy study) were obtained from children < 5 years with gastroenteritis attending general practice in England and Wales between 2007-2008. The aims of this study were to monitor the prevalence of infectious intestinal disease in pre-school children presenting to general practice and determine the burden of rotavirus disease to inform policy on childhood rotavirus vaccination.

IID2

Faecal samples from the second study of infectious intestinal disease in the community (IID2) were obtained from cases of gastroenteritis in the community or attending general practice in the UK between 2008-2009. The aims of this prospective community cohort study were to estimate overall and by organism, the incidence of infectious intestinal disease in the community, presenting to general practice and reported to national surveillance. This was to assess progress towards the UK Food Standards Agency task to reduce food-borne illness by 20% in its first five years of operation and also to

determine whether changes in healthcare provision might influence the interpretation of national statistics.

Malawi studies

Initially, faecal samples were collected as part of a two-year study (118) examining the extent of strain diversity among rotaviruses in Malawi, and the clinical presentation and outcome following infections with rotavirus, astrovirus and adenovirus 40/41 in children with and without HIV infection in Malawi. Subsequent samples were collected over an eight year period from the same population as part of on-going surveillance to monitor rotavirus strain diversity before implementation of a childhood rotavirus vaccine programme.

The QECH is a large, government-run, tertiary referral centre serving a predominantly urban population in Blantyre and its surrounding areas. Malawi is a land-locked developing country located in sub-Saharan southeast Africa. It is bordered by Tanzania to the northeast, Mozambique to the east, south and west and Zambia to the northwest. Malawi's largest city and capital is Lilongwe; the second largest city and financial centre is Blantyre. Malawi has a sub-tropical climate; with a warm wet season which runs from November to April and a cool dry season from May to August and a hot, dry season from September to October. In 2012 the population of Malawi was 15,906,000, gross national income per capita \$730, life expectancy at birth male/female 58/60 and infant mortality under five years 71/1000 live births (119).

2.2 Subject enrolment

As previously described (118). During the initial two-year study in Malawi, two research nurses collected faecal samples from children with acute diarrhoea who attended the under five year rehydration clinic at the hospital (outpatients) and children admitted to the paediatric wards (inpatients). Inclusion required age <5 years; a primary diagnosis of acute gastroenteritis (diarrhoea defined as at least three loose stools in 24 hours for less than two weeks) and moderate/severe dehydration. For inpatients, only children who were able to provide a stool sample within the first 48-hours following admission were enrolled. Informed, written consent was obtained for each subject from a parent or guardian prior to enrolment.

As a comparison group during the initial two-year study, faecal specimens were additionally collected from children admitted to the paediatric wards (inpatients) with a primary diagnosis other than gastroenteritis e.g. pneumonia, meningitis and malaria. Inclusion required: age <5-years and no history from the parent or guardian of diarrhoea in the preceding two weeks prior to admission. Although selected from the same age group, this comparison group was not matched for age or season. Following the initial two-year study, faecal samples continued to be obtained as described above, however as there was then only one research nurse involved, the intensity and consistency with which subjects were enrolled was reduced.

2.3 Specimen storage

Faecal specimens were labelled with a unique patient identifier and date of collection. Inpatient diarrhoeal specimens collected during the initial two year study were given a prefix of MW1, outpatients OP1 (these samples were not included in the research and analysis conducted as part of this thesis) and 'control' patients (comparison group without diarrhoea) CON1. Subsequent diarrhoeal samples collected over the course of the following eight years were given a prefix of MW2 and QEC. Faecal samples were collected in sterile containers with the aid of a spoon. On arrival in the laboratory, each stool sample was placed into a single, sterile, cryovial and faecal supernates were prepared for immediate rotavirus testing (as described previously and below). Remaining faecal samples were frozen at -80°C and transferred to the University of Liverpool for later testing as described below.

2.4 Preparation of faecal suspensions

As described previously (118). Approximately 10-20% faecal suspensions were prepared in 0.01M phosphate-buffered saline (PBS), pH 7.4. PBS solution was prepared by dissolving a single PBS tablet (Sigma, Poole, UK) in 50ml of distilled water. Approximately 100-200µl of liquid faeces (or a pea-sized portion of solid stool) was mixed thoroughly in a 1.5ml Eppendorf tube with 900µl of PBS, pH 7.4 by vortexing in a Class I Biosafety Cabinet. The mixture was then spun at 13000 rpm for 10 minutes in a bench-top microcentrifuge. The supernatant was pipetted off and decanted into a clean 1.5ml Eppendorf tube. 100µl of the 10-20% supernate was used immediately for rotavirus testing. The remainder was stored at -80°C until used for

adenovirus testing, astrovirus testing, rotavirus subgrouping and rotavirus dsRNA extraction, norovirus testing, campylobacter testing, and ETEC testing.

2.5 Identification of enteric pathogens

The identification of enteric pathogens relevant to this thesis is outlined below, due reference is given in those instances where previous work and the work of others was involved.

2.5.1 Enteric viruses

2.5.1.1 Rotavirus

As previously described (118), a commercially available ELISA (Rotaclone, Meridian Diagnostics, Cincinnati, OH, USA) was used to detect rotavirus antigen in 10-20% faecal supernates, according to manufacturer's instructions.

2.5.1.2 Norovirus

Nucleic acid extraction, reverse transcription and norovirus detection

For specimens collected in all studies, nucleic acid was obtained directly from faecal material, converted to randomly-primed cDNA and established as

norovirus-positive using previously described methods: IID1 study (113), IID2 study (46, 120), The Nappy study (117).

In more detail my own previous work at the Institute of Infection and Global Health, University of Liverpool examined the Malawi samples for norovirus. As previously described (121), nucleic acid was extracted from 10 % faecal suspensions in PBS using a high throughput automated extractor (QIAAsymphony ®, Qiagen UK). Reverse transcription of RNA was performed using random hexamers. Norovirus genogroups GI/GII were detected using a real-time PCR technique described by Kageyama et al (122) as modified by Amar et al (123).

Amplification of the GII.4 norovirus P2 domain for strain characterisation

In order to determine GII.4 strain types, the region of the ORF2 gene encoding the VP1 hypervariable P2 domain was amplified by PCR as previously described for samples collected from the UK (24, 124) and sequenced using a dideoxy chain-terminator method. For nucleic acid extracts from the Malawi studies the following methods were used.

Principle

Viral RNA was reverse transcribed to cDNA by using tag-primer TX30SXN and used as a template for PCR amplification using primer pair NVF7/NVR7 to obtain product that contains the norovirus P2 domain (Table 2).

Table 2. Primers used for amplification of the GII.4 norovirus P2 domain. Key:
nt = nucleotide position.

Primer	Sequence	Location (GII.4 Lordsdale sequence)	Reference
TX30SXN	GAC TAG TTC TAG ATC GCG AGC GGC CGC CCT TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT	3' end poly T	(125)
NVF7	AAT GCT GTA CAC ACC ACT TAG	nt 5624-5644	(126)
NVR7	GAA GTG CTG CAC CCA TTC CT	nt 6506-6487	(126)

Procedure

All PCR reaction mixes were prepared in a dedicated clean room using dedicated equipment (pipettes and tips). The PCR reaction mixtures were prepared as follows:

Denature

1x Denature (μ l components)

TX305XN (10 μ M)	1.0
dNTPs (10mM)	1.0
RNA	<u>8.0</u>
	10.0

Temperature conditions

65°C, 5 minutes

Reverse transcription

1x Reverse transcription (μ l components)

5x buffer	4.0
RNasin	1.0
DIM DTT	1.0
10mM dNTPs	1.0
50mM MgCl ₂	0.5
ssRT (Invitrogen)	0.5
H ₂ O	<u>2.0</u>
	10.0

Temperature conditions

42°C, 60 minutes

70°C, 15 minutes

P2 domain PCR

1x PCR (μ l components)

H ₂ O	34.25
10x buffer	5.0
50mM MgCl ₂	1.0
10mM dNTPs	1.0
10 μ M NVFR7	1.5
10 μ M NVR7	1.5
Taq polymerase (Promega)	0.75
cDNA	<u>5.0</u>
	50

PCR cycling conditions

1 cycle:	94°C, 5 minutes
35 cycles:	94°C, 30 seconds
	50°C, 1 minute
	72°C, 3 minutes
1 cycle:	72°C, 7 minutes

Sequencing

DNA was purified using a commercially available spin column, Illustra MicroSpin™ S-400 HR columns (GE Healthcare Life Sciences) according to the manufactures instructions prior to direct two-way sequencing by a commercial company, Cogenics (Beckman Coulter, Genomics, UK).

Phylogenetic analysis

Sequence analysis was performed using Bionumerics v6.1 (Applied Maths, Kortrijk,Belgium) and Molecular Evolutionary Genetic Analysis (MEGA) version 5 (127) to generate bootstrapped dendrograms from ClustalW predicted amino acid (aa) alignments using the neighbour-joining (NJ) and maximum parsimony (MP) methods. Amino acid motifs were designated by standard IUPAC single-letter amino acid code. Reference strains obtained from NCBI GenBank were included for comparison, accession numbers: AF145896, X86557, AY532115, AJ004864, AY532127, AY587988, AY502023, AB22092, EF126965, AB445395, GU445325, JX459908, DQ078794, EF126963. Sequences described in this study were deposited in Genbank; accession numbers KU312298-KU312306.

2.5.2 Enteric bacteria

2.5.2.1 *Campylobacter*

As previously described (112), following nucleic acid extraction from 10% faecal suspensions in PBS using a high throughput automated extractor (QIASymphony®, Qiagen UK), real time PCR was used to detect the *mapA* gene of *Campylobacter jejuni* and the *ceuE* gene of *Campylobacter coli*.

2.5.2.2 Enterotoxigenic *E coli*

Following nucleic acid extraction from 10% faecal suspensions in PBS using a high throughput automated extractor (Qiasymphony, Qiagen UK) as previously described (121), a qualitative PCR was used to detect ETEC heat-labile (LT) and heat-stable (STh and STp) enterotoxin genes according to a previously described method (128). ETEC strains NCTC 11602 and 11603 (obtained from PHE) were used as assay controls. Table 3 shows the toxigenic profile of the ETEC control strains.

Table 3. Toxigenic profile of ETEC Control strains.

ETEC control strain	LT	STh	STp
NCTC 11602	✓	✓	
NCTC 11603			✓

Genomic DNA purification from ETEC control strains

Genomic DNA was extracted from overnight broths of the ETEC control strains NCTC 11602 and 11603 and according to the manufactures instructions (Promega Wizard®).

Qualitative PCR for detection of ETEC

Detection of ETEC heat labile (LT) and heat stable (STh and STp) enterotoxins direct from faecal samples was accomplished following nucleic acid extraction from 10% faecal suspensions as previously described (121), and by adopting a previously published qualitative PCR technique with primers specific for the LT, STh and STp genes of ETEC (Table 4).

Optimisation of ETEC PCR sensitivity and extraction methods was achieved by examining tenfold serial dilutions of genomic DNA (from ETEC control strains NCTC 11602 and 11603) spiked in healthy volunteer stool. Specificity of the ETEC PCR assay was determined by using an in-house strain of *Salmonella enterica* serotype *enteritidis*.

Reagents

Autoclaved HPLC H₂O

10x PCR buffer

50mM MgCl₂

10mM dNTP solution

Taq polymerase (Promega)

ETEC primer mix (200 μ mol/ μ L) Sigma

Template (faecal suspension nucleic acid extract)

Procedure

All PCR reaction mixes were prepared in a dedicated clean room using dedicated equipment (pipettes and tips). Addition of template was done on a DNA processing bench. Each ETEC PCR experiment examined template from 45 individual cases, and included as positive controls the ETEC strains NCTC 11602 and 11603, and nuclease free water as a single negative control. The PCR reaction mixture was prepared as follows:

1x PCR (ul components)

H ₂ O	17.0
10x PCR buffer	2.5
50mM MgCl ₂	1.0
10mM dNTP	1.0
Taq polymerase	0.5
ETEC primer mix	1.0
Template	<u>2.0</u>
	25uL

PCR cycling conditions

1 cycle:	94°C, 5 minutes
30 cycles:	94°C, 30 seconds
	63°C, 30 seconds
	72°C, 1.5 minutes
1 cycle:	72°C, 5 minutes

Agarose gel electrophoresis of PCR products

PCR amplified fragments were separated on 1.5% agarose gels stained with ethidium bromide and visualised under UV light. A DNA ladder (Thermo Fisher Scientific, Invitrogen) was incorporated to allow estimation of target fragment sizes.

Table 4. Primers used for detection of ETEC

Target gene	Primer	Sequence (5` to 3`)	Product size (bp)
<i>LT</i>	MP2-LT-F	GAACAGGAGGTTTCTGCGTTAGGTG	655
	MP2-LT-R	CTTTCAATGGCTTTTTTTTTGGGAGTC	
<i>STp/estla</i>	MP4-ST1a-F	CCTCTTTTAGYCAGACARCTGAATCASTTG	157
	MP4-ST1a-R	CAGGCAGGATTACAACAAAGTTCACAG	
<i>STh/estlb</i>	MP2-ST1-F	TGTCTTTTTTACCTTTCGCTC	171
	MP2-ST1-R	CGGTACAAGCAGGATTACAACAC	

2.6 Statistical analysis

Categorical data were analysed using Chi-square test on IBM SPSS STATISTICS (version 20). A p value ≤ 0.05 was used to determine significance.

2.7 Ethics

At the time of sample collection, informed consent was obtained from patients or their guardians (if <18 years old) for storage of faecal samples for the purposes of enteric pathogen research. The Malawi Health Sciences and Research Committee gave ethical approval for the collection and use of faecal material. The IID Study Scientific Steering Committee granted approval for use of archive material. Faecal sample and data archives at the Institute of Infection and Global Health, University of Liverpool (Malawi study) and The Enteric Virus Unit, PHE (Nappy Study and IID studies) are stored in compliance with the UK human tissue act 2004 and Caldecott principles.

3 Enterotoxigenic *E coli* in hospitalised children with and without diarrhoea in Blantyre, Malawi

3.1 Introduction

Diarrhoeal disease is a leading cause of morbidity and mortality in children <5 years of age in low income countries, and after neonatal deaths is second only to respiratory tract infection (1, 129). Aetiology of childhood diarrhoea is diverse, including viral, bacterial and parasitic agents. The epidemiological importance of these various gastrointestinal pathogens in children in developing countries is poorly described, especially in Sub-Saharan Africa. Recently, reanalyses of two large diarrhoea aetiology studies (7, 9) using sensitive molecular diagnostic panels has identified that most attributable cases of diarrhoea in children <5 years in low income countries, including Africa and Asia, are due to four pathogens: *Shigella*; rotavirus; adenovirus 40/41, and ETEC.

Rotavirus is the leading cause of childhood diarrhoeal disease worldwide. However with the introduction of an effective rotavirus vaccine into many childhood immunisation programmes, the burden of rotavirus gastroenteritis should decrease. Monovalent rotavirus vaccine (RV1) was introduced in Malawi in October 2012, and since then it has been estimated that diarrhoea associated mortality has fallen by 31% in this paediatric population (130). In 2016 it was estimated that ETEC was the eighth leading cause of diarrhoea deaths globally in all age groups. Among bacterial causes of diarrhoea, ETEC

is one of the most important causes in children in low and middle-income countries. It was estimated in 2016, in children <5 years that 18,669 diarrhoea deaths (4.2% of all diarrhoea deaths) were due to ETEC, with the greatest burden seen in eastern sub-Saharan Africa (86). ETEC is also commonly implicated as a cause of diarrhoea in adult travellers from higher income countries visiting endemic settings (105). Unfortunately, no broadly protective vaccine against ETEC is currently available; however there are on-going research efforts in this area (131, 132). The WHO has identified an urgent need for vaccine development against ETEC as an important primary prevention strategy to reduce mortality from childhood diarrhoeal disease in developing countries (110). Vaccination remains the most effective approach to reduce mortality in low income settings, where millions of children currently have no access to safe water and sanitation infrastructures. However, vaccine development is dependent on an in depth understanding of the epidemiology of ETEC in the most affected populations. This issue is complicated by the fact that the aetiology of diarrhoeal disease is multi-factorial and accurate surveillance and determination of disease burden is dependent upon appropriate laboratory facilities being readily available in these settings.

ETEC causes diarrhoea through the actions of several enterotoxins, including a heat-labile (LT) and two heat-stable enterotoxins STh (human) and STp (porcine) (76). Detection methods for ETEC in faecal specimens have evolved over the past several decades from physiologic assays to more recent molecular methodologies including DNA probes and PCR (75).

3.2 Aims

The primary aims of this study were to determine the (a) prevalence, (b) age distribution, (c) seasonality and (d) toxigenic profile of circulating ETEC strains in hospitalised children (<5 years) in Blantyre, Malawi over a 10-year period.

A secondary aim was to explore mixed enteric infections in this paediatric population using data presented in this thesis and elsewhere.

3.3 Study design

A 10-year collection (July 1997 – June 2007) of faecal samples obtained from hospitalised children (<5 years) with and without diarrhoea in Blantyre, Malawi has previously been examined for rotavirus, norovirus and campylobacter (111, 112, 121). Remaining faecal samples were reused to detect ETEC heat labile (LT) and heat stable (STh and STp) enterotoxins by conventional PCR. Further detail on the materials and methods used to undertake this study are outlined in Chapter 2.

3.4 Results

3.4.1 Baseline characteristics of the study population

Faecal samples collected between July 1997-June 2007 were available for a total of 2448 children, comprising 1941 samples from children with diarrhoea and 507 without. For the two-year period (30th July 1997- 30th June 1999)

during which samples were collected from both children with and without diarrhoea, 1166 faecal samples were available, comprising 659 samples from children with diarrhoea and 507 without. Age and sex demographics for each of the population cohorts are presented in Table 5.

