Infectious intestinal disease (IID) in schools and school-aged children

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor of Philosophy by

Anna Louise Donaldson

January 2023

Department of Public Health and Policy

University of Liverpool

This thesis is dedicated to Tom, Joel and Zoe.

Acknowledgments

This research was funded by the National Institute for Health Research (NIHR) Health Protection Research Unit (HPRU) in Gastrointestinal Infections at the University of Liverpool in partnership with Public Health England (PHE), and in collaboration with the University of East Anglia, University of Oxford and the Institute of Food Research. I would like to thank the NIHR HPRU for supporting me in undertaking this work.

I would like to thank my supervisors, Professor Sarah O'Brien, Dr Roberto Vivancos and Dr John Harris, for their encouragement, guidance, and expertise throughout this research. I would also like to thank the other co-authors of publications within this thesis; Jo Hardstaff, Dan Hungerford, Ian Hall and Helen Clough, for their input and advice.

I am grateful for the support of PHE colleagues in accessing and extracting the data which underpin many of the analyses in this thesis. I am particularly thankful to Kathy Chandler from the Field Epidemiology Service for her assistance, patience, and invaluable knowledge of these datasets. I would like to thank colleagues from the Real Time Syndromic Surveillance Team for their time and effort in providing telehealth and general practice data to inform this work, in particular the support of Alex Elliot, Helen Hughes, Sally Harcourt and Paul Loveridge. I would also like to thank colleagues from the National GI team for the provision of HNORS data.

I am very grateful to the staff at St Helens, Liverpool and Sefton Local Authorities who supported the school absenteeism study. In particular, I would like to thank Steve Knuckey and Mike Roberts for their efforts in extracting and providing school data. Finally, I would like to thank Becky Glennon-Alty, for her help and knowledge in engaging with local schools in Merseyside.

Abstract

Background

Infectious intestinal disease (IID) is common, and children are more likely than adults both to suffer from IID and to transmit infection onto others. IID surveillance is primarily based on healthcare data and as such illness managed within the community will often go undetected. In this thesis, I explore the burden of IID in children and schools, and the potential utility of school attendance registers in enhancing IID surveillance.

Methods

I conducted an analysis of reported school outbreaks to identify factors associated with outbreak occurrence and attack rate. I then undertook a systematic review to explore the role of school-based surveillance in detecting infectious illness amongst children. This led to a retrospective analysis of school illness absence data to assess its association with IID surveillance indicators. I then used national surveillance data to identify whether cases and outbreaks of IID in children provided an early warning of seasonal norovirus infection. Finally, I explored whether symptoms alone could distinguish between different causes of IID and whether these symptom profiles altered across age groups.

Results

IID accounted for nearly half of all reported school outbreaks, with primary and all-through schools, and larger school size associated with an increased risk of outbreaks. In the systematic review, influenza-specific absences were found to provide a lead time of up to two weeks ahead of existing surveillance indicators, but no studies were identified which considered the role of school attendance registers in IID surveillance. An analysis comparing school illness absence data with IID surveillance indicators revealed a statistically significant positive association with viral IID laboratory reports, general practice consultations and telehealth calls for IID in children. Using national surveillance data, school IID outbreaks were found to provide a 3-week lead time ahead of care home and hospital outbreaks, and children provided a lead time ahead of adults for norovirus laboratory reports, and NHS 111 calls for both vomiting and diarrhoea. Symptom profiles were identified which distinguished bacterial from viral IID and these profiles were found to differ across age groups.

Conclusion

These findings highlight the importance of IID as a cause of school outbreaks and the potential for school outbreak data to provide an early warning of seasonal norovirus infection. Collecting syndrome-specific absence data could improve the utility of school data for surveillance purposes and symptom profiles, which differentiate viral and bacterial pathogens, could be used to enhance data collection from schools and existing syndromic surveillance systems.

