Introduction

The discovery of norovirus in 1972\(^1\), changed the understanding of the aetiology of gastroenteritis, making it the virus to be identified as an agent of gastroenteritis in humans.

Today, norovirus is recognised as one of the commonest human infections and estimated to be associated with 125 million cases and 35,000 deaths worldwide in 2010\(^2\). Better epidemiological surveillance and outbreak investigations\(^3\), coupled with wider implementation of molecular-based laboratory diagnostics\(^4\) are leading to better estimates of the burden of norovirus infections as well as improved outbreak control.

Data from challenge studies of prototype norovirus vaccines\(^5\)\(^6\) demonstrated that protection against infection and disease can be achieved, however there remain significant challenges to development of a norovirus vaccine. Recent advances in cell culture systems for norovirus\(^7\)\(^8\) and current research investigating the distribution of norovirus-associated disease in the population, for whom the disease burden is greatest, understanding host susceptibility factors, how to deploy novel technologies detecting norovirus in food and environmental matrices, and methodologies for ascertaining cases, are important in increasing our understanding of norovirus. Answers to these will help design strategies for vaccine and antiviral development, and how these might be best deployed to control norovirus infection.

Norovirus virology

At the time of discovery, the virus was referred to as the Norwalk agent, but as other related viruses were described in association with gastroenteritis, they became known as Norwalk-like viruses (NLVs), or – based on their morphology by electron microscopy – small, round, structured viruses (SLSVs). Following the cloning and sequencing of the Norwalk agent genome in 1993\(^9\), and subsequently other NLVs\(^10\), defining the genetic relatedness of these viruses led to their reclassification in the Caliciviridae family of viruses under the genus Norovirus in 2002\(^11\).

Classification of this genetically diverse group of viruses\(^12\)\(^13\) has described six established\(^14\) genogroups (GI-GVI), and a proposed seventh\(^15\)\(^16\). Two genogroups (GI and GII) are important pathogens of humans (GII also contains pathogens of animals, but there is no evidence of zoonotic transmission\(^17\)) and genogroups are further subdivided into genotypes: nine GI and twenty-two GII genotypes have been described\(^12\)\(^14\). Norovirus names are presented as genotypes, e.g. genogroup II-genotype 4 (GII.4) and strains are named for the place and year of their first description: e.g. GII.4/New Orleans 2009\(^18\) or GII.4/Sydney 2012\(^19\).

The norovirus genome is a single strand of positive-sense ribonucleic acid (+ssRNA) that is ~7700 nucleotides in length, organised as three open reading frames (ORF1-3)\(^20\). The 5' proximal ORF1 encodes a polyprotein that is post-translationally processed by the virus-encoded protease into six non-structural proteins, including a genome-linked protein (Vpg/NS5), protease (Pro/NS6) and an RNA-dependent RNA polymerase (RdRp/NS7). Both ORF2 and ORF3
encode a single protein each, VP1 and VP2 respectively, that are structural proteins involved in formation and stabilisation of the virus particle\textsuperscript{21}.

Norovirus particles are 27-35 nm in diameter and comprised of 180 copies of the VP1 protein, which itself is organised into three main domains: N-terminal (N), shell (S) and protruding (P), which is further arranged as P1 (subdivided as P1.1 and P1.2) and P2\textsuperscript{21}. In the mature infectious virus particle, the N domain is internal, and the P2 domain is the most external part of the virus, making it highly surface-exposed and placed to coordinate many of the interactions between norovirus and its host environment.

The primary host cell receptor for human norovirus is unknown, but can interact with histo-blood group antigens (HBGAs)\textsuperscript{22}; these are glycans expressed on epithelial cells and in mucosal secretions, which determine ABO blood type groups. Norovirus strains may use HBGAs as attachment factors or co-receptors\textsuperscript{23}, and sites in the VP1 P2 domain have been identified as HBGA binding sites\textsuperscript{24-29}. Synthesis of HBGAs occurs by sequential modification of a precursor, and the process is controlled by glycosyl-transferase enzymes from several genetic loci that exhibit polymorphism throughout the human population. The ABH and Lewis antigens are relevant to norovirus binding, and as such the phenotype of an individual for secretory ABH and Lewis antigens is a host susceptibility factor for norovirus infection. Specifically, individuals with a non-functional FUT2 gene, which encodes an α1,2-fucosyltransferase, have a ‘non-secretor’ phenotype, and are more resistant to norovirus infection that ‘secretor’ individuals\textsuperscript{30, 31}. Polymorphism at this locus may modulate susceptibility to other causes of diarrhoeal disease\textsuperscript{32}.