Table 5. Age and sex demographics for hospitalised children aged < 5 years with and without diarrhoea in Blantyre, Malawi 1997-2007.

10-year (1997- 2007)				2-year (1997-1999)		
Age (Months)	Mean	Median	Range	Mean	Median	Range
All	10.6	8.8	0.1-55.4	10	7.9	0.1-55.4
Diarrhoeic	10.8	9.2	0.1-55.4	10.4	8.5	0.1-55.4
Non-diarrhoeic	/	/	/	9.6	6	1.0-50
Sex	Male	Female	/	Male	Female	/
All	1338	1110	/	660	506	/
Diarrhoeic	1073	868	/	395	264	/
Non-diarrhoeic	265	242	/	265	242	/

3.4.2 Prevalence

A total of 1941 diarrhoeal samples collected from children <5 years of age, during a 10 year period from 1997-2007 were available for testing. The prevalence of ETEC was 10.6% (205/1941). For the two-year period (30th July

1997-30th June 1999) during which samples were collected from both children with and without diarrhoea, the detection rate of ETEC was significantly higher in children with diarrhoea than in those without (84/659 (12.7%) vs. 37/507 (7.3%) respectively; $p=0.002$).

3.4.3 Age distribution

The median age of children in the diarrhoeic group with ETEC infection was 10 months (range, <1-40 months). ETEC detections were observed in all age groups with a marked peak in the 6-11 month age group (Figure 2). When cases where ETEC was detected in the presence of another enteric pathogen (rotavirus, norovirus and *Campylobacter*) were excluded, no significant differences were seen in the proportions of ETEC detections by age group (Figure 3).

3.4.4 Seasonality

Detection of ETEC demonstrated circulation throughout the year with peaks in Jan and March and Sep-January (Figure 4).

3.4.5 Toxigenic profile of circulating ETEC strains

In children with diarrhoea, the most prevalent toxin type was STh (6.6%), followed by LT (2.1%) and STp (0.9%). A total of 19 (1%) specimens

contained mixed toxin types. A similar toxigenic profile was observed in children without diarrhoea (Tables 6 and 7).

3.4.6 Mixed Infections

Mixed or co-infections in which another enteric pathogen (rotavirus, norovirus or *Campylobacter*) was detected in association with ETEC were significantly higher in children with diarrhoea at 65.3% (134/205) compared to children without diarrhoea 40.5% (15/37) ($p < 0.004$). When mixed infections were excluded, there were no differences in the proportions of ETEC-detections in diarrhoeal or non-diarrhoeal samples (4% in both groups). Rotavirus co-infections were the most commonly seen among diarrhoeal cases (Figure 5). In ETEC positive diarrhoeal samples 37% (76), [29% (24) in the 2-year comparison period] were also positive for rotavirus with or without norovirus and/or *Campylobacter* compared with only 8% (3) of the non-diarrhoeal samples.

Figure 2. Age distribution of ETEC detections in children <5 years hospitalised with diarrhoea in Blantyre, Malawi 1997-2007 (m = months).

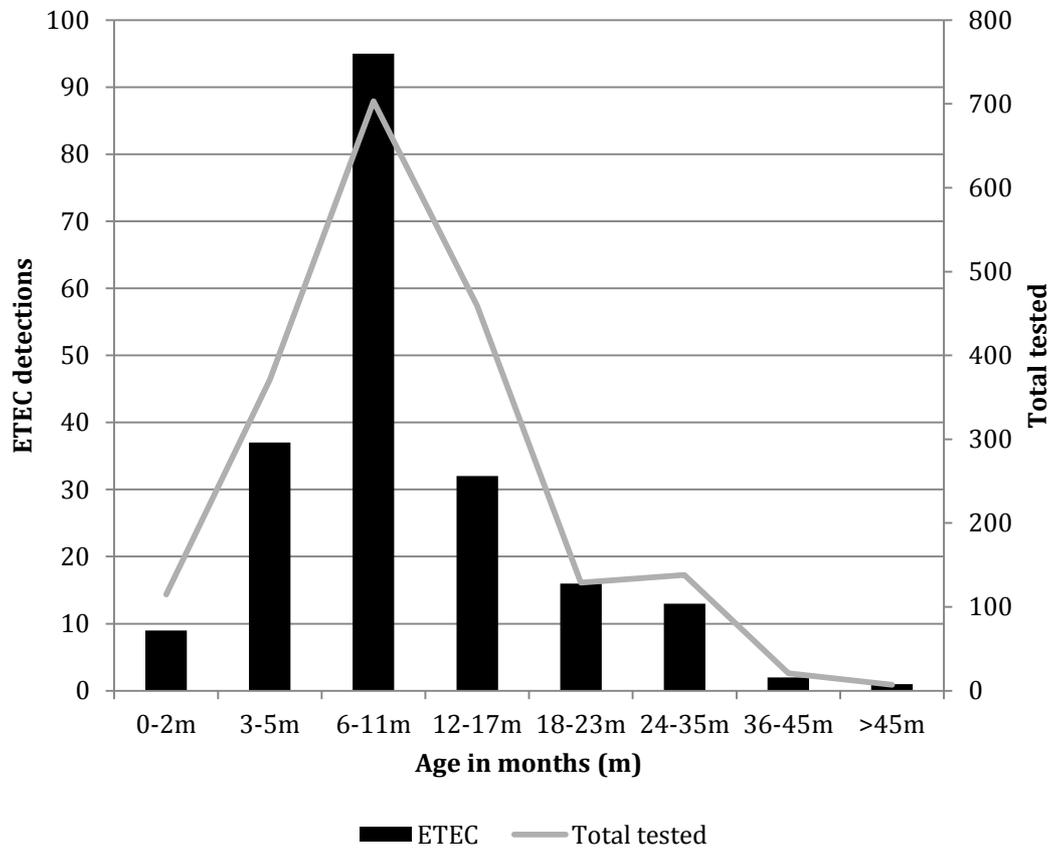


Figure 3. Age distribution of ETEC positive diarrhoeal children (m = months).

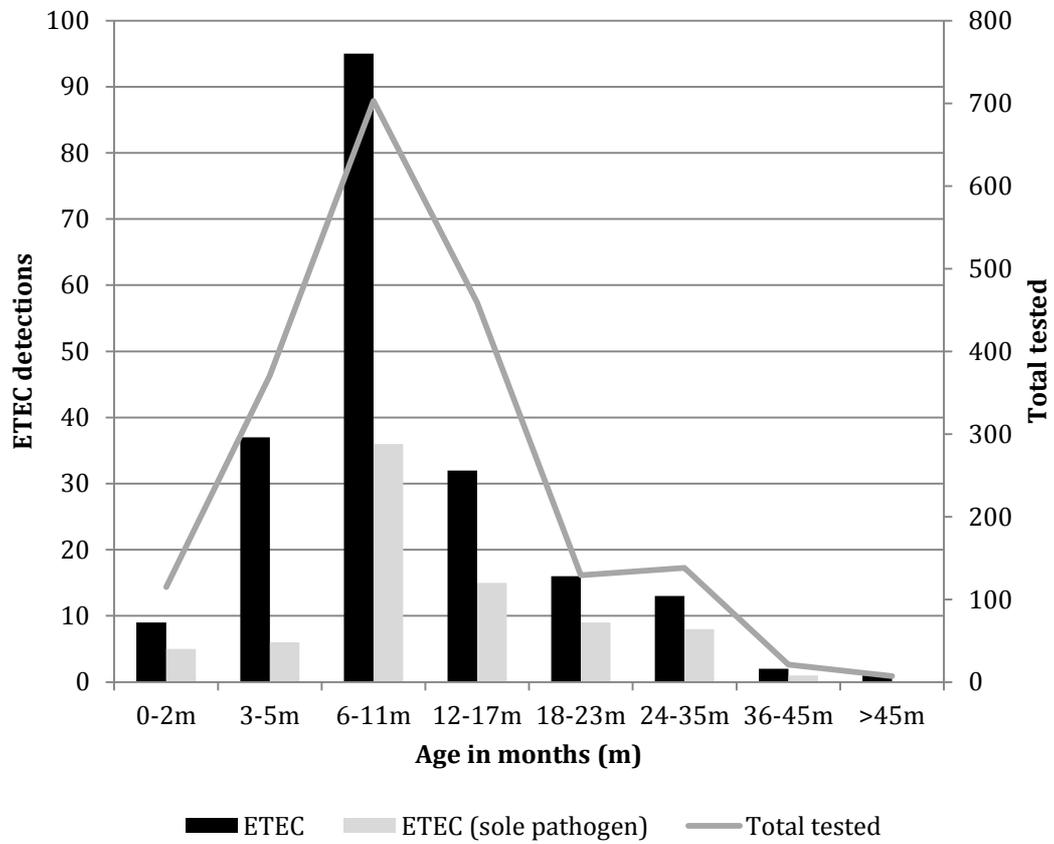


Figure 4. ETEC detections by month in children <5 years hospitalised with diarrhoea in Blantyre, Malawi 1997-2007.

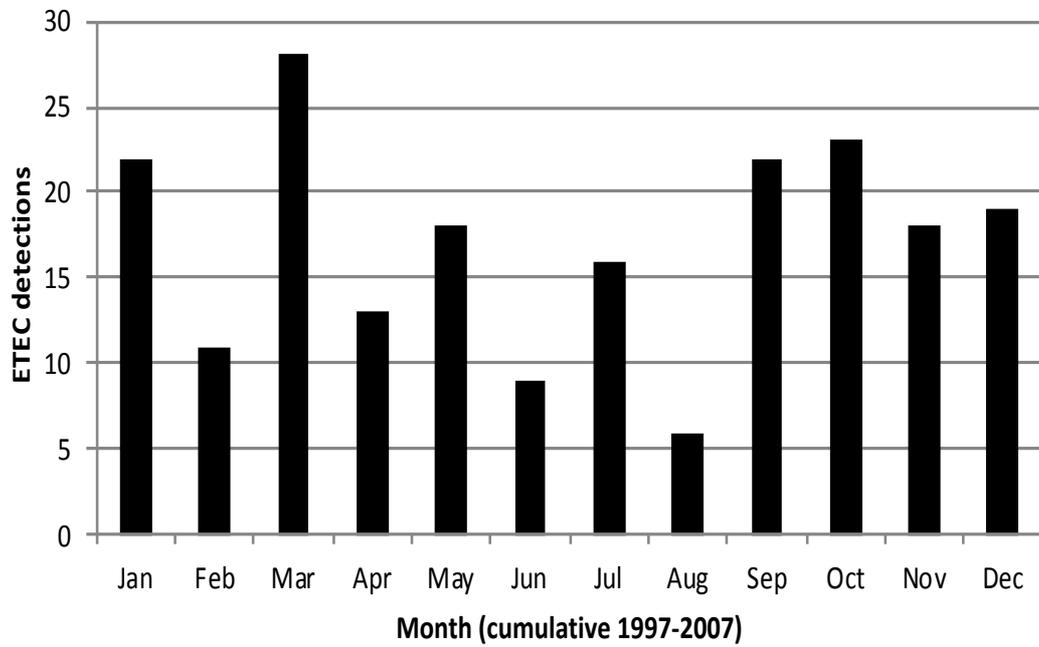


Figure 5. ETEC detections (%) in the presence or absence of rotavirus (RV) (111), norovirus (NoV) and/or *Campylobacter* (112), among stool samples collected from symptomatic and asymptomatic children.

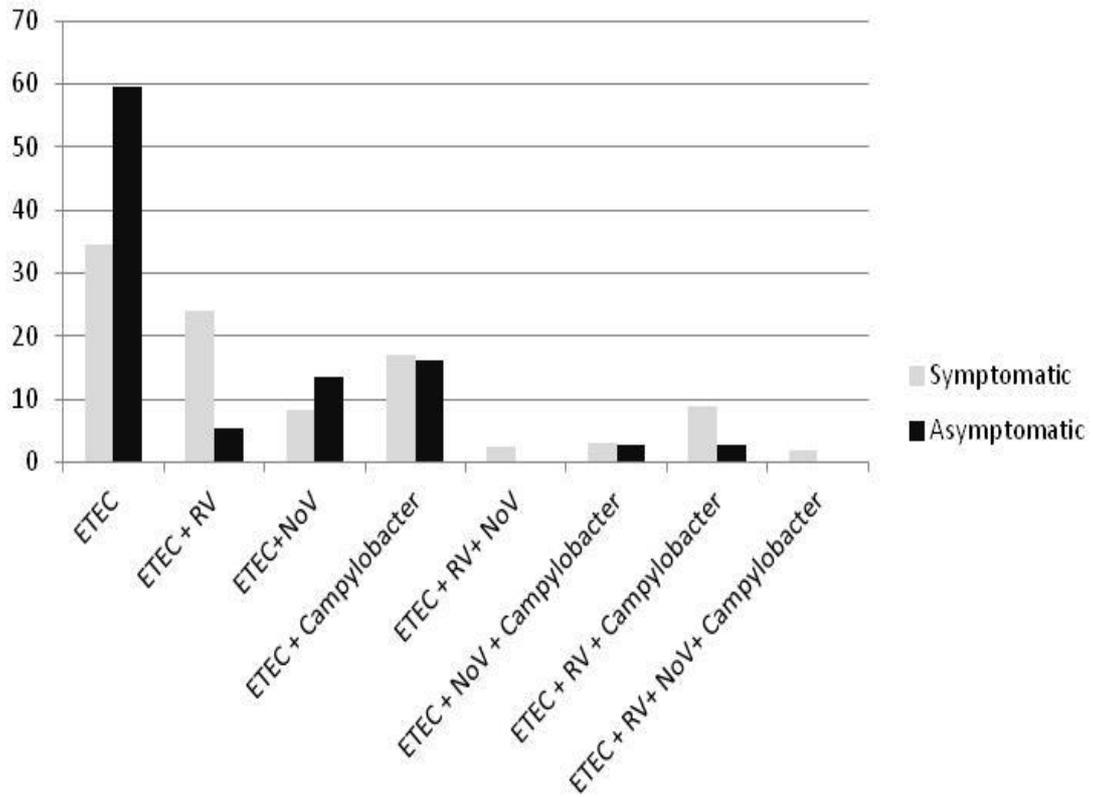


Table 6. Detection of ETEC and toxigenic profile of circulating strains in children <5 years hospitalised with diarrhoea in Blantyre, Malawi 1997-2007.

Group	ETEC Toxin	No detected (%)
Diarrhoeic 1997-2007	All	205 (10.6)
	LT	40 (2.1)
	ST	146 (7.5)
	STh	129 (6.6)
	STp	17 (0.9)
	Mixed	19 (1.0)
	LT/STh	16 (0.8)
	LT/STp	3 (0.2)

Table 7. Detection of ETEC and toxigenic profile of circulating strains in children <5 years hospitalised with and without diarrhoea in Blantyre, Malawi 1997-1999.

ETEC toxin	Diarrhoeic (%)	Non-diarrhoeic (%)	Chi-square (p-value)
All	84 (12.7)	37 (7.3)	9.147 (p=0.002)
LT	15 (2.3)	5 (1)	
ST	60 (9.1)	28 (5.5)	
STh	57 (8.6)	25 (5)	
STp	3 (0.5)	3 (0.6)	
Mixed	9 (1.4)	4 (0.8)	
LT/STh	8 (1.2)	1 (0.2)	
LT/STp	1 (0.2)	3 (0.6)	

Key:

LT: Heat labile toxin

ST: Heat stable toxin

STh: Human heat stable toxin

STp: Porcine heat stable toxin

3.5 Discussion

This 10-year study is one of the largest to date examining the role of ETEC using sensitive, molecular detection methods in children in Africa hospitalised with acute gastroenteritis, and is the only study to date conducted in Malawi. Additionally, it is amongst the few studies that have detected ETEC in Africa using molecular methods directly from stool samples, without first culturing *E.coli* from faecal samples.

Over the 10-year study period, ETEC was detected in 10.6% of children with diarrhoea. For the 2-year period during which faecal samples were collected from both children with and without diarrhoea, ETEC detection was significantly higher in children with diarrhoea, than in those without (12.7% vs. 7.3% $p=0.002$). There is a dearth of studies examining ETEC in children in Africa using PCR based methodologies. Many older studies have used various detections methods including DNA probes which have tended to have lower sensitivity and specificity than PCR (133), presenting difficulty when making direct comparisons. Globally, studies examining ETEC in infants (<1 year) diarrhoeal disease using various detection methods have tended to demonstrate detection rates ranging from 10-30% in one review (75). In the Middle East and Africa studies using molecular methods have demonstrated varied ETEC detection rates from 0% in Tripoli-Libya (134) to as high as 32.3% in Tunis, Tunisia (135) and 38.5% in South Africa (136). Additionally, in studies in Africa that included asymptomatic controls, ETEC has been detected significantly more in children with diarrhoea than in healthy controls

in Burkina Faso (137), but not in Ifakara, Tanzania (138) or Accra, Ghana (139). In Tunis, Tunisia (135) and South Western Nigeria (140) ST-EPEC, but not LT-EPEC strains were significantly associated with diarrhoea. The detection of enteric pathogens in both children with and without diarrhoea is not an isolated observation in less developed settings; this does however raise important questions about the role of these pathogens in childhood diarrhoea (141). Study design and methodology may go some way to explaining the observed differences, for example the definition of what constituted diarrhoea and the selection of asymptomatic or control patients not being homogenous between studies.