Contents

List of Figures

List of Tables

List of Boxes

Abbreviations

A&E	Accident and Emergency
AGE	Acute gastroenteritis
AIC	Akaike Information Criterion
AR	Attack rate
CCG	Clinical Commissioning Group
CI	Confidence interval
CPC	Chief presenting complaints
ED	Emergency Department
GI	Gastrointestinal
GP	General Practice
HNORS	Hospital Norovirus Outbreak Reporting System
IID	Infectious Intestinal Disease
ILI	Influenza-like-illness
IQR	Interquartile range
LSOA	Lower Super Output Area
NHS	National Health Service
OECD	The Organisation for Economic Co-operation and Development
Ofsted	Office for Standards in Education, Children's Services and Skills
ONS	Office for National Statistics
OR	Odds ratio
PHE	Public Health England*
PII	Personally Identifiable Information
PPI	Public and Patient Involvement
ReSST	Real-time Syndromic Surveillance Team
RR	Relative risk
SD	Standard deviation
SIMS	School Information Management System
STEM	Science, Technology, Engineering and Mathematics
UK	United Kingdom
URN	Unique Reference Number
WHO	World Health Organization

* Public Health England (PHE) was disbanded on the 1st October 2021, prior to the submission of this thesis, and its health protection functions were transferred to the newly created UK Health Security Agency (UKHSA). The research which informs this thesis was undertaken, written and published under Public Health England and therefore Public Health England is referred to throughout. To avoid confusion and ensure consistency, Public Health England is also referred to within the introduction and discussion sections. The functions of Public Health England which are described are still in existence within the UKHSA and the recommendations are equally as applicable to UKHSA and they were to PHE.

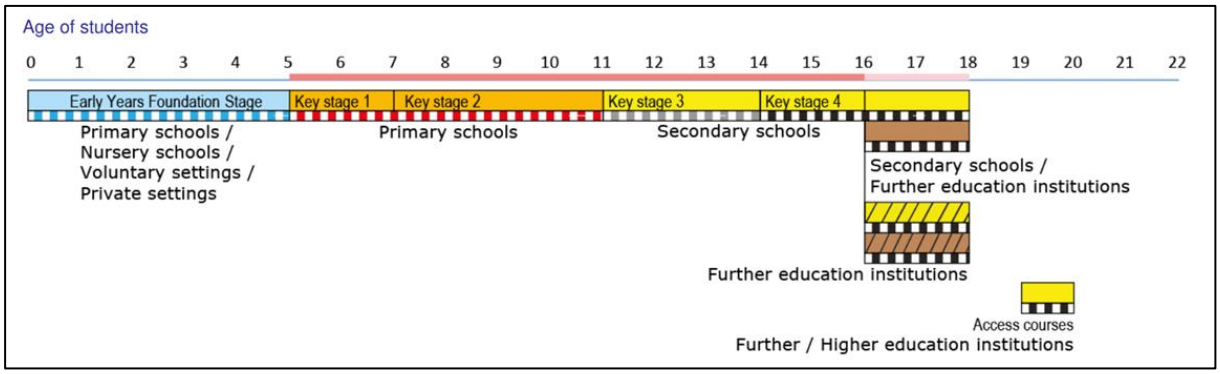
Chapter 1. Introduction

In this introductory chapter, I will provide the necessary background to the research presented in this thesis. I will start by outlining the structure of the education system in England, the regulation of schools and their role in supporting the health of pupils. I will then discuss infectious intestinal disease, its aetiology, incidence and impact on both individuals and wider society. Routes of disease transmission will be explored, including the use of social network analysis to understand transmission pathways between people and places. This will be used to make the case for the role of children in the wider community transmission of IID, with areas for further research identified. I will then consider infection prevention and control in schools, outlining factors that increase the risk of outbreaks in schools and highlighting gaps within the knowledge base. I will also explore the evidence base for interventions to reduce transmission and control outbreaks. Finally, surveillance of IID will be discussed in terms of the national surveillance systems currently in place and the ability of these systems to detect cases in children and outbreaks in schools. I will use evidence from the literature to explore alternative health surveillance in schools, identifying key evidence gaps in relation to IID. I will then outline the aims and objectives of this thesis and how the subsequent chapters will address these research questions.