Public health laboratory surveillance worldwide has demonstrated dominance of GI\textsubscript{4} viruses\textsuperscript{33-36}, however other norovirus genotypes circulate consistently, if at a lower level, in the population. The GI\textsubscript{4} cluster of norovirus strains have been the most commonly detected noroviruses circulating worldwide since the mid-1990s, over time, distinct variants of the GI\textsubscript{4} virus evolve, emerge, and then recede to be replaced by a new variant\textsuperscript{33}. Emergence of a new GI\textsubscript{4} variant is associated with higher levels of infection and illness in the population and increased numbers of outbreaks\textsuperscript{37}, although severity of disease does not necessarily increase. These emergence events may be geographically contained (e.g. a 2003 variant emerged in Asia, and a 2006 variant spread in Europe)\textsuperscript{38}, or may be global, with new variants emerging and spreading worldwide over the course of a single year (as seen in 2002\textsuperscript{39}, 2006\textsuperscript{40}, 2009\textsuperscript{41} and 2012\textsuperscript{19}).

Noroviruses have been shown to have high evolutionary rates, up to $10^{-2}$ substitutions/site/year in the VP1 protein\textsuperscript{41}, due to the error-prone nature of the virus-encoded RdRp\textsuperscript{42}. The rate of evolution is fastest in the P2 domain, which interacts with the host immune system. Immune response to norovirus infection appears to target this region of the virus capsid, and epitopes in this domain have been identified as important in defining the antigenic profile of GI\textsubscript{4} norovirus strains\textsuperscript{43-47}, leading to the emergence of antigenically distinct viruses in the population, associated with epidemic/pandemic waves of gastroenteritis\textsuperscript{33, 37, 48, 49}. 
The emergence of variant GII.4 strains is associated with mutations occurring in the virus at epitope A (VP1 amino acid positions 296-298, 368 & 372) and D (VP1 amino acid positions 393-395)\textsuperscript{44,45}. Mutations, particularly in epitopes A and D, will be selected for in the virus population if the mutations are such that existing immunity in the host population is evaded by the mutated virus, but the virus is otherwise not disadvantaged. Because much of the human population is exposed to antigenically similar noroviruses at a similar time, virus-specific immunity is likely similar between many people. As a result, the variant norovirus is advantaged, being more likely to successfully evade existing immunity, and subsequently establish more infections and be transmitted. In this way, the virus can spread quickly through the population. Eventually infections generate new immunological responses, which ultimately limit the success of this variant in the population, but in turn creates an ecological niche favourable for a new variant, and the process cycles again. This process has been observed for GII.4 noroviruses throughout the 1990s, 2000s and 2010s\textsuperscript{39,33,37,45,47}.

**Norovirus surveillance**

Surveillance of norovirus is complicated because most people do not contact medical services when they are ill. In the UK, it is suggested that for each laboratory report of norovirus around 300 cases go unreported\textsuperscript{52}. This is largely related to the nature of the illness itself. The virus is highly infectious with an estimated infectious dose of around 10-100 virus particles (virions) needed to cause infection\textsuperscript{52}, with a high probability of infection from ingesting a single particle\textsuperscript{53}. It has a short incubation period, anywhere between 12 and 72 hours, and symptoms typically last for around 24-48 hours\textsuperscript{54}. Despite these difficulties it is still recognised as the commonest cause of gastrointestinal disease, not just in the UK, but worldwide\textsuperscript{55}. In the UK it is estimated between 3 and 4 million cases occur annually\textsuperscript{53,56}, at a cost of £106 million to patients and the health care services. In the USA this estimate is around 21 million domestically acquired cases\textsuperscript{57}. Infections with norovirus occur in all age groups, however, the highest incidence is in children aged less than five years\textsuperscript{56,58}.

The illness is often described as generally mild and self-limiting. The description of a mild infection can trivialise the effect of the illness; in England it has been estimated that 3000 admissions occur annually as a result of norovirus infection in adults\textsuperscript{59} or 0.3% of emergency admissions in those aged over 65, and 0.1% in adults aged 16-64 years. The consequences of infection are also greater in vulnerable populations. In a study in the county of Avon, UK, hospital patients were ill for longer than care home residents and staff working in the hospitals or care home, with around 10% of inpatients affected still showing symptoms 7 days after becoming ill\textsuperscript{60}. There is also evidence that norovirus can contribute to mortality in the elderly. Modelling of deaths suggests that norovirus is associated with 20% in those aged over 65 years who died of infectious intestinal disease, and that 13% of deaths caused by non-infectious intestinal disease\textsuperscript{61}.