EPEC infection was observed in all age groups, with a peak in detection among 6-11 month old children. The median age of children in the diarrhoeic group with EPEC infection was 10 months (range, <1 – 40 months). This therefore suggests that the majority of disease in this population occurs in children less than 2-years of age, and is consistent with previous observations of EPEC as an important cause of weaning diarrhoea. Which affects children in poor sanitation settings from 6-24 months of age weaning from maternal breast milk (93). Additionally the GEMS reanalysis (7) identified EPEC as the most important bacterial cause of diarrhoea in children 0-11 months in Kenya and the second most important bacterial cause in the three other African study sites (Gambia, Mali and Mozambique). The observed decrease in EPEC infections beyond 2-years of age may be due to both a): environmental and/or b): host immunological factors.

Circulation of ETEC was noted throughout the year. No clear seasonality was demonstrated in this study; however detection of ETEC did peak in March, which is during the hot rainy season in Malawi, which stretches from November to April. Previously ETEC infections have been shown to cluster during warm, wet months that present favourable conditions for proliferation of the pathogen in the environment, as well as facilitating contamination of surface water with faecal material, including in the African countries of Egypt (142) and Guinea-Bissau (80). In other studies in Africa, ETEC diarrhoea was reported as statistically more likely to occur in the dry than rainy season in Tanzania (143) and some studies have noted no seasonality at all in Southern Mozambique (144) and Kenya (145). It is worth noting however that none of the quoted studies had the advantage of being conducted over a long period of time, as is the case with data presented here, an important factor when interpreting data presented on seasonality.

Analysis of circulating toxin types demonstrated a predominance of ST producing ETEC strains, which is consistent with the recent GEMS reanalysis (7), and has been described in several other studies in African children (137, 143, 146). The fact that ST producing ETEC strains appear more strongly associated with diarrhoea than LT strains may be related to an observed lower prevalence of CFs on LT producing ETEC strains (84). This is significant in terms of vaccine development as ST is only immunogenic when conjugated, unlike LT vaccines, which are comparatively easier to design. When ST toxin type distribution is examined, STh was more predominant than STp (6.6% vs 0.9%) in children with diarrhoea over the 10-year study period.

A similar pattern was observed for the 2-year period when faecal samples were collected from children with and without diarrhoea. Very few studies have looked to differentiate ST toxin types, including the recent GEMS reanalysis (7) (which only reported STh to be consistent with the original GEMS study, although the TaqMan™ card did include STp). In studies that have, STh or STp has been observed to predominate in different geographical settings. A study of children in urban Somalia (147) noted predominance of STh and a study in Libyan (146) children noted predominance of STp producing strains. Interestingly two cohort studies, one of children in Guinea-Bissau (80) and the other in Southern Israel (148) have noted that STh appeared more pathogenic than STp producing strains. Detection of specific ETEC toxin types may have important implications for vaccine development and further work is needed to fully explore this in different epidemiological settings, including the characterisation of colonisation factors in circulating strains.

ETEC was the fourth most common enteric pathogen detected to date in this paediatric Malawian population at 10.6%, after rotavirus (32.1%), *Campylobacter* (21%) and norovirus (11.3%). The high prevalence of mixed infections with ETEC among children with diarrhoea suggests that symptoms in a significant proportion of ETEC-positive cases may have been due to the presence of another pathogen, particularly rotavirus. In previous studies in Africa, detection of ETEC in mixed or co-infections ranged from 10.6% to 56.2% (143, 144). It is important to note that the detection methods used to

examine enteric pathogens in these Malawian studies varied, which makes comparisons difficult; in particular, ELISA was used to detect rotavirus. While it is likely that the use of more sensitive molecular detection methods would have led to increased detections, there is evidence that detection of rotavirus by ELISA correlates well with rotavirus diarrhoea, and avoids detection of low viral loads usually associated with prolonged shedding or asymptomatic carriage (150, 151).

In common with data presented here, interpretation of population-based case control studies examining prevalence of enteric pathogens in various settings is often complicated by detection of similar proportions of pathogen in both cases with and controls without diarrhoea. This raises important questions about the role of these pathogens in diarrhoeal disease. A number of gastrointestinal pathogens are known to be detectable after acute infection has resolved (141), and it is possible that a proportion of isolates detected in controls simply reflects chronic shedding following acute infection weeks or even months previously. However this alone is unlikely to explain these observations, especially in prospective studies where no diarrhoea is recorded in cases over a prolonged period of time. In case control studies, the calculation of an odds ratio or the attributable fraction have been used to quantify the degree of association of the pathogen being examined as a cause of diarrhoeal disease.

There are various host and pathogen specific factors that may influence the observation of excretion of certain enteric pathogens in stool in individuals

who do not experience diarrhoea. These factors are explored in detail in an excellent review by Levine and Robins-Browne (141), an overview of which is discussed here.

Pathogen factors

Certain enteric pathogens have been associated with prolonged or extended periods of excretion in stool following an episode of acute infection, including; *Salmonella*, *Campylobacter jejuni* and Norovirus G1 and GII. There is also an observed heterogeneity in the pathogenicity i.e ability to cause clinical diarrhoea of circulating enteric pathogens in human populations which has been observed in volunteer trials. This is notable for ETEC, where presence of various colonisation factors and differential expression of the LT and ST toxins impact upon the virulence of different strains. This observation has also been made for various strains of *Vibrio cholerae* O1, EPEC, *C jejuni* and *G lamblia*. A lack of complete understanding of the factors that determine clinically significant virulence in these pathogens has no doubt limited studies aimed at identifying their role in the cause of infectious diarrhoeal disease. It has been demonstrated in animal, but not human studies that the presence of two enteric pathogens may act synergistically to cause clinically apparent or more severe diarrhoea, e.g. with ETEC and rotavirus in pigs (152, 153).

Host factors

The presence or absence of various protein or carbohydrate receptors expressed on host mucosal surfaces can also play a role in the propensity to develop diarrhoeal disease following exposure to an enteric pathogen. This is most striking when we examine norovirus and cholera infection. Individuals with blood group O and hypochlorhydria are at increased risk of more severe disease following ingestion of *vibrio cholerae*. And in norovirus, expression of ABO blood group and secretor status influences host susceptibility to infection. Other non-specific host factors may influence the ability of an enteric pathogen to cause clinically apparent disease or a state of asymptomatic colonisation including; status of the host intestinal microbiome, mucus and epithelial layer, and immune response. It is also possible that a control case may simply be incubating the infection at the time of stool collection, and has not yet developed diarrhoea.

Environmental factors

Ingestion of an enteric pathogen at a dose insufficient to cause clinical infection may explain detection in an asymptomatic individual; this may be particularly relevant for pathogens like ETEC that rely on a food vehicle for transmission and where infection is dose dependent. Ingestion of animal enteric pathogens unable to cause human disease may lead to them being mistaken for human strains, such as porcine ETEC strain 263.

Characterisation of ETEC colonisation factors is needed to aid differentiation of human from animal ETEC strains.

Other factors

The sensitivity of the diagnostic test used may influence detection. For example, very sensitive molecular tests may detect the transient passage of tiny quantities of ingested pathogens insufficient to cause disease. Disruption of the host intestinal microbiome with oral antibiotics may alter host susceptibility to infection following exposure to an enteric pathogen. And finally, micronutrient deficiency, for example zinc and vitamin A can increase propensity to cause clinically overt diarrhoeal disease following enteric pathogen exposure, that otherwise may lead to a subclinical state in a non-deficient individual.

An advantage of this study is that PCR was performed directly on stool rather than culture, which may be less sensitive as it relies on a) growing the organism and b) selecting *E coli* colonies visually, which may lead to selection of non-diarrhoeic strains. However, it should be noted that the faecal samples used in this study had already undergone several freeze thaw cycles, which may have impacted on the ability to detect ETEC using even sensitive molecular methods. An additional advantage of this study is that it is one of few that have looked for both types of ST (STh and STp) toxin, which have been implicated in human disease. It is possible that occasional PCR contamination could have contributed to the data presented in this study,

especially as only a single negative PCR control was used, this however is likely to be minor. Both in light of the fact that there was a significant difference in the observed number of detections between children with and without diarrhoea, and also as compared to results presented in a recent case-control study in the same paediatric population in Blantyre, Malawi (149). That study demonstrated similar results in terms of a significantly higher proportion of ETEC detections in cases compared to controls; indeed, 21.2% of cases were PCR positive for ETEC, twice the proportion detected in data presented here, although that study was conducted after the introduction of rotavirus vaccine and a more sensitive real-time PCR was used.

Vaccine development in addition to improved levels of sanitation and hygiene is an important goal in efforts to reduce the global burden of ETEC infection. Data from this study suggests that any potential vaccine should include protection against ST or STh associated CFs and would best be administered to children less than 2-years where disease burden is highest. In addition, enhanced epidemiological surveillance of ETEC infection, similar to the Rotavirus Surveillance Networks is required from Sub-Saharan Africa and South East Asia where disease burden is greatest, to better inform future vaccine targets as well as well as monitor their effectiveness.

4 Molecular epidemiology of GII.4 norovirus strains in Malawi and the UK

4.1 Introduction

Global deaths from diarrhoeal diseases have declined considerably over the past several decades, from 2.6 million in 1990 to 1.3 million in 2013 (154).

This decline in mortality has been particularly noted in children < 5 years, with deaths falling from 1.6 million to 0.5 million per year over the same time period. These achievements have largely come about as a result of economic development in middle/low income settings where disease burden is highest; these achievements were highlighted in the progress that was made towards the UN Millennium Development Goals by 2015 (155). Improvements in sanitation, hygiene, access to clean drinking water and oral rehydration solution, as well as the global roll out of rotavirus vaccine into childhood immunization programmes have all had a positive impact (156). Despite these improvements, diarrhoeal disease remains the fourth most common cause of mortality in children < 5 years and the second most common cause of morbidity (157).

Norovirus is estimated to be associated with 18% (95% CI: 17-20%) of all diarrhoeal disease worldwide (38). Norovirus is pervasive and affects individuals across all age groups and income settings, however the highest rates of norovirus gastroenteritis are seen among young children in lower income settings (45). Global deaths from norovirus have been estimated at

approximately 200,000, with 70,000 or more of those deaths seen among children in developing countries (158).

Virology

Noroviruses comprise a diverse group of non-enveloped, positive sense, single stranded RNA viruses belonging to the *Caliciviridae* (16). The ~7.5kb genome contains three open reading frames (ORF 1-3) (159). ORF1 encodes six non-structural proteins including an RNA-dependent RNA polymerase (RdRp), while ORF2 and ORF3 encode the major VP1 and minor VP2 capsid proteins, respectively (15, 160, 161). The VP1 capsid protein consists of a conserved shell (S) domain located at the base of the capsid and a protruding (P) domain, further divided into two subdomains P1 and P2. The P2 subdomain is the most exposed portion of the viral capsid and is believed to contain antigenic and proposed receptor binding sites (162, 163). Noroviruses are classified into seven distinct genogroups (GI-VII) based upon sequence diversity within the VP1 protein, with genogroups GI, GII and GIV associated with human disease (19). Genogroups can further be divided into genotypes on the basis of diversity within the capsid and polymerase sequences (21).

Norovirus diversity and evolution

Following the introduction of molecular surveillance for infectious intestinal disease in the early 1990s, it quickly became apparent that, at least since 1995, norovirus variants of a single genetic lineage (GII.4) represented the

predominant-circulating virus causing human disease worldwide (164, 165). Since then GII.4 noroviruses have been confirmed as being responsible for the majority of norovirus outbreaks and sporadic disease observed globally. It is worth noting however, that disease surveillance is heavily biased towards outbreak-associated disease in high-income settings with established and well resourced public health systems. Notable established outbreak surveillance programmes include CaliciNet in the United States; ViroNet in Canada; the Foodborne Viruses in Europe network (FBVE) and the global norovirus network, NoroNet (166). Since the mid-1990s there have been six successive GII.4 variants that have demonstrated global outbreak or pandemic potential: 'US1995/96' in 1996 (164); 'Farmington Hills' in 2002 (167); 'Hunter' in 2004 (168); 'Den Haag' or '2006b' in 2007/08 (169); 'New Orleans' in 2009/12 (166) and most recently 'Sydney' in 2012 (170). Additionally, several other GII.4 variants have been identified which have been associated with regional epidemics, including 'Asia 2003' and 'Yerseke or 2006a' (165). It is interesting to note that we have not seen the emergence of a new globally important GII.4 strain since Sydney GII.4. However there has been recent emergence of a GII.17 strain in Asia and a recombinant strain GII.P16-GII.2 in other countries globally (171).

Novel epidemic GII.4 variants appear to emerge every two or three years. Although successive GII.4 variants appear to replace the previously dominant circulating strain, the old strain does appear to co-circulate, albeit at low levels within the population prior to going extinct. The reasons for this are not entirely clear, but may be due to a short-lived non-sterilising immunity to

norovirus (172) or perhaps these strains cause asymptomatic disease and do not come to the attention of current surveillance systems.

Norovirus evolution and diversity is influenced by a complex interaction of viral and host characteristics. Histo-blood group antigens (HBGAs), which are postulated to be involved in host susceptibility and possibly virus receptor binding are differentially expressed within the population (48). Host immunity is short lived following initial infection, and cross protection is limited at best (172). The huge diversity seen among noroviruses is driven by several virus and population factors; genetic drift occurs as a consequence of the intrinsic proof reading errors that occur during RNA replication, homologous recombination between ORF1 and ORF2 and within ORF2, and immune pressure acting on the hyper variable P2 sub-domain (124, 173, 174).

Our current understanding of the molecular mechanisms underpinning the persistence and emergence of successive GII.4 norovirus lineages in human populations has come about from a series of studies that have examined changes in the genetic diversity of capsid sequences and HBGA binding of successive GII.4 lineages. These studies will be discussed in more detail.

In a study by Siebenga *et al* in 2007 (175) the authors examined genetic and structural changes in the capsid proteins of successive GII.4 strains collected over a 13-year period in the Netherlands. The capsid protein was chosen, as being the most exposed portion of the norovirus virion, it was believed to represent likely antigenic or immune recognition sites. The study defined 'key

informative sites' as those where at least two strains had an identical amino acid or nucleotide mutation in the alignment.

The most exposed portion of the capsid protein, the P2 domain, was found to have the highest density of informative nucleotide (predominantly replacement mutations) and amino acid mutations, 24% of all amino acids in the domain. Changes in informative sites coincided with the emergence of new epidemic GII.4 variants, and *in silico* structural analysis demonstrated that the majority of these informative sites mapped to the surface of the P2 domain (the most exposed portion of the viral capsid). These findings suggested that antigenic drift in the capsid region occurs in response to host population immunity; and this results in the emergence of new variants capable of herd immune escape and the potential to cause global epidemics.

A subsequent and related study by Allen *et al in 2008* (124) described amino acid variation in the P2 domain of GII.4 variants collected from outbreak surveillance in the UK over a 10-year period. An advantage of this study was that, as oppose to *in silico* analysis using the crystal structure of the G.1 Norwalk virus as used in the Siebenga study; the authors modelled changes onto the crystal structure of a GII.4 norovirus strain belonging to the Grimsby cluster (VA 387) to describe potentially immunologically significant variant specific epitopes in the P2 domain of the capsid protein.

Again significant diversity was seen across the P2 region with significant changes coinciding with the emergence of new epidemic GII.4 variants. The

authors identified two sites, consisting of six amino acid positions where significant changes would impact upon the biochemical properties and structure of the P2 domain, noted as site A (amino acid positions 296-298) and site B (amino acid positions 393-395). Significantly these regions are found in exposed loops in the 3D capsid structure, and site A is close to a HBGA binding site identified by Cao *et al* (176). Additionally, an amino acid substitution at position 340 was noted to coincide with emergence of the GII.4 2002 epidemic strain. Homology modelling was used to examine variation in biochemical and structural properties of the capsid protein. Significant structural and electrostatic changes related to changes at sites A and B were observed. These changes may allow masking from potentially neutralising host antibody responses and for antibody escape mutants to emerge at the population level.

In a subsequent study the diversity seen at sites A and B has been further characterised (177). From analysis of amino acid changes over time at these sites the authors characterised three site A motif clusters and noted greater diversity at site B. They predicted that site A is a major antigenic epitope responsible for defining different epochs of norovirus activity, while site B may have a more minor role in maintaining the diversity needed to preserve the observed year on year epidemics within pre-exposed populations, between the major epidemic waves or “epochs”.

Another study by Lindesmith *et al* in 2008 (174) examined GII.4 virus like particle-carbohydrate ligand binding patterns in epidemic strains from the

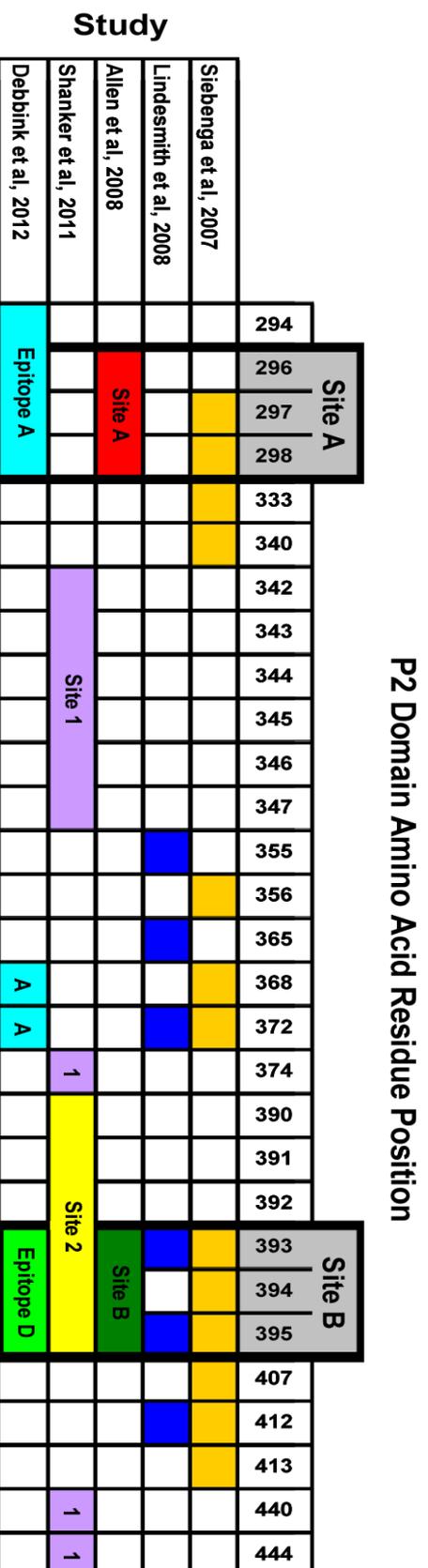
previous two decades. The authors demonstrated that these binding patterns changed over time and included carbohydrates regulated by human FUT2 and FUT3 pathways, suggesting that strain sensitivity to human susceptibility alleles will vary. Findings from this study suggest that noroviruses are under heavy immune selective pressure and that the P2 domain is evolving at a faster rate than other regions in the viral capsid.