1.1 Education system in England, UK

There are approximately 24 000 registered schools in England, supporting a pupil population of 8.89 million.[1] Education is divided into five stages which takes children from birth through to 18 years of age (Figure 1.1).[2] The first phase of education, Early Years Foundation, is optional and is for children aged 0-4 years. It is largely provided by private nursery schools and preschools which may or may not be linked to a primary school. All children aged 3-4 years are entitled to free early years education, but full-time compulsory education does not start until children are 5 years old.[3] The next two phases of education, Key Stage 1 and Key Stage 2, are for children aged 5-10 years and are taught by primary schools. Secondary schools teach Key Stages 3 and 4 to children aged 11-16 years. Some schools may cover both primary and secondary education (all-through schools). Compulsory education ends at age 16, but students must stay in some form of education or apprenticeship until they are 18 years old.[4]

Figure 1.1: Structure of the education system in England, UK



Reproduced under a CC-BY licence from The Structure of the European Education Systems 2018/19 [2]

The majority of schools in England are state-funded and in addition to mainstream primary and secondary schools there are special schools for children with special educational needs, and pupil referral units for children who are unable to attend a mainstream school.[1] Across the country there are also independent schools and non-maintained special schools which are registered but do not receive government funding and there is often a fee for pupils to attend.[1] These make up less than 10% of schools, with state-funded primary schools accounting for the largest proportion of all schools in England (69%).

Schools vary significantly in size, with state primary schools having an average of 281 pupils per school compared with an average size of 986 pupils for state secondary schools.[1] Special schools and pupil referral units tend to be the smallest schools, with an average of 129 and 44 pupils respectively. Regardless of the overall size of the school, class sizes are limited to approximately 30 pupils, with larger schools having more classes per year group. Approximately a third of pupils in primary and secondary schools are in minority ethnic groups and around 1 in 6 are eligible for free school meals, which are offered to families on low incomes who are in receipt of state benefits.[1]

State schools receive funding from their Local Government or directly from central government. There are different types of state school, which vary in terms of their governance and curriculum.[5] Community schools are run by the Local Government and follow the National Curriculum, which outlines programmes of study and attainment targets for the school.[6] Foundation schools and voluntary schools are run by an elected governing body and have more flexibility to change how they run the school, but still follow the National Curriculum. This includes many faith schools. Academies and free schools are funded by central government but run by independent trusts. They can decide how they run the school as well as what curriculum to follow.

All state schools are inspected by the Office for Standards in Education, Children's Services and Skills (Ofsted) and must follow the same rules on admissions, special educational needs, and exclusions, including sitting the same exams. Ofsted are responsible for inspecting, regulating, and reporting on the quality and effectiveness of education provided by state schools as well as some independent schools which may register to be inspected by Ofsted. Those that do not register with Ofsted are inspected by the Independent Schools Inspectorate (ISI), which itself is monitored by Ofsted.[7]

Whilst schools are first and foremost an educational setting, they also have health and safety responsibilities towards their pupils. This includes ensuring pupils are safe within the school environment and when undertaking school activities.[8] State schools in the UK support the delivery of national public health programmes, such as the National Child Measurement Programme and vaccination programmes.[9–11] In addition to these, all children are offered a hearing and eye sight screen in their first year of school.[12,13] The links between health and education are well recognised, with poorer health in childhood associated with lower educational attainment.[14–16] Poorer health status can also result in increased absence from school, which in turn affects educational outcomes.[15] Lower educational attainment is linked to poorer health in adult life, with increased rates of morbidity and mortality.[17–19] Education and health interact to affect employment opportunities, with poorer health and lower education associated with a higher risk of unemployment.[20] Unemployment itself is also associated with poorer health outcomes.[21]

This close interaction between education and health can create both positive and negative cycles, with good health in childhood leading to better education and improved long term health outcomes. Conversely, poorer health in childhood can lead to lower educational attainment, increased risk of unemployment and poorer health in adulthood. This highlights the importance of optimising childhood health, attendance, and attainment at school to improve long term health, social and economic outcomes.