Public Health England have conducted surveillance of gastrointestinal disease outbreaks since 1992\textsuperscript{62}. Analysis of the first nine years of data highlighted the importance of norovirus outbreaks in hospitals; over 80% of all reported
outbreaks in hospitals were suspected or confirmed as norovirus, and 25% of all general outbreaks occurred in hospitals\textsuperscript{63}.

Recent developments

Surveillance

Since the recognition of the importance of norovirus as a cause of GI disease a more detailed online surveillance system was set up in 2009\textsuperscript{63}. The online system increased ascertainment of outbreaks in hospitals, with more outbreaks reported in the first year than the whole of the preceding system\textsuperscript{63}. Both systems highlighted the increased activity during the winter months, and the considerable burden it places on NHS hospitals in England. The online system suggests around 13000 patients and 3000 staff are affected each year, moreover, almost 9000 bed days are lost because of restrictions to admissions during outbreaks\textsuperscript{63}.

The key to surveillance of norovirus is allaying the epidemiology with surveillance of virology. It is often difficult to achieve this. Recording the number of outbreaks, and laboratory reports indicates levels of infection, but they cannot directly ally this knowledge of circulating strains of the virus. The activity recorded in both Public Health England’s hospital outbreak reporting scheme and laboratory reports suggests that seasonal activity varies considerably. The reasons for changes in seasonal activity need unpicking and modelling of changes in the circulating strains of norovirus against laboratory reporting provided evidence that modifications within the virus itself leads to changes in the epidemiology. In the autumn/winter of 2012 PHE recorded increased levels of norovirus activity; later attributed to the emergence of the Sydney 2012 strain\textsuperscript{37}. However, other reasons have been proposed, such as changes in winter conditions such as falling temperature\textsuperscript{64}.

Given the difficulty in surveillance of norovirus infections from direct sources, other developments need to be explored. For example, social media could provide early indications of increasing activity. There are a number of publications looking at the use of internet search and social media postings to provide information on increased disease activity\textsuperscript{65-67}. Other forms of syndromic surveillance have been used such as the use of telephone helpline data to map diarrhoea and vomiting\textsuperscript{68}, difficulties with this approach fall mainly on disentangling the causes of the illness from syndromes (diarrhoea and vomiting). Norovirus is not the only cause of D&V and has a seasonality similar to that of rotavirus, similarly sapovirus has similar illness characteristics to norovirus.

Virus culture systems

Understanding the interactions of norovirus with host cells has been limited by the lack of an in vitro laboratory cell culture system. Attempts to establish conventional cell culture approaches were unsuccessful\textsuperscript{69}, after which alternative approaches were developed\textsuperscript{70, 71}, however, these were limited in their usefulness.
More recently, progress has been made towards development of laboratory culture systems for human norovirus. Two systems have been described: one describes human norovirus replication in B cells⁸, and a second which describes human norovirus replication supported by stem cell-derived human enteroids⁹. These systems present exciting new opportunities to understand how norovirus interacts both with the host cell and with the host environment.

The system using human enteroids⁹ provides a model for processes of norovirus replication such as attachment/entry, genome replication, and virus assembly/release can be interrogated in a biologically relevant cell type. Advances in these areas will be crucial for identifying targets for virus-specific interventions, and evaluating how effective different antiviral therapies can limit norovirus replication. Further insights into virus entry and egress will enhance understanding of the interactions between virus and host receptors and identify novel interactions between virus and host that serve as intervention targets, for example antibodies which interfere with attachment or release processes, thus neutralising free virus.

The second system, in which norovirus replication is supported in B cells, uses commensal bacteria that express HBGAs to facilitate virus replication in this model⁸,⁷². Analysis of norovirus replication in this system could enhance understanding of the interaction between norovirus and HBGAs – and identify how these interactions might be disrupted— but also what interactions might occur with, and what role might be played by, the microbiome during norovirus infection⁷³.

Norovirus vaccines

Modelling studies have shown that norovirus vaccination would offer healthcare and economic benefit⁷⁴. These could help control and prevent the large-scale and often protracted outbreaks often seen in healthcare settings⁷⁵ and other settings such as in the military⁷⁶,⁷⁷.

Until very recently, development of candidate vaccines focussed on recombinant protein systems; expression of the norovirus capsid protein VP1 in vitro leads to self-assembly of the protein into virus-like particles (VLPs) that are antigenically and morphologically identical to infectious virus, but lacking a genome, VLPs are entirely non-infectious⁷⁸.