Finally, in two similar studies by Shanker *et al* in 2011 (178) and Debbink *et al* in 2012 (179) the authors examined HBGA binding patterns in contemporary and historic GII.4 norovirus outbreak strains in more detail to understand how changes in binding patterns may contribute to a proposed model of epochal norovirus evolution as suggested by Lindesmith *et al* (174). Shanker *et al* identified two sites where HBGA binding occurred, site 1 and 2 (Figure 6); and found that the strength of binding at site 2 could change over time, suggesting a role in epochal evolution by enhanced targeting of Lewis-positive, secretor-positive individuals. As well as examining HBGA binding, Debbink *et al* also examined antigenic binding profiles using mouse and human monoclonal antibodies and sera, they identified two epitopes A and B (Figure 6).

Taken together these studies provide evidence for a model of epochal evolution driven by genetic drift in GII.4 noroviruses, as also seen in Influenza viruses (180). In this model, periods of phenotypic stasis characterised by selectively neutral mutations occur in a population of antigenically equivalent viruses; until sufficient diversity is generated to allow for antigenic escape and generation of a new epidemic strain. There is likely a complex and as yet

incompletely understood interaction between changes in residues coding for antigenicity and HBGA binding that influence the model of GII.4 evolution described above. A disadvantage of these studies however is that there are heavily biased towards outbreak strains and may not provide a complete picture of the molecular evolutionary processes that culminate in emergence of new epidemic GII.4 strains.

Figure 6. Key areas of the viral capsid (amino acid positions) that have been predicted to be involved in virus-host interactions. Taken from Zakikhany *et al* (177) (Open access).



Norovirus in Africa

The prevalence and diversity of norovirus in Africa is poorly understood, especially in adults where there is a paucity of data. However, early seroprevalence studies have suggested that norovirus infection is widespread among the population and that initial infection occurs early in life (181, 182). There is wide variation in studies in terms of their duration, population groups and size, inclusion of asymptomatic controls and availability of genotyping data.

A recent review by Mans *et al* (183) has examined studies in Africa (the majority of which were in children), that took place over a period greater than six months, had a study size greater than fifty, and used molecular detection (RT-PCR) methods. The authors estimated a mean overall prevalence of 13.5% (range 0.8-25.5%) in individuals with gastroenteritis and 9.7% (range 7-31%) in asymptomatic controls. Although lacking in Africa, available norovirus genotyping data suggested the most prevalent genotype was GII.4, and other important genotypes identified were GII.3 and GII.6.

In children ≤ 2 years, norovirus prevalence was significantly higher than those ≤ 5 years (18% vs. 11.8%). The prevalence was also significantly higher in symptomatic cases compared to asymptomatic controls, 13.5% vs. 9.7%. This difference is however significantly smaller when compared to a global estimate by Ahmed *et al* (38) of 20% vs. 7%. The reasons for this observed discrepancy are unclear, but may be due to how asymptomatic controls were

defined in the African studies. It is known that norovirus can be shed in stool for up to three weeks after acute infection (184, 185) (longer in immunosuppressed populations e.g. with HIV), and interestingly only two (121, 186) out of nineteen studies specified a diarrhoea free period prior to specimen collection. For example, the GEMS study (187) used a seven day diarrhoea free period before enrolment of controls. In addition, children in developing countries may be significantly more exposed to norovirus from their environment and have multiple exposures, some of which may be clinically asymptomatic. There was no general trend in terms of co-infections reported by the Mans study, and reporting of mixed infections varied greatly, with norovirus/rotavirus co-infections most common.

Genotype data was available from eighteen African countries (183). The prevalence of GII norovirus from all studies was 84.1% (range 71-100) and GI 13.9% (range 0-29); three studies reported mixed GI/GII infections.

Genotyping was undertaken using either analysis of capsid sequences alone or in combination with RNA dependent RNA-polymerase (RdRp). GII.4 was the most commonly detected and widely distributed genotype across Africa; the second most commonly detected was GII.3 and the second most widely distributed was GI.3. Other capsid genotypes with wide distribution, but relatively low prevalence, included GII.2, GII.6, GII.14 and GII.17.

Between 1998 and 2013 twelve GII.4 variants were reported in norovirus studies from Africa. Several of these studies in paediatric African populations have suggested circulation of pandemic GII.4 strains years before their

subsequent global recognition, including Farmington Hills 2002 and Den Haag 2006b in Malawi (121) in 1999 and 2001 respectively; Osaka 2007 in Ghana (188) in 2000; New Orleans 2009 in Botswana (189) in 2006 and more recently Sydney 2012 in Cameroon (190) in 2009.

Few studies in Africa have reported combined RdRp/capsid typing data, therefore information on the prevalence of norovirus recombinants is lacking. There is a need to standardise the approach to norovirus surveillance in Africa and genotype the strains fully.

Great Norovirus genotype diversity is observed in studies of paediatric populations in Africa, even in those with small numbers (183). There is very little data from adults or elderly individuals. The high HIV prevalence observed in Sub-Saharan Africa raises the question, does this large number of immunosuppressed individuals affect norovirus evolution and diversity, and clearly this needs further investigation.

4.2 Aims

Long-term surveillance of rotavirus in children in Africa has provided an opportunity to explore the impact and epidemiological features of norovirus in this population (111). The aims of the present study were to examine the hypervariable P2 sub domain of these and other sporadic GII.4 norovirus strains collected between 1993-2013, in order to further explore their evolutionary relationship with GII.4 global outbreak strains.

4.3 Methods

4.3.1 Specimens

This retrospective study included 294 anonymised GII.4 norovirus PCR positive samples (from archived faecal material or cDNA extracted from them). Sample collection and definitions are described in detail in Chapter 2. Briefly, samples used in this study (Table 8) were collected from three sources (i): children <5 years old in Blantyre, Malawi, hospitalised with diarrhoea, (ii): children presenting to general practice in England and Wales with gastroenteritis (Nappy study), (iii): children and adults with gastroenteritis in the community or attending general practice in the UK (Infectious Intestinal Disease (IID) studies I and II). Samples from IID I also included healthy community controls.

Table 8. Origin of norovirus positive samples analysed in this study.

	Infectious Intestinal disease study in England	Malawi study	Structured surveillance of Infectious Intestinal Disease in Pre-School Children in the Community	Second study of Infectious Intestinal Disease in the Community
Study	IID1	Malawi study	Nappy study	IID2
Sampling years	1993-1996	1998-2002	2007-2008	2008-2009
Location	England	Malawi	England	UK
Number of samples	87	3	19	185

Ethics

The IID Study Scientific Steering Committee granted approval for use of archive material from the IID studies. Samples were anonymised and no patient identifiable information was available. Samples from the Malawi study were stored at the Institute of Infection and Global Health, University of Liverpool, UK and IID and Nappy study samples were stored at the Enteric Virus Unit, Public Health England (PHE), Colindale, London, UK. Data and sample archives held at the Institute of Infection and Global Health, University of Liverpool and The Enteric Virus Unit, PHE are compliant with the standards set out by the UK human health tissue act and Caldecott principles. At the time of sample collection, informed consent was gained from patients (or their guardians if age less than 18 years) to store samples for the purposes of future research.

4.3.2 Nucleic acid extraction, reverse transcription and norovirus testing

For specimens collected in all studies, nucleic acid was obtained directly from faecal material, converted to randomly-primed cDNA and established as norovirus-positive using previously described methods: IID1 study (113), IID2 study (46, 120), The Nappy study (117).

In more detail, my own previous work at the Institute of Infection and Global Health, University of Liverpool, examined the Malawi samples for norovirus. As previously described (121), nucleic acid was extracted from 10 % faecal

suspensions in PBS using a high throughput automated extractor (QIAAsymphony ®, Qiagen UK). Reverse transcription of RNA was performed using random hexamers. Norovirus genogroups GI/GII were detected using a real-time PCR technique described by Kageyama et al (122) as modified by Amar et al (123).

4.3.3 Amplification of the GII.4 norovirus P2 domain for strain characterisation

In order to determine GII.4 strain types, the region of the ORF2 gene encoding the VP1 hypervariable P2 domain was amplified by PCR as previously described for samples collected from the UK (24, 124). Full-length P2 domain amplicons were sequenced using a dideoxy chain-terminator method. For nucleic acid extracts from the Malawi studies the reader is referred to Chapter 2 for further details.

4.4 Results

For the period 1993-2009 a total of 294 GII.4 norovirus strains were characterised from four studies conducted in the UK and Malawi (Table 9).

Table 9. Numbers of GII.4 strains analysed and variant assignment according to phylogenetic analysis.

Strains that do not align chronologically with the GII.4 variant circulating during the study period are highlighted (bold & underlined). Adapted from Allen DJ and Trainor E *et al* 2016 (191) (Open access).

Study/ GII.4 strain variant	IID1 1993-1996	Malawi 1998-2002	Nappy 2007-2008	IID 2 2008-2009	Total
Camberwell/1987	4				4
Lordsdale/1993	6				6
Grimsby/1995	43				43
Dresden/1997	22				22
Bochum/1997	6				6
Farmington Hills/2002	<u>1</u>				1
Hunter/2004	<u>3</u>	<u>3</u>	2		5
2002-2004 Cluster					3
Yerseke/2006			9	34	43
Den Haag/2006			8	99	107
Appledorn/2007				51	51
New Orleans/2009	<u>1</u>			<u>1</u>	2
Sydney/2012	<u>1</u>				1
Total	87	3	19	185	294

4.4.1 Molecular epidemiology of GII.4 strains detected in UK and Malawi

Analysis of IID1 study specimens demonstrated that almost half the strains $n = 43/87$ (49%) could be assigned to the Grimsby/1995 genetic cluster, and a quarter $n = 22/87$ (25%) to the Dresden/1997 cluster (Table 9). The Grimsby/1995 and Dresden/1997 strains were commonly detected globally circulating GII.4 norovirus strains during the study period.

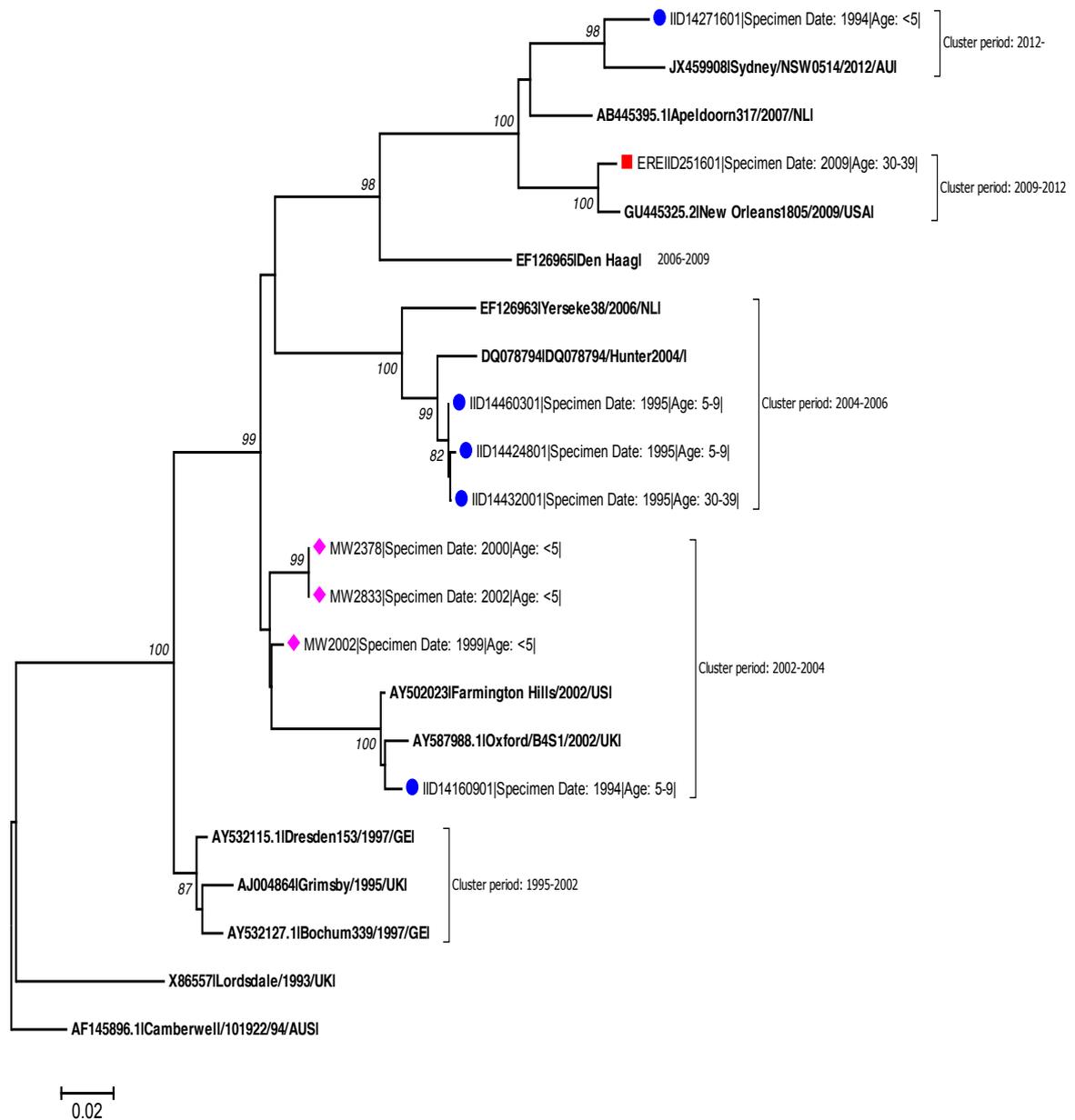
Analysis of Nappy Study specimens demonstrated that n=9/19 (47%) of strains could be assigned to the Yerseke 2006a genetic cluster, and n=8/19 (42%) to the Den Haag 2006a cluster (Table 9); again both these strains were commonly circulating globally during the study period.

Finally, analysis of the IID2 study specimens demonstrated that n= 99/185 (53%) could be assigned to the Den Haag 2006a genetic cluster, and n=51/185 (28%) to the Apeldoorn/2007 cluster (Table 9), and as previously, both these strains were commonly detected globally during the study period.

GII.4 norovirus strains from three of the four studies (IID1, Malawi and IID2) were detected in a time period that predated their subsequent identification as strains with global epidemic importance (Table 9 and Figure 7).

Figure 7. Clustering of norovirus strains identified in sporadic cases of gastroenteritis from Malawi (labelled with pink diamond) or the UK (labelled with a blue circle or red square) against global GII.4 strains representative of the epidemic strain variants that emerged between 1995 and 2012.

Nucleotide sequences analysed correspond to 470 nt of the P domain of the VP1 encoding gene and encompass the entire length of P2 domain. Taken from Allen DJ and Trainor E *et al* 2016 (191). (Open access).



In further detail, analysis of GII.4 strains from the IID1 study (1993-1996) identified four strains that clustered with strains not identified as having global epidemic importance until after 2002. Taking into account the date of initial sample collection and when these strains were subsequently recognised as global epidemic strains, retrospective detection of Farmington Hills/2002, Hunter/2004, New Orleans/2009 and Sydney/2012 by eight, nine, fifteen and eighteen years respectively is demonstrated by these data (Table 9 and Figure 7).

Analysis of the three GII.4 strains detected during the Malawi study demonstrated that they clustered together and were related to strains later detected after 2002 (Figure 7). Phylogenetic analysis demonstrated that all three strains were closely related to GII.4 strains identified in 2002 and 2004. Two of these strains (MW2002 and MW2378), were identified from faecal species collected in 1999 and 2000, respectively. These data therefore suggest that GII.4 norovirus strains similar to Farmington Hills/2002 were circulating in a paediatric population in Malawi at least 2 to 3 years prior to their subsequent identification as strains with global epidemic importance (Table 9 and Figure 7).

Analysis of GII.4 strains from the IID2 study (2008-2009), identified one strain identical to New Orleans/2009; this was detected in a faecal sample collected in Summer 2009, immediately prior to the recognition of New Orleans/2009 as a new global epidemic strain (Table 9 and Figure 7).

4.4.2 Analysis of amino acid variation in the hyper variable domain of the major capsid protein

Examination of variable amino acid positions along the hypervariable P2 domain in those GII.4 norovirus isolates detected in a time period that predated their identification as strains with global epidemic importance has suggested that they may represent genetically transitional GII.4 strains, that later emerged globally (Figure 8).

The Farmington Hills/2002 prototype strain (GenBank accession number AY5022023) is indistinguishable from that detected from the IID1 study, indicating that this strain was circulating in the UK at least eight years earlier than its eventual association with global emergence.

Figure 8. Alignment of the deduced amino acid sequencing obtained from sporadic strains detected in Malawi and the UK with prototype strains representative of the epidemic strain variants that emerged between 1987 and 2012.