1.2 Infectious intestinal disease

Definition and aetiology

Infectious intestinal disease (IID) is characterised by the acute onset of diarrhoea and/or vomiting in otherwise healthy people caused by an infectious, transmissible organism.[22] Whilst the majority of cases are mild and self-limiting, IID can be accompanied by systemic symptoms, such as fever and dehydration.[23] IID should be distinguished from both gastroenteritis and gastro-intestinal infection, and can be defined as the occurrence of gastro-intestinal symptoms in the presence of an infectious

organism.[22] In contrast, gastroenteritis has many non-infectious causes, such as inflammatory bowel disease and food intolerances. Likewise, gastro-intestinal infection does not always give rise to gastro-intestinal symptoms, as is the case with *Helicobacter pylori* infection.[24] IID can be caused by bacteria, viruses and protozoa. Whilst there are a wide range of pathogens that cause IID, the most common organisms within the UK are shown in Box 1.1.

Box 1.1: Common causes of infectious intestinal disease (IID) in the UK

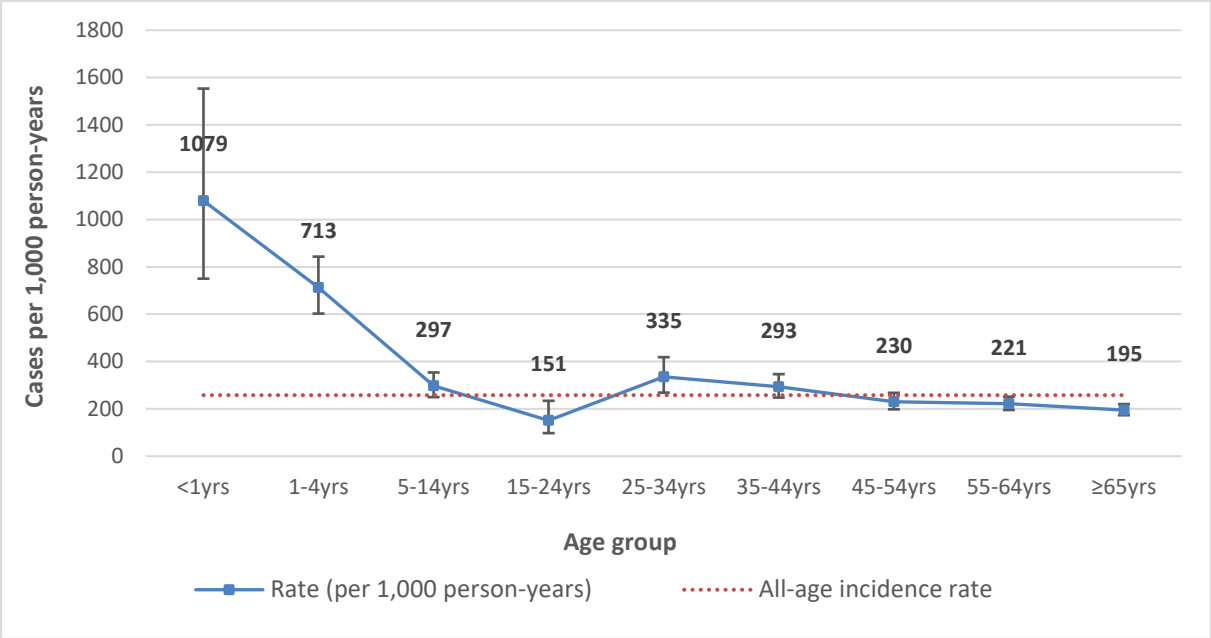
Bacteria	Protozoa	Viruses
▪ *Clostridium difficile* ▪ *Clostridium perfringens* ▪ *Campylobacter* spp. ▪ *Escherichia coli* O157 VTEC ▪ *Escherichia coli* non-O157 VTEC ▪ Enteroaggregative *E.coli* ▪ *Listeria* spp. ▪ *Salmonella* spp. ▪ *Shigella* spp.	▪ Cryptosporidium ▪ Giardia	▪ Adenovirus types 40 and 41 ▪ Astrovirus ▪ Norovirus ▪ Rotavirus ▪ Sapovirus