Early clinical studies of responses in humans to immunisation with VLPs demonstrated they were immunogenic when delivered orally⁷⁹,⁸⁰ or intranasally⁸¹. A randomised, double-blind placebo-controlled trial conducted in healthy, susceptible adult volunteers investigated the safety and efficacy of vaccination using norovirus VLPs, followed by challenge with a homologous norovirus strain⁶. This trial demonstrated 70% of vaccine recipients had a virus-specific IgA response, and vaccination reduced the frequency of both infection and disease between placebo control group and vaccine recipients⁶.
However, the prototype vaccine (and challenge strain) used in this trial was based on a single norovirus strain – the prototype GI.1 Norwalk virus/1968 – which is uncommon, detected in <1% of norovirus strains characterised per year in surveillance programmes in developed countries. As the most significant disease burden is associated with the GI.4 genocluster, any candidate vaccine would need to elicit immunity to GI.4 norovirus strains, and cross-react to antigenically distinct GI.4 variants. A chimeric VLP was developed incorporating epitopes from antigenically distinct GI.4 viruses\textsuperscript{82}, and induced broadly-reactive antibody responses\textsuperscript{83}.

A subsequent trial incorporated the chimeric VLP into a GI.1/GI.4 bivalent vaccine formulation and demonstrated vaccine induced seroconversion in 90% of vaccine recipients, and reduced gastroenteritis following challenge\textsuperscript{5}. However, the predefined primary endpoints were not achieved in this study, and further studies are necessary to assess how effective this candidate vaccine would be in the general population, and specifically in paediatric and elderly populations. Furthermore, studies must address both the duration of and the inter-/intra-genotype breadth of protection.

**Perspectives**

Clearly, significant progress has been made in understanding the virology and epidemiology of norovirus in humans: but there remain significant gaps in our knowledge, important for development of therapeutic and preventative interventions, and ascertaining norovirus disease burden to understand how these should be utilised, and to measure their effectiveness. This is true in all economic settings, but especially in low economic settings.

One key question is to understand the emergence of norovirus strains. With no animal reservoir\textsuperscript{17}, the virus must be sustained – and continuously evolve – in the human population. Using genomics approaches to measure and monitor virus diversity among circulating strains, and to characterise and measure whether observed genetic changes induce phenotypic changes will be crucial in developing the systems needed to understand and monitor emergence events, particularly those that lead to rapid pandemic spread of norovirus strains. There is increasing evidence that children may act as important reservoirs of norovirus, and the virus may exploit the more naïve immunological background in children to explore antigenic diversity, ultimately leading to virus diversification and subsequent emergence of novel strains\textsuperscript{84}.

Second, more detailed understanding of the burden of the disease, transmission dynamics and pathogenesis in risk groups, both those at risk of more severe disease (immunocompromised\textsuperscript{85}, elderly\textsuperscript{61}), and those more likely to come into contact with or are at higher risk of transmitting the virus (food-handlers\textsuperscript{86}, healthcare workers\textsuperscript{87}, military personnel\textsuperscript{75}) is needed. There are complex epidemiological and virological questions relating to the distribution of norovirus-associated disease in the population, for whom the disease burden is greatest, as well as understanding host susceptibility factors. Integrated laboratory and epidemiological studies are crucial to investigate how norovirus is transmitted, disease attribution via different transmission pathways, how infections can be tracked in the
population and during outbreaks, and what role susceptibility factors such as HBGA phenotype or the individual microbiome composition may play in norovirus infection, development of disease and outcomes.

Third, alongside data on the direct burden of disease, enhanced data are needed to understand where interventions may alleviate transmission and disease overall, as many settings are interlinked. For example, administering a norovirus vaccine to patients in long term care homes might help prevent outbreaks in this environment, but might have limited effects on the population as a whole. However, it may be a worthwhile strategy if vaccination in care homes subsequently prevents outbreaks in hospitals and reduces bed blocking.

With the recent advances in laboratory culture systems for norovirus\textsuperscript{7,88}, next generation sequencing technologies\textsuperscript{89}, improved diagnostics\textsuperscript{4} and measuring phenotypic characteristics of noroviruses\textsuperscript{90}, there are new opportunities to advance understanding of this common and important human pathogen.

Acknowledgments

The research was funded by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Gastrointestinal Infections at University of Liverpool in partnership with Public Health England (PHE), in collaboration with University of East Anglia, University of Oxford and the Institute of Food Research. DJA is based at the London School of Hygiene & Tropical Medicine. JPH is based at University of Liverpool. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health or Public Health England.
References