Successive variants are colour coded and prototype strains are listed chronologically with earliest variants at the top. Study sequences are identified according to origin as in Table 9. Single amino acid identities of the study strains against prototype strains are colour coded using the colour assigned to each epidemic strain variant. Taken from Allen DJ and Trainor E *et al*/2016 (191). (Open access).

Strain Name	Year Collected	284	Envelope A										Envelope A					Envelope D						
			286	287	288	333	335	340	346	355	356	365	368	372	376	377	378	393	394	395	407	412	413	
Camberwell/1987	1987	V	S	H	D	L	T	A	A	S	V	V	T	N	Q	T	G	D	-	H	N	T	G	
Lordsdale/1993	1993	A	S	H	D	L	T	A	A	S	V	V	T	N	Q	A	G	D	-	H	N	T	G	
Grimsby/1995	1995	A	S	H	D	M	T	E	A	S	V	V	T	N	Q	T	G	N	-	N	N	T	G	
Dresden/1997	1997	A	S	H	D	M	T	E	A	S	V	V	T	N	Q	T	G	N	-	N	N	T	G	
Bochum/1997	1997	A	S	H	D	T	T	E	A	S	V	V	A	T	N	Q	T	G	N	-	N	N	T	G
Farmington Hills/2002	2002	A	T	H	N	M	T	G	G	D	V	I	N	N	E	T	G	N	G	T	S	T	G	
Oxford/2002	2002	A	T	H	N	M	T	G	G	D	V	I	N	N	E	T	G	N	G	T	S	T	G	
KU312299 [Farmington Hills/2002]	1993-1996	A	T	H	N	M	T	G	G	D	V	I	N	N	E	T	G	N	G	T	S	T	G	
KU312304 [Cluster 2002-2004]	1999	A	T	H	D	V	S	G	G	S	V	V	T	N	E	T	G	N	S	T	N	T	G	
KU312305 [Cluster 2002-2004]	2000	A	T	H	D	V	S	G	G	S	V	V	T	N	E	T	G	N	S	T	N	T	G	
KU312306 [Cluster 2002-2004]	2002	A	T	H	D	V	S	G	G	S	V	V	T	N	E	T	G	N	S	T	N	T	G	
Hunter/2004	2004	A	A	Q	N	V	T	R	G	S	V	V	S	S	E	T	G	S	T	T	D	D	S	
KU312302 [Hunter/2004]	1993-1996	A	T	Q	N	V	T	R	G	S	V	V	S	S	E	T	G	S	T	T	D	D	S	
KU312301 [Hunter/2004]	1993-1996	A	T	Q	N	V	T	R	G	S	V	V	S	S	E	T	G	S	T	T	D	D	S	
KU312303 [Hunter/2004]	1993-1996	A	T	Q	N	V	T	R	G	S	V	V	S	S	E	T	G	S	T	T	D	D	S	
Den Haag/2006b	2006	A	S	R	N	V	T	G	G	S	V	V	S	S	E	T	H	S	T	T	D	D	S	
Yerske/2006a	2006	A	T	Q	E	V	T	R	G	S	V	V	S	S	E	T	G	S	T	T	D	D	S	
Agadjoon/2007	2007	A	S	R	N	V	T	A	G	S	V	V	S	S	E	T	A	N	D	T	A	S	N	
New Orleans/2009	2009	P	S	R	N	V	T	T	G	S	A	V	A	D	E	T	N	D	T	A	S	N	I	
KU351252 [New Orleans/2009]	1994	P	S	R	N	V	T	T	G	S	A	V	A	D	E	T	N	D	T	A	S	N	I	
KU312298 [New Orleans/2009]	2006-2007	P	S	R	N	V	T	T	G	S	A	V	A	D	E	T	N	D	T	A	S	N	I	
Sydney/2012	2012	T	S	R	N	V	T	T	G	S	A	V	A	D	E	T	N	D	T	A	S	N	I	
KU312300 [Sydney/2012]	1993-1996	T	S	R	N	M	T	T	G	S	A	V	A	D	E	T	N	D	T	A	S	N	I	

Analysis of GII.4 strains collected in Malawi demonstrates that although they are extremely similar to the Farmington Hills/2002 prototype, they show some variances that may suggest they are intermediary between this and the Grimsby/1995 prototype (GenBank accession number AJ004864). Key differences are noted at amino acid positions within the hypervariable P2 domain of the norovirus major capsid protein, specifically position 298, part of epitope A, which is believed to be an important antigenic site in emergent GII.4 norovirus strains, and amino acid position 368, an important site in the emergence of the Sydney/2012 pandemic GII.4 strain. Other amino acid positions show greater similarity to Grimsby/1995 than to Farmington Hills/2002, specifically positions 355, 365 and 407. Finally, of note are two further amino acid positions, 335 (a serine residue) that is unique to the Malawi strains and position 333 which is more similar to the later Hunter/2004 prototype (GenBank accession number DQ078794) than either Grimsby/1995 or Farmington Hills/2002.

Hunter/2004 strains identified during the IID1 study appear very similar to the prototype strain, with the notable exception of a difference at amino acid position 296, part of epitope A, which is more similar to the preceding global epidemic strain Farmington Hills/2002.

Analysis of the New Orleans 2009 strain detected in the IID2 study demonstrated it is indistinguishable from the prototype strain (GenBank accession number GU445325), indicating that the New Orleans/2009 strain was circulating in the UK at least one year prior to its eventual recognition as a strain with global epidemic importance.

Finally, analysis of the Sydney/2012 strain detected during the IID1 study demonstrates it is very similar to the prototype strain (GenBank accession number JX459908), with the exception of two significant amino acid positions. Firstly, amino acid position 333, which is more related to the Grimsby/1995 prototype (the principal circulating GII.4 strain during the IID1 study period). And secondly, amino acid position 393, this is part of Epitope D, another antigenic site believed to be important in emergent GII.4 norovirus strains, which appears more related to the Hunter/2004 prototype.

4.5 Discussion

This large retrospective study examined GII.4 norovirus strains collected over a 25-year period, including from previous studies of sporadic community acquired gastroenteritis in the UK and Malawi (1993-2009), and global epidemic strains (1987-2012). Two of the most recently described GII.4 norovirus strains of global epidemic importance, GII.4 Sydney/2012 and GII.4 New Orleans/2009 were detected among sporadic community cases of gastroenteritis up to 18 and 15 years respectively, prior to their initial recognition as emergent global epidemic strains. These data confirm the previous observation that GII.4 strains of global epidemic importance were circulating among sporadic cases of paediatric diarrhoea in Malawi prior to their subsequent recognition as emergent epidemic strains (121). The three strains detected during the Malawi study are similar to Farmington Hills/2002, suggesting that this or closely related, perhaps intermediary strains were circulating amongst paediatric populations in Malawi several years prior to their subsequent recognition or emergence.

Phylogenetic analysis of these strains therefore suggests that there may be a greater diversity of GII.4 norovirus strains circulating among sporadic community gastroenteritis cases compared to those causing large outbreaks and subsequently identified through current national and international outbreak surveillance systems. These data suggest that there is low level circulation of a diverse population of GII.4 strains in sporadic community cases years before some of these emerge as strains with global epidemic significance. These observations lend further support to a proposed model (175) of epochal evolution driven by antigenic drift in the viral capsid sequence that occurs in response to host immune pressure, which selects for a diverse population of GII.4 variants, some of which may subsequently escape herd immunity and emerge as strains with global epidemic potential.

Analysis of GII.4 norovirus strains characterised after the year 2002 demonstrates an amino acid insertion at position 394. This insertion is also seen in a minority of UK and Malawi GII.4 strains that were circulating in the community pre-2002. It is not clear why there was potentially a delay in the spread of this amino acid motif. One possible reason is that other single amino acid changes around position 394, and other proposed epitopes (see Figure 6), may have a role in defining viral fitness and host immune escape. Data presented here may therefore be of utility in trying to further delineate the amino acid positions alongside the VP1 protein that are involved in strain specific immune responses. Further analysis of the entire genome may also identify regions e.g. non-structural coding genes, which act as important determinants in the emergence of strains with global epidemic potential. The GII.4 norovirus strains analysed in this study were collected from both adults and children;

however no age specific associations were identified in those strains that clustered with global epidemic strains identified years later.

Current norovirus surveillance programmes are heavily biased towards analysing strains from outbreak cases. It is possible these strains represent the end of an evolutionary process where founder variants are selected out following one or more of a series of bottleneck events in the population. This work has demonstrated the importance of characterising norovirus strains from sporadic community cases, so that a more complete picture of strains circulating in the population can be captured. This will allow better understanding of the evolutionary processes that contribute to the emergence of strains with global epidemic importance. By examining strains from sporadic community cases of gastroenteritis it has been possible to characterise strains that are intermediate between known epidemic variants.

These data suggest that the general population may be the reservoir in which the intermediate strains are generated and circulate, albeit at low levels, due to reduced viral fitness, until sufficient antigenic diversity occurs that allows for host immune escape. However, this may not be the complete picture, as alternative sources have also been suggested including animal reservoirs and chronically infected immunosuppressed patients (192, 193). Therefore surveillance programmes that include strains from sporadic cases of norovirus gastroenteritis may allow earlier detection of novel strains or emergent motifs that confer selective advantages; as well as information on how these strains evolve and establish at the population level. Along with a more detailed understanding of those epitopes involved in strain

specific immune responses, the ability to predict future global epidemics may emerge.

This work has potential to inform on-going efforts to develop a norovirus vaccine, as a greater understanding of the degree of cross protection and the role of strain diversity in evading pre-existing host immune responses will be required. It is interesting to note that in a recent vaccine trial in healthy adult volunteers using a consensus multivalent norovirus virus like particle derived vaccine, antibodies were induced in vaccine candidates that could block carbohydrate ligand binding to GII.4 variants that were not yet circulating at the time of vaccination (194). The work presented here offers a possible explanation to this observation, which is that global epidemic GII.4 strains circulate, albeit at low levels, in the population before a subsequent explosive emergence. This could potentially contribute to the observation of broad antibody responses in those vaccine candidates previously exposed to these strains.

Further work is needed to better characterise circulating GII.4 norovirus strains. This should include more comprehensive sequence analysis of the whole genome or sequences spanning the entire P2 domain. There is also a need to design more robust and inclusive norovirus surveillance networks that can capture strains from sporadic as well as outbreak disease and in more of the global population, including from individuals in low-income settings.

5 General discussion and recommendations

5.1 Summary of main findings

Using sensitive molecular diagnostic methods, I have investigated two well-recognised bacterial and viral causes of acute diarrhoeal disease in a low income paediatric population; specifically ETEC and norovirus in Blantyre, Malawi. The epidemiology of ETEC was investigated among hospitalised children, and molecular epidemiological processes involved in the evolution of GII.4 norovirus strains were explored in this and other settings including both adults and children in the UK.

Enterotoxigenic E coli

Key findings from the ETEC study that examined archived faecal specimens collected from children \leq 5 years with diarrhoea over a 10-year period (1997-2007) was a prevalence of ETEC of 10.6%; ETEC was detected significantly more often in children with (12.7%) than in those without diarrhoea (7.3%) for the 2-years (1997-1999) when both specimen types were available for examination. Additionally, analysis of the toxigenic distribution of ETEC strains in this paediatric population identified STh as most prevalent, followed by LT then STp.

In addition to the work presented here examining ETEC and norovirus, this 10- year collection of archived faecal samples from Malawi have also previously been examined for rotavirus (111); norovirus (121) and *Campylobacter* (112). It is worth noting that not all samples originally examined for rotavirus were subsequently

available for examination of the other enteric pathogens; only 79% of these samples were subsequently examined for norovirus, *Campylobacter* and ETEC. This was due to technical reasons related to the long term storage of the specimens, specifically the quantity of faecal material available for nucleic extraction.

When combining work from this thesis with that published earlier and comparing the relative prevalence of the different enteric pathogens, rotavirus was most prevalent at 32.1%; followed by *Campylobacter* (21%); norovirus (11.3%), then ETEC (10.6%). The median age of detections also varied; viral pathogens were more commonly seen in younger age groups (rotavirus age 7 and norovirus age 6 months) and bacterial pathogens tended to be identified in older age groups (*Campylobacter* age 11 and ETEC age 9.2 months). Malawi experiences a tropical climate; with a dry season May through October and a rainy season from November through April. The viral pathogens demonstrated a more discernable seasonality than the bacterial pathogens; rotavirus detections peaked during the dry season and norovirus peaked towards the end of the dry and wet seasons. *Campylobacter* demonstrated no seasonality and for ETEC infections, although no clear seasonality was observed, detections peaked in January, March and September through December.

Norovirus

I examined GII.4 norovirus strains over a 25-year period (1987-2012) from studies of sporadic acute diarrhoea in children in Malawi, adults and children in the UK, as well as contemporarily and historically recognised global epidemic strains. Key findings included the demonstration that GII.4 strains of global epidemic importance,

including the two most recently described strains GII.4 New Orleans 2009 and Sydney 2012, were circulating in sporadic cases of paediatric community diarrhoea 15 and 18-years respectively prior to their subsequent and first recognition as strains with global epidemic or pandemic significance.

5.2 Strengths and limitations

These studies have examined two well-recognised enteric pathogens in a low income, high disease burden African paediatric population. Retrospective analysis has allowed for examination of ETEC over a 10-year period and GII.4 norovirus strains from sporadic and outbreak surveillance over a 25 year period. Although there was some variation in the research effort, as a small research team undertook collection of faecal samples in Malawi, the surveillance approach used e.g. inclusion/exclusion criteria and laboratory methods utilised was generally consistent throughout.

Sensitive molecular diagnostic techniques were used for the detection of both norovirus and ETEC in these studies. The molecular examination for ETEC was undertaken directly from faecal samples, without first culturing for *E.coli*; this is one of the largest studies to date in Africa which has examined for ETEC in this way, and is the only study which has examined for ETEC in children in Malawi. In addition to the ETEC LT toxin this is also one of few studies that has examined for both types of ST toxin (STp and STh); both of which have been implicated in human disease. However, the study did not examine for colonisation factors in these ETEC strains, and this information would be useful to inform future vaccine development.

A sensitive conventional PCR method was used for the detection of ETEC; therefore it has not been possible to provide quantitative data to examine for PCR cut off/Ct values. Regardless, cut off values have not been established for ETEC to differentiate between infective and carriage states, and the use of qPCR would not have provided additional data to answer this question. Further longitudinal studies are needed to establish cut off values for ETEC and other enteric pathogens.

It is worth noting that these studies in Malawi have not looked at a fully comprehensive range of gastrointestinal pathogens (including viral, bacterial and parasitic) and that varied diagnostic modalities were used including ELISA, convention and real time PCR. In addition, it has not been possible to calculate population attributable fractions, so it is not possible to fully determine whether all pathogen detections were the cause of the diarrhoeal episode under examination. Further, the relative importance and epidemiology of the various enteric pathogens under examination may have changed since the introduction of rotavirus vaccine in this and other settings; highlighting the need for robust surveillance systems in this and other low-income settings with the highest disease burdens, including post vaccine surveillance.

Archived stool samples examined in these studies had already undergone frequent freeze-thaw cycles; which may have resulted in degradation of the pathogen genomes (especially RNA, which is less stable) and reduced the sensitivity of the methods of detection utilised.

5.3 Further work and recommendations

There are no licensed vaccines against either norovirus or ETEC, and the only licensed vaccines currently available to protect against childhood diarrhoeal disease are for rotavirus and cholera. However, there are currently norovirus and ETEC vaccines at clinical trial stage. My data suggest that in Malawi, in order of relative importance vaccines against campylobacter, norovirus and ETEC would appear attractive candidates to further reduce the burden of childhood diarrhoeal disease following on from the successful introduction of rotavirus vaccine.

The results from reanalyses of two large global studies using sensitive molecular diagnostic panels to examine for gastrointestinal pathogens in childhood diarrhoeal disease in high burden settings have noted the following. In the GEMS re-analysis (7), six enteric pathogens together accounted for 77.8% of all attributable diarrhoea, in order of importance *Shigella* spp./EIEC; rotavirus; adenovirus 40/41; ST-ETEC; cryptosporidium spp. and *Campylobacter* spp. And in the MAL-ED cohort study re-analysis (9) ten enteric pathogens together accounted for 95.7% of all attributable diarrhoea, in order of importance *Shigella*, sapovirus, rotavirus, adenovirus 40/41, ETEC, norovirus, astrovirus, *Campylobacter*, cryptosporidium and typical enteropathogenic *E coli*. Taken together in terms of disease burden, three pathogens from these studies appear to be attractive vaccine candidates (*Shigella*, adenovirus 40/41 and ETEC), with *Shigella* and ETEC the most important causes of bacterial childhood diarrhoeal disease. Indeed the global health community have recently prioritised the development of vaccines against *Shigella* and ETEC through PATH (Global Health Organisation) (110).

Norovirus as an important cause of childhood diarrhoea in low income settings does not feature strongly in these large study reanalyses, most prominently as only the sixth most important pathogen in the MAL-ED study. This raises the question of where a norovirus vaccine would be most useful, likely in more developed settings where disease burden is higher and/or disease causes significant economic impact e.g. elderly and immunosuppressed in high-income settings, travellers, and the military.

Any successful norovirus vaccine is likely to need to be replaced or updated every few years, as is the case for the influenza virus (195), due to the rapid strain replacements seen in both these RNA viruses. The work presented here makes a case for more robust and intelligent surveillance systems, that not only capture strains from outbreaks of norovirus gastroenteritis, but also from sporadic disease. Indeed this type of surveillance may allow for the detection of novel strains prior to their epidemic or pandemic emergence; this will clearly be of importance in informing any future vaccine updates.