Adapted under the Open Government Licence from The Second Study of Infectious Intestinal Disease in the Community [22]

Incidence rate and seasonality

In the UK, it has been estimated that 1 in 4 people suffer from an episode of IID each year (258 cases per 1 000 person-years).[22] Young children and infants have considerably higher rates of IID compared to older children and adults (Figure 1.2). The highest rate of IID is seen in children less than 1 year of age (1 079 cases per 1 000 person-years), with almost all infants having at least one episode of IID in their first 12 months of life. For school-aged children (5-14 years) the incidence rate is just under 300 cases per 1 000 person-years, meaning almost 1 in 3 children will experience diarrhoea and vomiting in each academic year. Older teenagers and young adults (15-24 years) have an incidence rate which is lower than the all-age incidence rate; a difference which is statistically significant.

Figure 1.2: Incidence rate of IID (with 95% confidence intervals), by age, from the IID2 Cohort Study



Adapted under the Open Government Licence from The Second Study of Infectious Intestinal Disease in the Community [22]

Norovirus is the single most common cause of IID in high-income countries, accounting for approximately 11-16% of community cases.[25–28] In the UK, it affects nearly 5% of the population every year.[29] Norovirus infection occurs all year round but is more common during the winter months (December to February for the Northern Hemisphere).[30] Prior to the vaccine, rotavirus was the most common cause of IID in young children, affecting nearly all children globally by 5 years of age.[31,32] In the UK, it was responsible for half of hospital admissions and approximately one quarter of GP consultations and telehealth calls for gastroenteritis in children under 5 years each year.[33] Like norovirus, it is common in the winter months and peaked between winter and early spring each year.[33] However, vaccination has altered the seasonal pattern of rotavirus, changing the size, timing and duration of seasonal peaks.[34] *Campylobacter* is one of the most common bacterial causes of IID globally [35] resulting in 0.5 million cases each year in the UK.[29] *Campylobacter* has a well-defined seasonal trend, with a peak incidence between May and June each year in the UK.[36]

Individual and societal impact

Norovirus infection can cause vomiting, watery diarrhoea, abdominal cramps and fever, with symptoms typically lasting two to three days.[37] Whilst norovirus is typically a mild, self-limiting infection, the severity of disease and duration of symptoms can be affected by age, co-morbidity and hospitalisation.[38–40] Norovirus is highly transmissible due to the low infectious dose and high levels

of viral shedding.[41] Consequently, it spreads easily in semi-enclosed environments, resulting in outbreaks in a multitude of settings, including schools, nursing homes and hospitals.[42,43] Each year norovirus causes widespread disruption to healthcare services and it has been estimated to cost the global economy $4.2 billion in healthcare costs and $60.3 billion in societal costs each year.[44] In the UK, norovirus is estimated to cause between 6 000 and 18 000 hospital admissions, 30 000 A&E attendances, 160 000 GP consultations and 56 000 calls to telehealth services each year.[45] The total cost to the healthcare service and patients is estimated at £81 million per annum, with a cost per case of approximately £30.[45]