In addition to previous work on pathogen-specific burden of diarrhoeal disease as presented in the recent GEMS studies, these studies presented here provide country specific epidemiology for Malawi. This information will be important to guide vaccine development to settings with the highest disease burden.

References

1. World Health Organization. Global Health Observatory (GHO) data. Causes of child mortality 2017 [cited 2018 10 December]. Available from: https://http://www.who.int/gho/child_health/mortality/causes/en/.
2. Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2017 (GBD 2017) Results. 2017 [cited 2018 10 Dec]. Available from: <http://ghdx.healthdata.org/gbd-results-tool>.
3. Mwenda JM, Mandomando I, Jere KC, Cunliffe NA, Duncan Steele A. Evidence of reduction of rotavirus diarrheal disease after rotavirus vaccine introduction in national immunization programs in the African countries: Report of the 11(th) African rotavirus symposium held in Lilongwe, Malawi. Vaccine. 2019 Apr 24. PubMed PMID: 31029514.
4. Levine MM, Kotloff KL, Nataro JP, Muhsen K. The Global Enteric Multicenter Study (GEMS): impetus, rationale, and genesis. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2012 Dec;55 Suppl 4:S215-24. PubMed PMID: 23169934.
5. Rogawski ET, Liu J, Platts-Mills JA, Kabir F, Lertsethtakarn P, Siguas M, et al. Use of quantitative molecular diagnostic methods to investigate the effect of enteropathogen infections on linear growth in children in low-resource settings: longitudinal analysis of results from the MAL-ED cohort study. The Lancet Global health. 2018 Dec;6(12):e1319-e28. PubMed PMID: 30287125.
6. Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, Panchalingam S, 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 Jul 20;382(9888):209-22. PubMed PMID: 23680352.
7. Liu J, Platts-Mills JA, Juma J, Kabir F, Nkeze J, Okoi C, 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 Sep 24;388(10051):1291-301. PubMed PMID: 27673470.
8. Investigators M-EN. The MAL-ED study: a multinational and multidisciplinary approach to understand the relationship between enteric pathogens, malnutrition, gut physiology, physical growth, cognitive development, and immune responses in infants and children up to 2 years of age in resource-poor environments. Clinical infectious diseases : an official

publication of the Infectious Diseases Society of America. 2014 Nov 1;59 Suppl 4:S193-206. PubMed PMID: 25305287.

9. Platts-Mills JA, Liu J, Rogawski ET, Kabir F, Lertsethtakarn P, Siguas M, et al. Use of quantitative molecular diagnostic methods to assess the aetiology, burden, and clinical characteristics of diarrhoea in children in low-resource settings: a reanalysis of the MAL-ED cohort study. *The Lancet Global health*. 2018 Dec;6(12):e1309-e18. PubMed PMID: 30287127.

10. Kapikian AZ, Wyatt RG, Dolin R, Thornhill TS, Kalica AR, Chanock RM. Visualization by immune electron microscopy of a 27-nm particle associated with acute infectious nonbacterial gastroenteritis. *Journal of virology*. 1972 Nov;10(5):1075-81. PubMed PMID: 4117963.

11. Jiang X, Wang M, Graham DY, Estes MK. Expression, self-assembly, and antigenicity of the Norwalk virus capsid protein. *Journal of virology*. 1992 Nov;66(11):6527-32. PubMed PMID: 1328679.

12. Greenberg HB, Valdesuso JR, Kalica AR, Wyatt RG, McAuliffe VJ, Kapikian AZ, et al. Proteins of Norwalk virus. *Journal of virology*. 1981 Mar;37(3):994-9. PubMed PMID: 6785451.

13. Prasad BV, Rothnagel R, Jiang X, Estes MK. Three-dimensional structure of baculovirus-expressed Norwalk virus capsids. *Journal of virology*. 1994 Aug;68(8):5117-25. PubMed PMID: 8035511.

14. Jiang X, Wang M, Wang K, Estes MK. Sequence and genomic organization of Norwalk virus. *Virology*. 1993 Jul;195(1):51-61. PubMed PMID: 8391187.

15. Bertolotti-Ciarlet A, Crawford SE, Hutson AM, Estes MK. The 3' end of Norwalk virus mRNA contains determinants that regulate the expression and stability of the viral capsid protein VP1: a novel function for the VP2 protein. *Journal of virology*. 2003 Nov;77(21):11603-15. PubMed PMID: 14557646.

16. Green KY, Ando T, Balayan MS, Berke T, Clarke IN, Estes MK, et al. Taxonomy of the caliciviruses. *The Journal of infectious diseases*. 2000 May;181 Suppl 2:S322-30. PubMed PMID: 10804145.

17. Clarke IN, Green KY, Hansman GS, Knowles NJ, Koopmans MK, Matson DO, Meyers G, Neill JD, Radford A, Smith AW, Studdert MJ, Thiel HJ, and Vinjé J. Caliciviridae. 2012. In: *Virus Taxonomy: Classification and Nomenclature of*

Viruses: Ninth Report of the International Committee on Taxonomy of Viruses [Internet]. San Diego: Elsevier; [977-86].

18. Ettayebi K, Crawford SE, Murakami K, Broughman JR, Karandikar U, Tenge VR, et al. Replication of human noroviruses in stem cell-derived human enteroids. *Science*. 2016 Sep 23;353(6306):1387-93. PubMed PMID: 27562956.

19. Vinje J. Advances in laboratory methods for detection and typing of norovirus. *Journal of clinical microbiology*. 2015 Feb;53(2):373-81. PubMed PMID: 24989606.

20. Kroneman A, Vega E, Vennema H, Vinje J, White PA, Hansman G, et al. Proposal for a unified norovirus nomenclature and genotyping. *Archives of virology*. 2013 Oct;158(10):2059-68. PubMed PMID: 23615870.

21. Hoa Tran TN, Trainor E, Nakagomi T, Cunliffe NA, Nakagomi O. Molecular epidemiology of noroviruses associated with acute sporadic gastroenteritis in children: global distribution of genogroups, genotypes and GII.4 variants. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology*. 2013 Mar;56(3):185-93. PubMed PMID: 23218993.

22. Kroneman A, Vennema H, Deforche K, v d Avoort H, Penaranda S, Oberste MS, et al. An automated genotyping tool for enteroviruses and noroviruses. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology*. 2011 Jun;51(2):121-5. PubMed PMID: 21514213.

23. Zheng DP, Ando T, Fankhauser RL, Beard RS, Glass RI, Monroe SS. Norovirus classification and proposed strain nomenclature. *Virology*. 2006 Mar 15;346(2):312-23. PubMed PMID: 16343580.

24. Xerry J, Gallimore CI, Iturriza-Gomara M, Allen DJ, Gray JJ. Transmission events within outbreaks of gastroenteritis determined through analysis of nucleotide sequences of the P2 domain of genogroup II noroviruses. *Journal of clinical microbiology*. 2008 Mar;46(3):947-53. PubMed PMID: 18216210.

25. Lopman B, Zambon M, Brown DW. The evolution of norovirus, the "gastric flu". *PLoS medicine*. 2008 Feb;5(2):e42. PubMed PMID: 18271623.

26. Donaldson EF, Lindesmith LC, Lobue AD, Baric RS. Viral shape-shifting: norovirus evasion of the human immune system. *Nature reviews Microbiology*. 2010 Mar;8(3):231-41. PubMed PMID: 20125087.
27. Sanjuan R, Nebot MR, Chirico N, Mansky LM, Belshaw R. Viral mutation rates. *Journal of virology*. 2010 Oct;84(19):9733-48. PubMed PMID: 20660197.
28. Nachman MW, Crowell SL. Estimate of the mutation rate per nucleotide in humans. *Genetics*. 2000 Sep;156(1):297-304. PubMed PMID: 10978293.
29. Lauring AS, Frydman J, Andino R. The role of mutational robustness in RNA virus evolution. *Nature reviews Microbiology*. 2013 May;11(5):327-36. PubMed PMID: 23524517.
30. Lauring AS, Andino R. Quasispecies theory and the behavior of RNA viruses. *PLoS pathogens*. 2010 Jul 22;6(7):e1001005. PubMed PMID: 20661479.
31. Vashist S, Bailey D, Putics A, Goodfellow I. Model systems for the study of human norovirus Biology. *Future virology*. 2009 Jul;4(4):353-67. PubMed PMID: 21516251.
32. Thorne LG, Goodfellow IG. Norovirus gene expression and replication. *The Journal of general virology*. 2014 Feb;95(Pt 2):278-91. PubMed PMID: 24243731.
33. Donaldson EF, Lindesmith LC, Lobue AD, Baric RS. Norovirus pathogenesis: mechanisms of persistence and immune evasion in human populations. *Immunological reviews*. 2008 Oct;225:190-211. PubMed PMID: 18837783.
34. Agus SG, Dolin R, Wyatt RG, Tousimis AJ, Northrup RS. Acute infectious nonbacterial gastroenteritis: intestinal histopathology. Histologic and enzymatic alterations during illness produced by the Norwalk agent in man. *Annals of internal medicine*. 1973 Jul;79(1):18-25. PubMed PMID: 4721173.
35. Levy AG, Widerlite L, Schwartz CJ, Dolin R, Blacklow NR, Gardner JD, et al. Jejunal adenylate cyclase activity in human subjects during viral gastroenteritis. *Gastroenterology*. 1976 Mar;70(3):321-5. PubMed PMID: 765182.

36. Meeroff JC, Schreiber DS, Trier JS, Blacklow NR. Abnormal gastric motor function in viral gastroenteritis. *Annals of internal medicine*. 1980 Mar;92(3):370-3. PubMed PMID: 6766695.
37. Karandikar UC, Crawford SE, Ajami NJ, Murakami K, Kou B, Ettayebi K, et al. Detection of human norovirus in intestinal biopsies from immunocompromised transplant patients. *The Journal of general virology*. 2016 Sep;97(9):2291-300. PubMed PMID: 27412790.
38. Ahmed SM, Hall AJ, Robinson AE, Verhoef L, Premkumar P, Parashar UD, et al. Global prevalence of norovirus in cases of gastroenteritis: a systematic review and meta-analysis. *The Lancet infectious diseases*. 2014 Aug;14(8):725-30. PubMed PMID: 24981041.
39. Phillips G, Tam CC, Conti S, Rodrigues LC, Brown D, Iturriza-Gomara M, et al. Community incidence of norovirus-associated infectious intestinal disease in England: improved estimates using viral load for norovirus diagnosis. *American journal of epidemiology*. 2010 May 01;171(9):1014-22. PubMed PMID: 20360244.
40. Saito M, Goel-Apaza S, Espetia S, Velasquez D, Cabrera L, Loli S, et al. Multiple norovirus infections in a birth cohort in a Peruvian Periurban community. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2014 Feb;58(4):483-91. PubMed PMID: 24300042.
41. Atmar RL, Opekun AR, Gilger MA, Estes MK, Crawford SE, Neill FH, et al. Determination of the 50% human infectious dose for Norwalk virus. *The Journal of infectious diseases*. 2014 Apr 01;209(7):1016-22. PubMed PMID: 24253285.
42. Harris JP, Edmunds WJ, Pebody R, Brown DW, Lopman BA. Deaths from norovirus among the elderly, England and Wales. *Emerging infectious diseases*. 2008 Oct;14(10):1546-52. PubMed PMID: 18826817.
43. Trivedi TK, Desai R, Hall AJ, Patel M, Parashar UD, Lopman BA. Clinical characteristics of norovirus-associated deaths: a systematic literature review. *American journal of infection control*. 2013 Jul;41(7):654-7. PubMed PMID: 23266383.
44. Parashar UD, Monroe SS. "Norwalk-like viruses" as a cause of foodborne disease outbreaks. *Reviews in medical virology*. 2001 Jul-Aug;11(4):243-52. PubMed PMID: 11479930.

45. Patel MM, Widdowson MA, Glass RI, Akazawa K, Vinje J, Parashar UD. Systematic literature review of role of noroviruses in sporadic gastroenteritis. *Emerging infectious diseases*. 2008 Aug;14(8):1224-31. PubMed PMID: 18680645.
46. Tam CC, Rodrigues LC, Viviani L, Dodds JP, Evans MR, Hunter PR, et al. Longitudinal study of infectious intestinal disease in the UK (IID2 study): incidence in the community and presenting to general practice. *Gut*. 2012 Jan;61(1):69-77. PubMed PMID: 21708822.
47. Parrino TA, Schreiber DS, Trier JS, Kapikian AZ, Blacklow NR. Clinical immunity in acute gastroenteritis caused by Norwalk agent. *The New England journal of medicine*. 1977 Jul 14;297(2):86-9. PubMed PMID: 405590.
48. Lindesmith L, Moe C, Marionneau S, Ruvoen N, Jiang X, Lindblad L, et al. Human susceptibility and resistance to Norwalk virus infection. *Nature medicine*. 2003 May;9(5):548-53. PubMed PMID: 12692541.
49. de Graaf M, van Beek J, Koopmans MP. Human norovirus transmission and evolution in a changing world. *Nature reviews Microbiology*. 2016 Jul;14(7):421-33. PubMed PMID: 27211790.
50. Thorven M, Grahn A, Hedlund KO, Johansson H, Wahlfrid C, Larson G, et al. A homozygous nonsense mutation (428G-->A) in the human secretor (FUT2) gene provides resistance to symptomatic norovirus (GGII) infections. *Journal of virology*. 2005 Dec;79(24):15351-5. PubMed PMID: 16306606.
51. Le Pendu J, Ruvoen-Clouet N, Kindberg E, Svensson L. Mendelian resistance to human norovirus infections. *Seminars in immunology*. 2006 Dec;18(6):375-86. PubMed PMID: 16973373.
52. Wyatt RG, Dolin R, Blacklow NR, DuPont HL, Buscho RF, Thornhill TS, et al. Comparison of three agents of acute infectious nonbacterial gastroenteritis by cross-challenge in volunteers. *The Journal of infectious diseases*. 1974 Jun;129(6):709-14. PubMed PMID: 4209723.
53. Lopman BA, Reacher MH, Vipond IB, Sarangi J, Brown DW. Clinical manifestation of norovirus gastroenteritis in health care settings. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2004 Aug 1;39(3):318-24. PubMed PMID: 15306997.
54. Rockx B, De Wit M, Vennema H, Vinje J, De Bruin E, Van Duynhoven Y, et al. Natural history of human calicivirus infection: a prospective cohort

study. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2002 Aug 1;35(3):246-53. PubMed PMID: 12115089.

55. Glass RI, Parashar UD, Estes MK. Norovirus gastroenteritis. *The New England journal of medicine*. 2009 Oct 29;361(18):1776-85. PubMed PMID: 19864676.

56. World Health Organization. *The Treatment of diarrhoea : a manual for physicians and other senior health workers 2005* [cited 2016 15 Jun]. Available from:

<http://apps.who.int/iris/bitstream/handle/10665/43209/9241593180.pdf;jsessionid=884220BB2CB1E9DAA4E073001AF9EF37?sequence=1>.

57. Kaplan JE, Feldman R, Campbell DS, Lookabaugh C, Gary GW. The frequency of a Norwalk-like pattern of illness in outbreaks of acute gastroenteritis. *American journal of public health*. 1982 Dec;72(12):1329-32. PubMed PMID: 6291414.

58. Kirby A, Iturriza-Gomara M. Norovirus diagnostics: options, applications and interpretations. Expert review of anti-infective therapy. 2012 Apr;10(4):423-33. PubMed PMID: 22512752.

59. Platts-Mills JA, Liu J, Houpt ER. New concepts in diagnostics for infectious diarrhea. *Mucosal immunology*. 2013 Sep;6(5):876-85. PubMed PMID: 23881355.

60. Dunbar SA. Applications of Luminex xMAP technology for rapid, high-throughput multiplexed nucleic acid detection. *Clinica chimica acta; international journal of clinical chemistry*. 2006 Jan;363(1-2):71-82. PubMed PMID: 16102740.

61. McAuliffe G, Bissessor L, Williamson D, Moore S, Wilson J, Dufour M, et al. Use of the EntericBio Gastro Panel II in a diagnostic microbiology laboratory: challenges and opportunities. *Pathology*. 2017 Jun;49(4):419-22. PubMed PMID: 28438393.

62. Kirby A, Ashton L, Hart IJ. Detection of norovirus infection in the hospital setting using vomit samples. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology*. 2011 May;51(1):86-7. PubMed PMID: 21376659.

63. Kirby A, Dove W, Ashton L, Hopkins M, Cunliffe NA. Detection of norovirus in mouthwash samples from patients with acute gastroenteritis. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology*. 2010 Aug;48(4):285-7. PubMed PMID: 20637439.
64. Lazzarini M, Ronfani L. Oral zinc for treating diarrhoea in children. *The Cochrane database of systematic reviews*. 2013;1:CD005436. PubMed PMID: 23440801.
65. Banyai K, Estes MK, Martella V, Parashar UD. Viral gastroenteritis. *Lancet*. 2018 Jul 14;392(10142):175-86. PubMed PMID: 30025810.
66. Brown LK, Clark I, Brown JR, Breuer J, Lowe DM. Norovirus infection in primary immune deficiency. *Reviews in medical virology*. 2017 Mar 8. PubMed PMID: 28271593.
67. Lucero Y, Vidal R, O'Ryan GM. Norovirus vaccines under development. *Vaccine*. 2018 Aug 28;36(36):5435-41. PubMed PMID: 28668568.
68. PHE. Guidelines for the management of norovirus outbreaks in acute and community health and social care settings. 2012.
69. CDC. Guideline for the prevention and control of norovirus gastroenteritis outbreaks in healthcare settings. 2011.
70. Shulman ST, Friedmann HC, Sims RH. Theodor Escherich: the first pediatric infectious diseases physician? *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2007 Oct 15;45(8):1025-9. PubMed PMID: 17879920.
71. Kauffmann F. The serology of the coli group. *Journal of immunology*. 1947 Sep;57(1):71-100. PubMed PMID: 20264689.
72. Levine MM. Escherichia coli that cause diarrhea: enterotoxigenic, enteropathogenic, enteroinvasive, enterohemorrhagic, and enteroadherent. *The Journal of infectious diseases*. 1987 Mar;155(3):377-89. PubMed PMID: 3543152.
73. Hunter PR. Drinking water and diarrhoeal disease due to Escherichia coli. *Journal of water and health*. 2003 Jun;1(2):65-72. PubMed PMID: 15382735.

74. Dubreuil JD. The whole Shebang: the gastrointestinal tract, *Escherichia coli* enterotoxins and secretion. *Current issues in molecular biology*. 2012;14(2):71-82. PubMed PMID: 22368232.
75. Nataro JP, Kaper JB. Diarrheagenic *Escherichia coli*. *Clinical microbiology reviews*. 1998 Jan;11(1):142-201. PubMed PMID: 9457432.
76. Fleckenstein JM, Hardwidge PR, Munson GP, Rasko DA, Sommerfelt H, Steinsland H. Molecular mechanisms of enterotoxigenic *Escherichia coli* infection. *Microbes and infection / Institut Pasteur*. 2010 Feb;12(2):89-98. PubMed PMID: 19883790.
77. Kesty NC, Mason KM, Reedy M, Miller SE, Kuehn MJ. Enterotoxigenic *Escherichia coli* vesicles target toxin delivery into mammalian cells. *The EMBO journal*. 2004 Nov 24;23(23):4538-49. PubMed PMID: 15549136.
78. Vaandrager AB, Tilly BC, Smolenski A, Schneider-Rasp S, Bot AG, Edixhoven M, et al. cGMP stimulation of cystic fibrosis transmembrane conductance regulator Cl⁻ channels co-expressed with cGMP-dependent protein kinase type II but not type Iβ. *The Journal of biological chemistry*. 1997 Feb 14;272(7):4195-200. PubMed PMID: 9020133.
79. Hug MJ, Tamada T, Bridges RJ. CFTR and bicarbonate secretion by [correction of to] epithelial cells. *News in physiological sciences : an international journal of physiology produced jointly by the International Union of Physiological Sciences and the American Physiological Society*. 2003 Feb;18:38-42. PubMed PMID: 12531931.
80. Steinsland H, Valentiner-Branth P, Perch M, Dias F, Fischer TK, Aaby P, et al. Enterotoxigenic *Escherichia coli* infections and diarrhea in a cohort of young children in Guinea-Bissau. *The Journal of infectious diseases*. 2002 Dec 15;186(12):1740-7. PubMed PMID: 12447759.
81. Yamanaka H, Nomura T, Fujii Y, Okamoto K. Need for TolC, an *Escherichia coli* outer membrane protein, in the secretion of heat-stable enterotoxin I across the outer membrane. *Microbial pathogenesis*. 1998 Sep;25(3):111-20. PubMed PMID: 9790870.
82. Chao AC, de Sauvage FJ, Dong YJ, Wagner JA, Goeddel DV, Gardner P. Activation of intestinal CFTR Cl⁻ channel by heat-stable enterotoxin and guanylin via cAMP-dependent protein kinase. *The EMBO journal*. 1994 Mar 1;13(5):1065-72. PubMed PMID: 7510634.

83. Isidean SD, Riddle MS, Savarino SJ, Porter CK. A systematic review of ETEC epidemiology focusing on colonization factor and toxin expression. *Vaccine*. 2011 Aug 26;29(37):6167-78. PubMed PMID: 21723899.
84. Gaastra W, Svennerholm AM. Colonization factors of human enterotoxigenic *Escherichia coli* (ETEC). *Trends in microbiology*. 1996 Nov;4(11):444-52. PubMed PMID: 8950814.
85. Huffman SL, Combest C. Role of breast-feeding in the prevention and treatment of diarrhoea. *Journal of diarrhoeal diseases research*. 1990 Sep;8(3):68-81. PubMed PMID: 2243179.
86. Khalil IA, Troeger C, Blacker BF, Rao PC, Brown A, Atherly DE, et al. Morbidity and mortality due to shigella and enterotoxigenic *Escherichia coli* diarrhoea: the Global Burden of Disease Study 1990-2016. *The Lancet infectious diseases*. 2018 Nov;18(11):1229-40. PubMed PMID: 30266330.
87. DuPont HL, Formal SB, Hornick RB, Snyder MJ, Libonati JP, Sheahan DG, et al. Pathogenesis of *Escherichia coli* diarrhea. *The New England journal of medicine*. 1971 Jul 1;285(1):1-9. PubMed PMID: 4996788.
88. Kean BH. The Diarrhea of Travelers to Mexico. Summary of Five-Year Study. *Annals of internal medicine*. 1963 Nov;59:605-14. PubMed PMID: 14082713.
89. Gorbach SL, Kean BH, Evans DG, Evans DJ, Jr., Bessudo D. Travelers' diarrhea and toxigenic *Escherichia coli*. *The New England journal of medicine*. 1975 May 1;292(18):933-6. PubMed PMID: 163964.
90. Hill DR, Beeching NJ. Travelers' diarrhea. *Current opinion in infectious diseases*. 2010 Oct;23(5):481-7. PubMed PMID: 20683261.
91. Steffen R, Hill DR, DuPont HL. Traveler's diarrhea: a clinical review. *Jama*. 2015 Jan 6;313(1):71-80. PubMed PMID: 25562268.
92. Barrett J, Brown M. Travellers' diarrhoea. *Bmj*. 2016 Apr 19;353:i1937. PubMed PMID: 27094342.
93. Qadri F, Svennerholm AM, Faruque AS, Sack RB. Enterotoxigenic *Escherichia coli* in developing countries: epidemiology, microbiology, clinical features, treatment, and prevention. *Clinical microbiology reviews*. 2005 Jul;18(3):465-83. PubMed PMID: 16020685.

94. Zhang W, Sack DA. Progress and hurdles in the development of vaccines against enterotoxigenic *Escherichia coli* in humans. Expert review of vaccines. 2012 Jun;11(6):677-94. PubMed PMID: 22873126.
95. Levine MM, Nalin DR, Hoover DL, Bergquist EJ, Hornick RB, Young CR. Immunity to enterotoxigenic *Escherichia coli*. Infection and immunity. 1979 Mar;23(3):729-36. PubMed PMID: 378834.
96. Svennerholm AMaQF. Mucosal immune responses against enterotoxigenic *Escherichia coli* [ETEC] in humans. In: Vajdy M, editor. Immunity against mucosal pathogens: Springer; 2008. p. 153-71.
97. Nalin DR, McLaughlin JC, Rahaman M, Yunus M, Curlin G. Enterotoxigenic *Escherichia coli* and idiopathic diarrhoea in Bangladesh. Lancet. 1975 Dec 6;2(7945):1116-9. PubMed PMID: 53601.
98. Levine MM, Caplan ES, Waterman D, Cash RA, Hornick RB, Snyder MJ. Diarrhea caused by *Escherichia coli* that produce only heat-stable enterotoxin. Infection and immunity. 1977 Jul;17(1):78-82. PubMed PMID: 328397.
99. De SN, Bhattacharya K, Sarkar JK. A study of the pathogenicity of strains of *Bacterium coli* from acute and chronic enteritis. The Journal of pathology and bacteriology. 1956 Jan;71(1):201-9. PubMed PMID: 13307349.
100. Dean AG, Ching YC, Williams RG, Harden LB. Test for *Escherichia coli* enterotoxin using infant mice: application in a study of diarrhea in children in Honolulu. The Journal of infectious diseases. 1972 Apr;125(4):407-11. PubMed PMID: 4553083.
101. Donta ST, Smith DM. Stimulation of steroidogenesis in tissue culture by enterotoxigenic *Escherichia coli* and its neutralization by specific antiserum. Infection and immunity. 1974 Mar;9(3):500-5. PubMed PMID: 4593468.
102. Guerrant RL, Brunton LL, Schnaitman TC, Rebhun LI, Gilman AG. Cyclic adenosine monophosphate and alteration of Chinese hamster ovary cell morphology: a rapid, sensitive in vitro assay for the enterotoxins of *Vibrio cholerae* and *Escherichia coli*. Infection and immunity. 1974 Aug;10(2):320-7. PubMed PMID: 4368545.
103. Giannella RA, Drake KW, Luttrell M. Development of a radioimmunoassay for *Escherichia coli* heat-stable enterotoxin: comparison with the suckling mouse bioassay. Infection and immunity. 1981 Jul;33(1):186-92. PubMed PMID: 7021423.

104. Sack RB. Enterotoxigenic *Escherichia coli*: identification and characterization. *The Journal of infectious diseases*. 1980 Aug;142(2):279-86. PubMed PMID: 6106031.
105. Al-Abri SS, Beeching NJ, Nye FJ. Traveller's diarrhoea. *The Lancet infectious diseases*. 2005 Jun;5(6):349-60. PubMed PMID: 15919621.
106. PATH. The case for investment in enterotoxigenic *Escherichia coli* vaccines. Seattle, WA: PATH; 2011 [cited 2016 Feb]. Available from: <http://www.path.org/publications/files/VAC-etec-investment-rpt.pdf>.
107. Ahmed T, Bhuiyan TR, Zaman K, Sinclair D, Qadri F. Vaccines for preventing enterotoxigenic *Escherichia coli* (ETEC) diarrhoea. *The Cochrane database of systematic reviews*. 2013;7:CD009029. PubMed PMID: 23828581.
108. Evans DJ, Jr., Evans DG, Opekun AR, Graham DY. Immunoprotective oral whole cell vaccine for enterotoxigenic *Escherichia coli* diarrhea prepared by in situ destruction of chromosomal and plasmid DNA with colicin E2. *FEMS microbiology immunology*. 1988 Jan;1(1):9-18. PubMed PMID: 3078575.
109. Walker RI. An assessment of enterotoxigenic *Escherichia coli* and *Shigella* vaccine candidates for infants and children. *Vaccine*. 2015 Feb 18;33(8):954-65. PubMed PMID: 25482842.
110. Walker RI, Wierzba TF, Mani S, Bourgeois AL. Vaccines against *Shigella* and enterotoxigenic *Escherichia coli*: A summary of the 2016 VASE Conference. *Vaccine*. 2017 Dec 14;35(49 Pt A):6775-82. PubMed PMID: 28987444.
111. Cunliffe NA, Ngwira BM, Dove W, Thindwa BD, Turner AM, Broadhead RL, et al. Epidemiology of rotavirus infection in children in Blantyre, Malawi, 1997-2007. *The Journal of infectious diseases*. 2010 Sep 1;202 Suppl:S168-74. PubMed PMID: 20684698.
112. Mason J, Iturriza-Gomara M, O'Brien SJ, Ngwira BM, Dove W, Maiden MC, et al. *Campylobacter* infection in children in Malawi is common and is frequently associated with enteric virus co-infections. *PloS one*. 2013;8(3):e59663. PubMed PMID: 23555739.
113. Wheeler JG, Sethi D, Cowden JM, Wall PG, Rodrigues LC, Tompkins DS, et al. Study of infectious intestinal disease in England: rates in the community, presenting to general practice, and reported to national surveillance. *The*

Infectious Intestinal Disease Study Executive. Bmj. 1999 Apr 17;318(7190):1046-50. PubMed PMID: 10205103.

114. Cunliffe NA, Gondwe JS, Broadhead RL, Molyneux ME, Woods PA, Bresee JS, et al. Rotavirus G and P types in children with acute diarrhea in Blantyre, Malawi, from 1997 to 1998: predominance of novel P[6]G8 strains. Journal of medical virology. 1999 Mar;57(3):308-12. PubMed PMID: 10022804.

115. Cunliffe NA, Gentsch JR, Kirkwood CD, Gondwe JS, Dove W, Nakagomi O, et al. Molecular and serologic characterization of novel serotype G8 human rotavirus strains detected in Blantyre, Malawi. Virology. 2000 Sep 1;274(2):309-20. PubMed PMID: 10964774.

116. Cunliffe NA, Dove W, Jiang B, Thinwda Cert BD, Broadhead RL, Molyneux ME, et al. Detection of group C rotavirus in children with acute gastroenteritis in Blantyre, Malawi. The Pediatric infectious disease journal. 2001 Nov;20(11):1088-90. PubMed PMID: 11734719.

117. Iturriza-Gomara M, Elliot AJ, Dockery C, Fleming DM, Gray JJ. Structured surveillance of infectious intestinal disease in pre-school children in the community: 'The Nappy Study'. Epidemiology and infection. 2009 Jul;137(7):922-31. PubMed PMID: 19017426.

118. Cunliffe NA, Gondwe JS, Graham SM, Thindwa BD, Dove W, Broadhead RL, et al. Rotavirus strain diversity in Blantyre, Malawi, from 1997 to 1999. Journal of clinical microbiology. 2001 Mar;39(3):836-43. PubMed PMID: 11230392.

119. Kanyuka M, Ndawala J, Mleme T, Chisesa L, Makwemba M, Amouzou A, et al. Malawi and Millennium Development Goal 4: a Countdown to 2015 country case study. The Lancet Global health. 2016 Mar;4(3):e201-14. PubMed PMID: 26805586.

120. O'Brien SJ, Rait G, Hunter PR, Gray JJ, Bolton FJ, Tompkins DS, et al. Methods for determining disease burden and calibrating national surveillance data in the United Kingdom: the second study of infectious intestinal disease in the community (IID2 study). BMC medical research methodology. 2010;10:39. PubMed PMID: 20444246.

121. Trainor E, Lopman B, Iturriza-Gomara M, Dove W, Ngwira B, Nakagomi O, et al. Detection and molecular characterisation of noroviruses in hospitalised children in Malawi, 1997-2007. Journal of medical virology. 2013 Jul;85(7):1299-306. PubMed PMID: 23918547.

122. Kageyama T, Kojima S, Shinohara M, Uchida K, Fukushi S, Hoshino FB, et al. Broadly reactive and highly sensitive assay for Norwalk-like viruses based on real-time quantitative reverse transcription-PCR. *Journal of clinical microbiology*. 2003 Apr;41(4):1548-57. PubMed PMID: 12682144. Epub 2003/04/19. eng.
123. Amar CF, East CL, Gray J, Iturriza-Gomara M, Maclure EA, McLauchlin J. Detection by PCR of eight groups of enteric pathogens in 4,627 faecal samples: re-examination of the English case-control Infectious Intestinal Disease Study (1993-1996). *European journal of clinical microbiology & infectious diseases* : official publication of the European Society of Clinical Microbiology. 2007 May;26(5):311-23. PubMed PMID: 17447091. Epub 2007/04/21. eng.
124. Allen DJ, Gray JJ, Gallimore CI, Xerry J, Iturriza-Gomara M. Analysis of amino acid variation in the P2 domain of the GII-4 norovirus VP1 protein reveals putative variant-specific epitopes. *PloS one*. 2008;3(1):e1485. PubMed PMID: 18213393.
125. Kojima S, Kageyama T, Fukushi S, Hoshino FB, Shinohara M, Uchida K, et al. Genogroup-specific PCR primers for detection of Norwalk-like viruses. *Journal of virological methods*. 2002 Feb;100(1-2):107-14. PubMed PMID: 11742657.
126. Chhabra P, Walimbe AM, Chitambar SD. Complete genome characterization of Genogroup II norovirus strains from India: Evidence of recombination in ORF2/3 overlap. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 2010 Oct;10(7):1101-9. PubMed PMID: 20633703.
127. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution*. 2011 Oct;28(10):2731-9. PubMed PMID: 21546353.
128. Muller D, Greune L, Heusipp G, Karch H, Fruth A, Tschape H, et al. Identification of unconventional intestinal pathogenic *Escherichia coli* isolates expressing intermediate virulence factor profiles by using a novel single-step multiplex PCR. *Applied and environmental microbiology*. 2007 May;73(10):3380-90. PubMed PMID: 17400780.
129. Kosek M, Bern C, Guerrant RL. The global burden of diarrhoeal disease, as estimated from studies published between 1992 and 2000. *Bulletin of the World Health Organization*. 2003;81(3):197-204. PubMed PMID: 12764516.

130. Bar-Zeev N, King C, Phiri T, Beard J, Mvula H, Crampin AC, et al. Impact of monovalent rotavirus vaccine on diarrhoea-associated post-neonatal infant mortality in rural communities in Malawi: a population-based birth cohort study. *The Lancet Global health*. 2018 Sep;6(9):e1036-e44. PubMed PMID: 30103981.
131. Bourgeois AL, Wierzba TF, Walker RI. Status of vaccine research and development for enterotoxigenic *Escherichia coli*. *Vaccine*. 2016 Jun 3;34(26):2880-6. PubMed PMID: 26988259.
132. Riddle MS, Chen WH, Kirkwood CD, MacLennan CA. Update on vaccines for enteric pathogens. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*. 2018 Oct;24(10):1039-45. PubMed PMID: 29964231.
133. Scaletsky IC, Fabbricotti SH, Aranda KR, Morais MB, Fagundes-Neto U. Comparison of DNA hybridization and PCR assays for detection of putative pathogenic enteroadherent *Escherichia coli*. *Journal of clinical microbiology*. 2002 Apr;40(4):1254-8. PubMed PMID: 11923341.
134. Rahouma A, Klena JD, Krema Z, Abobker AA, Treesh K, Franka E, et al. Enteric pathogens associated with childhood diarrhea in Tripoli-Libya. *The American journal of tropical medicine and hygiene*. 2011 Jun;84(6):886-91. PubMed PMID: 21633024.
135. Al-Gallas N, Bahri O, Bouratbeen A, Ben Haasen A, Ben Aissa R. Etiology of acute diarrhea in children and adults in Tunis, Tunisia, with emphasis on diarrheagenic *Escherichia coli*: prevalence, phenotyping, and molecular epidemiology. *The American journal of tropical medicine and hygiene*. 2007 Sep;77(3):571-82. PubMed PMID: 17827382.
136. Geyer A, Crewe-Brown HH, Greeff AS, Fripp PJ, Steele AD, Van Schalkwyk TV, et al. The microbial aetiology of summer paediatric gastroenteritis at Ga-Rankuwa Hospital in South Africa. *East African medical journal*. 1993 Feb;70(2):78-81. PubMed PMID: 8513746.
137. Bonkougou IJ, Lienemann T, Martikainen O, Dembele R, Sanou I, Traore AS, et al. Diarrhoeagenic *Escherichia coli* detected by 16-plex PCR in children with and without diarrhoea in Burkina Faso. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*. 2012 Sep;18(9):901-6. PubMed PMID: 21985619.

138. Gascon J, Vargas M, Schellenberg D, Urassa H, Casals C, Kahigwa E, et al. Diarrhea in children under 5 years of age from Ifakara, Tanzania: a case-control study. *Journal of clinical microbiology*. 2000 Dec;38(12):4459-62. PubMed PMID: 11101580.
139. Opintan JA, Bishar RA, Newman MJ, Okeke IN. Carriage of diarrhoeagenic *Escherichia coli* by older children and adults in Accra, Ghana. *Transactions of the Royal Society of Tropical Medicine and Hygiene*. 2010 Jul;104(7):504-6. PubMed PMID: 20307897.
140. Okeke IN, Lamikanra A, Steinruck H, Kaper JB. Characterization of *Escherichia coli* strains from cases of childhood diarrhea in provincial southwestern Nigeria. *Journal of clinical microbiology*. 2000 Jan;38(1):7-12. PubMed PMID: 10618054.
141. Levine MM, Robins-Browne RM. Factors that explain excretion of enteric pathogens by persons without diarrhea. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2012 Dec;55 Suppl 4:S303-11. PubMed PMID: 23169942.
142. Rao MR, Abu-Elyazeed R, Savarino SJ, Naficy AB, Wierzba TF, Abdel-Messih I, et al. High disease burden of diarrhea due to enterotoxigenic *Escherichia coli* among rural Egyptian infants and young children. *Journal of clinical microbiology*. 2003 Oct;41(10):4862-4. PubMed PMID: 14532244.
143. Vargas M, Gascon J, Casals C, Schellenberg D, Urassa H, Kahigwa E, et al. Etiology of diarrhea in children less than five years of age in Ifakara, Tanzania. *The American journal of tropical medicine and hygiene*. 2004 May;70(5):536-9. PubMed PMID: 15155987.
144. Mandomando IM, Macete EV, Ruiz J, Sanz S, Abacassamo F, Valles X, et al. Etiology of diarrhea in children younger than 5 years of age admitted in a rural hospital of southern Mozambique. *The American journal of tropical medicine and hygiene*. 2007 Mar;76(3):522-7. PubMed PMID: 17360878.
145. Saidi SM, Iijima Y, Sang WK, Mwangudza AK, Oundo JO, Taga K, et al. Epidemiological study on infectious diarrheal diseases in children in a coastal rural area of Kenya. *Microbiology and immunology*. 1997;41(10):773-8. PubMed PMID: 9403500.
146. Ali MM, Mohamed ZK, Klena JD, Ahmed SF, Moussa TA, Ghenghesh KS. Molecular characterization of diarrheagenic *Escherichia coli* from Libya. *The American journal of tropical medicine and hygiene*. 2012 May;86(5):866-71. PubMed PMID: 22556089.

147. Casalino M, Yusuf MW, Nicoletti M, Bazzicalupo P, Coppo A, Colonna B, et al. A two-year study of enteric infections associated with diarrhoeal diseases in children in urban Somalia. *Transactions of the Royal Society of Tropical Medicine and Hygiene*. 1988;82(4):637-41. PubMed PMID: 3076719.
148. Porat N, Levy A, Fraser D, Deckelbaum RJ, Dagan R. Prevalence of intestinal infections caused by diarrheagenic *Escherichia coli* in Bedouin infants and young children in Southern Israel. *The Pediatric infectious disease journal*. 1998 Jun;17(6):482-8. PubMed PMID: 9655539.
149. Iturriza-Gomara M, Jere KC, Hungerford D, Bar-Zeev N, Shioda K, Kanjerwa O, et al. Aetiology of diarrhoea among hospitalised children in Blantyre, Malawi following rotavirus vaccine introduction: a case-control study. *The Journal of infectious diseases*. 2019 Feb 28. PubMed PMID: 30816414.
150. Phillips G, Lopman B, Tam CC, Iturriza-Gomara M, Brown D, Gray J. Diagnosing norovirus-associated infectious intestinal disease using viral load. *BMC infectious diseases*. 2009 May 14;9:63. PubMed PMID: 19442278.
151. Phillips G, Lopman B, Rodrigues LC, Tam CC. Asymptomatic rotavirus infections in England: prevalence, characteristics, and risk factors. *American journal of epidemiology*. 2010 May 1;171(9):1023-30. PubMed PMID: 20392863.
152. Neog BK, Barman NN, Bora DP, Dey SC, Chakraborty A. Experimental infection of pigs with group A rotavirus and enterotoxigenic *Escherichia coli* in India: gross, histopathological and immunopathological study. *Veterinaria italiana*. 2011 Apr-Jun;47(2):117-28. PubMed PMID: 21706463.
153. Tzipori S, Chandler D, Makin T, Smith M. *Escherichia coli* and rotavirus infections in four-week-old gnotobiotic piglets fed milk or dry food. *Australian veterinary journal*. 1980 Jun;56(6):279-84. PubMed PMID: 6252884.
154. Mortality GBD, Causes of Death C. 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 Jan 10;385(9963):117-71. PubMed PMID: 25530442.
155. Campbell DA. An Update on the United Nations Millennium Development Goals. *Journal of obstetric, gynecologic, and neonatal nursing : JOGNN*. 2017 May - Jun;46(3):e48-e55. PubMed PMID: 28285005.

156. Burnett E, Jonesteller CL, Tate JE, Yen C, Parashar UD. Global Impact of Rotavirus Vaccination on Childhood Hospitalizations and Mortality From Diarrhea. *The Journal of infectious diseases*. 2017 Jun 1;215(11):1666-72. PubMed PMID: 28430997.
157. Global Health Data Exchange 2013. Available from: <http://vizhub.healthdata.org/irank/arrow.php>.
158. Lanata CF, Fischer-Walker CL, Olascoaga AC, Torres CX, Aryee MJ, Black RE, et al. Global causes of diarrheal disease mortality in children <5 years of age: a systematic review. *PloS one*. 2013;8(9):e72788. PubMed PMID: 24023773.
159. Dingle KE, Lambden PR, Caul EO, Clarke IN. Human enteric Caliciviridae: the complete genome sequence and expression of virus-like particles from a genetic group II small round structured virus. *The Journal of general virology*. 1995 Sep;76 (Pt 9):2349-55. PubMed PMID: 7561776.
160. Hardy ME. Norovirus protein structure and function. *FEMS microbiology letters*. 2005 Dec 1;253(1):1-8. PubMed PMID: 16168575.
161. Prasad BV, Hardy ME, Dokland T, Bella J, Rossmann MG, Estes MK. X-ray crystallographic structure of the Norwalk virus capsid. *Science*. 1999 Oct 8;286(5438):287-90. PubMed PMID: 10514371.
162. Tan M, Hegde RS, Jiang X. The P domain of norovirus capsid protein forms dimer and binds to histo-blood group antigen receptors. *Journal of virology*. 2004 Jun;78(12):6233-42. PubMed PMID: 15163716.
163. Lochridge VP, Jutila KL, Graff JW, Hardy ME. Epitopes in the P2 domain of norovirus VP1 recognized by monoclonal antibodies that block cell interactions. *The Journal of general virology*. 2005 Oct;86(Pt 10):2799-806. PubMed PMID: 16186235.
164. Noel JS, Fankhauser RL, Ando T, Monroe SS, Glass RI. Identification of a distinct common strain of "Norwalk-like viruses" having a global distribution. *The Journal of infectious diseases*. 1999 Jun;179(6):1334-44. PubMed PMID: 10228052.
165. Siebenga JJ, Vennema H, Zheng DP, Vinje J, Lee BE, Pang XL, et al. Norovirus illness is a global problem: emergence and spread of norovirus GII.4 variants, 2001-2007. *The Journal of infectious diseases*. 2009 Sep 1;200(5):802-12. PubMed PMID: 19627248.

166. Vega E, Barclay L, Gregoricus N, Williams K, Lee D, Vinje J. Novel surveillance network for norovirus gastroenteritis outbreaks, United States. *Emerging infectious diseases*. 2011 Aug;17(8):1389-95. PubMed PMID: 21801614.
167. Lopman B, Vennema H, Kohli E, Pothier P, Sanchez A, Negredo A, et al. Increase in viral gastroenteritis outbreaks in Europe and epidemic spread of new norovirus variant. *Lancet*. 2004 Feb 28;363(9410):682-8. PubMed PMID: 15001325.
168. Bull RA, Tu ET, McIver CJ, Rawlinson WD, White PA. Emergence of a new norovirus genotype II.4 variant associated with global outbreaks of gastroenteritis. *Journal of clinical microbiology*. 2006 Feb;44(2):327-33. PubMed PMID: 16455879.
169. Eden JS, Bull RA, Tu E, McIver CJ, Lyon MJ, Marshall JA, et al. Norovirus GII.4 variant 2006b caused epidemics of acute gastroenteritis in Australia during 2007 and 2008. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology*. 2010 Dec;49(4):265-71. PubMed PMID: 20888289.
170. van Beek J, Ambert-Balay K, Botteldoorn N, Eden JS, Fonager J, Hewitt J, et al. Indications for worldwide increased norovirus activity associated with emergence of a new variant of genotype II.4, late 2012. *Euro surveillance : bulletin European sur les maladies transmissibles = European communicable disease bulletin*. 2013;18(1):8-9. PubMed PMID: 23305715.
171. Tohma K, Lepore CJ, Ford-Siltz LA, Parra GI. Phylogenetic Analyses Suggest that Factors Other Than the Capsid Protein Play a Role in the Epidemic Potential of GII.2 Norovirus. *mSphere*. 2017 May-Jun;2(3). PubMed PMID: 28529975.
172. Matsui SM, Greenberg HB. Immunity to calicivirus infection. *The Journal of infectious diseases*. 2000 May;181 Suppl 2:S331-5. PubMed PMID: 10804146.
173. Bull RA, Tanaka MM, White PA. Norovirus recombination. *The Journal of general virology*. 2007 Dec;88(Pt 12):3347-59. PubMed PMID: 18024905.
174. Lindesmith LC, Donaldson EF, Lobue AD, Cannon JL, Zheng DP, Vinje J, et al. Mechanisms of GII.4 norovirus persistence in human populations. *PLoS medicine*. 2008 Feb;5(2):e31. PubMed PMID: 18271619.

175. Siebenga JJ, Vennema H, Renckens B, de Bruin E, van der Veer B, Siezen RJ, et al. Epochal evolution of GGII.4 norovirus capsid proteins from 1995 to 2006. *Journal of virology*. 2007 Sep;81(18):9932-41. PubMed PMID: 17609280.
176. Cao S, Lou Z, Tan M, Chen Y, Liu Y, Zhang Z, et al. Structural basis for the recognition of blood group trisaccharides by norovirus. *Journal of virology*. 2007 Jun;81(11):5949-57. PubMed PMID: 17392366.
177. Zakikhany K, Allen DJ, Brown D, Iturriza-Gomara M. Molecular evolution of GII-4 Norovirus strains. *PloS one*. 2012;7(7):e41625. PubMed PMID: 22844506.
178. Shanker S, Choi JM, Sankaran B, Atmar RL, Estes MK, Prasad BV. Structural analysis of histo-blood group antigen binding specificity in a norovirus GII.4 epidemic variant: implications for epochal evolution. *Journal of virology*. 2011 Sep;85(17):8635-45. PubMed PMID: 21715503.
179. Debbink K, Donaldson EF, Lindesmith LC, Baric RS. Genetic mapping of a highly variable norovirus GII.4 blockade epitope: potential role in escape from human herd immunity. *Journal of virology*. 2012 Jan;86(2):1214-26. PubMed PMID: 22090110.
180. Koelle K, Cobey S, Grenfell B, Pascual M. Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in humans. *Science*. 2006 Dec 22;314(5807):1898-903. PubMed PMID: 17185596.
181. Taylor MB, Parker S, Grabow WO, Cubitt WD. An epidemiological investigation of Norwalk virus infection in South Africa. *Epidemiology and infection*. 1996 Apr;116(2):203-6. PubMed PMID: 8620912.
182. Smit TK, Bos P, Peenze I, Jiang X, Estes MK, Steele AD. Seroepidemiological study of genogroup I and II calicivirus infections in South and southern Africa. *Journal of medical virology*. 1999 Oct;59(2):227-31. PubMed PMID: 10459161.
183. Mans J, Armah GE, Steele AD, Taylor MB. Norovirus Epidemiology in Africa: A Review. *PloS one*. 2016;11(4):e0146280. PubMed PMID: 27116615.
184. Atmar RL, Opekun AR, Gilger MA, Estes MK, Crawford SE, Neill FH, et al. Norwalk virus shedding after experimental human infection. *Emerging infectious diseases*. 2008 Oct;14(10):1553-7. PubMed PMID: 18826818.

185. Murata T, Katsushima N, Mizuta K, Muraki Y, Hongo S, Matsuzaki Y. Prolonged norovirus shedding in infants ≤ 6 months of age with gastroenteritis. *The Pediatric infectious disease journal*. 2007 Jan;26(1):46-9. PubMed PMID: 17195705.
186. Moyo S, Hanevik K, Blomberg B, Kommedal O, Vainio K, Maselle S, et al. Genetic diversity of norovirus in hospitalised diarrhoeic children and asymptomatic controls in Dar es Salaam, Tanzania. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 2014 Aug;26:340-7. PubMed PMID: 24960396.
187. Kotloff KL, Blackwelder WC, Nasrin D, Nataro JP, Farag TH, van Eijk A, 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. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. 2012 Dec;55 Suppl 4:S232-45. PubMed PMID: 23169936.
188. Armah GE, Gallimore CI, Binka FN, Asmah RH, Green J, Ugoji U, et al. Characterisation of norovirus strains in rural Ghanaian children with acute diarrhoea. *Journal of medical virology*. 2006 Nov;78(11):1480-5. PubMed PMID: 16998875.
189. Mattison K, Sebunya TK, Shukla A, Noliwe LN, Bidawid S. Molecular detection and characterization of noroviruses from children in Botswana. *Journal of medical virology*. 2010 Feb;82(2):321-4. PubMed PMID: 20029818.
190. Ayukekbong J, Lindh M, Nenonen N, Tah F, Nkuo-Akenji T, Bergstrom T. Enteric viruses in healthy children in Cameroon: viral load and genotyping of norovirus strains. *Journal of medical virology*. 2011 Dec;83(12):2135-42. PubMed PMID: 22012721.
191. Allen DJ, Trainor E, Callaghan A, O'Brien SJ, Cunliffe NA, Iturriza-Gomara M. Early Detection of Epidemic GII-4 Norovirus Strains in UK and Malawi: Role of Surveillance of Sporadic Acute Gastroenteritis in Anticipating Global Epidemics. *PloS one*. 2016;11(4):e0146972. PubMed PMID: 27115152.
192. Farkas T, Nakajima S, Sugieda M, Deng X, Zhong W, Jiang X. Seroprevalence of noroviruses in swine. *Journal of clinical microbiology*. 2005 Feb;43(2):657-61. PubMed PMID: 15695660.
193. Gallimore CI, Lewis D, Taylor C, Cant A, Gennery A, Gray JJ. Chronic excretion of a norovirus in a child with cartilage hair hypoplasia (CHH). *Journal*

of clinical virology : the official publication of the Pan American Society for Clinical Virology. 2004 Jun;30(2):196-204. PubMed PMID: 15125877.

194. Lindesmith LC, Ferris MT, Mullan CW, Ferreira J, Debbink K, Swanstrom J, et al. Broad blockade antibody responses in human volunteers after immunization with a multivalent norovirus VLP candidate vaccine: immunological analyses from a phase I clinical trial. PLoS medicine. 2015 Mar;12(3):e1001807. PubMed PMID: 25803642.

195. Krammer F. The human antibody response to influenza A virus infection and vaccination. Nature reviews Immunology. 2019 Mar 5. PubMed PMID: 30837674.

Appendix 1: Publications arising from thesis data

Chapter 3.

Trainor E, Iturriza-Gomara M, Ngwira B, Cunliffe N. Detection of enterotoxigenic E. coli in hospitalised children with and without diarrhoea in Blantyre, Malawi. Paediatrics and international child health. 2016 Jan 29:1-4. PubMed PMID: 27077614.

Chapter 4.

Allen DJ, Trainor E, Callaghan A, O'Brien SJ, Cunliffe NA, Iturriza-Gomara M. Early Detection of Epidemic GII-4 Norovirus Strains in UK and Malawi: Role of Surveillance of Sporadic Acute Gastroenteritis in Anticipating Global Epidemics. PloS one. 2016;11(4):e0146972. PubMed PMID: 27115152. Pubmed Central PMCID: 4846118.

Appendix 2: Related publications

Trainor E, Lopman B, Iturriza-Gomara M, Dove W, Ngwira B, Nakagomi O, et al. Detection and molecular characterisation of noroviruses in hospitalised children in Malawi, 1997-2007. *Journal of medical virology*. 2013 Jul;85(7):1299-306. PubMed PMID: 23918547.