Before the introduction of the vaccine, rotavirus was the most common cause of severe diarrhoea in young children globally,[46] causing sudden onset diarrhoea and vomiting accompanied by fever and abdominal pain.[31] Symptoms could last for three to seven days and young children were particularly susceptible to dehydration which could be severe and life threatening.[31] However, the licensing of the rotavirus vaccine in 2006 has significantly diminished the incidence and impact of rotavirus infection in young children, resulting in a 42% median reduction in acute gastroenteritis mortality, a 38% median reduction in acute gastroenteritis hospitalisations, and a 67% median reduction in rotavirus hospitalisations and emergency department visits in children under five globally.[47] Rotavirus vaccination has also led to a reduction in acute gastroenteritis hospitalisations in older unvaccinated age groups.[48] However, despite the reduction in severe cases, rotavirus activity has continued and infection is now more common in unvaccinated older children and the elderly.[49]

Campylobacter can cause watery or bloody diarrhoea, fever and abdominal pain, with symptoms lasting approximately six days.[50] Infection has also been linked to subsequent Guillain-Barré syndrome, which is thought to affect 1 in every 5 000 cases.[51] In the UK, *Campylobacter* infection has been estimated to result in approximately 11 000 telehealth calls, 100 000 GP consultations, nearly 6 000 A&E attendances, and between 850 and 2 700 hospital admissions every year, costing the health service an estimated £4 to £5 million per annum.[45] Costs to patients include those from medication, transport to healthcare appointments, childcare, and lost income and has been estimated at £45 million per annum.[45]

For children, infectious intestinal disease is a major cause of mortality and morbidity worldwide.[52] The most frequent and serious complication is dehydration, which occurs more commonly in children under 5 years and in those with poor nutrition.[53,54] In high-income countries mortality from IID is rare but it is still a major cause of healthcare usage and hospitalisations, especially in young children and infants.[55] For school-aged children, illness is generally milder and of short duration but an

estimated 5% will still require hospitalisation, particularly due to *E.coli*, *Shigella* and *Listeria*.[56] In addition to the health consequences of infection, there is an economic burden associated with childhood infection which results from healthcare appointments, medication and parental loss of income associated with caring for an unwell child at home.[56] Due to the high incidence of IID in childhood, these impacts are magnified to create a significant burden on society in terms of productivity losses and missed days at school and work each year.[57] Frequent school absenteeism has been associated with poorer educational outcomes,[58] and there is evidence that every day missed from school has the potential to adversely affect performance, especially in those already at risk of underperforming.[59] Therefore, the management and control of IID could have important implications on longer term educational outcomes, especially for the most vulnerable children.

1.3 The role of children and schools in disease transmission

Transmission pathways

Infectious intestinal disease is primarily spread by faecal-oral transmission but there is evidence that aerosolised viral particles from vomiting can also transmit pathogens.[60] The most common pathway of infection is direct person-to-person contact, but transmission can also occur via food, water, animal contact and environmental contamination.[61,62] The predominant transmission pathway varies by organism, with viral IID primarily transmitted through person-to-person contact, although foodborne transmission can also occur.[62] People infected with viral IID, such as norovirus and rotavirus, shed large numbers of viral particles which can remain infective on surfaces for days to weeks, depending on the surface material and temperature.[63–65] Both norovirus and rotavirus have very low infective doses, with as few as ten viral particles sufficient to cause infection.[65] When contaminated surfaces are touched, infection can be transmitted by subsequent hand to mouth contact. Whilst rotavirus is primarily spread through faecal-oral transmission,[34] norovirus can be widely dispersed by vomit and can transmit to others via inhalation, contamination of surfaces or direct contamination of hands.[60,65] Norovirus can also be readily transmitted through contaminated food and is a leading cause of foodborne outbreaks.[64] The most common source of contaminated food is via an infected food handler, but some foods may be contaminated at source by irrigation water polluted with human faeces.[64] Shellfish are most commonly implicated in direct foodborne transmission,[66] as some molluscs can accumulate norovirus in their bodies if they are grown in water contaminated with human faeces.[67]

Animals are the major reservoir of many bacterial causes of IID, which can be transmitted to humans from undercooked meats, inadequate storage of cooked meats, or cross-contamination of raw food

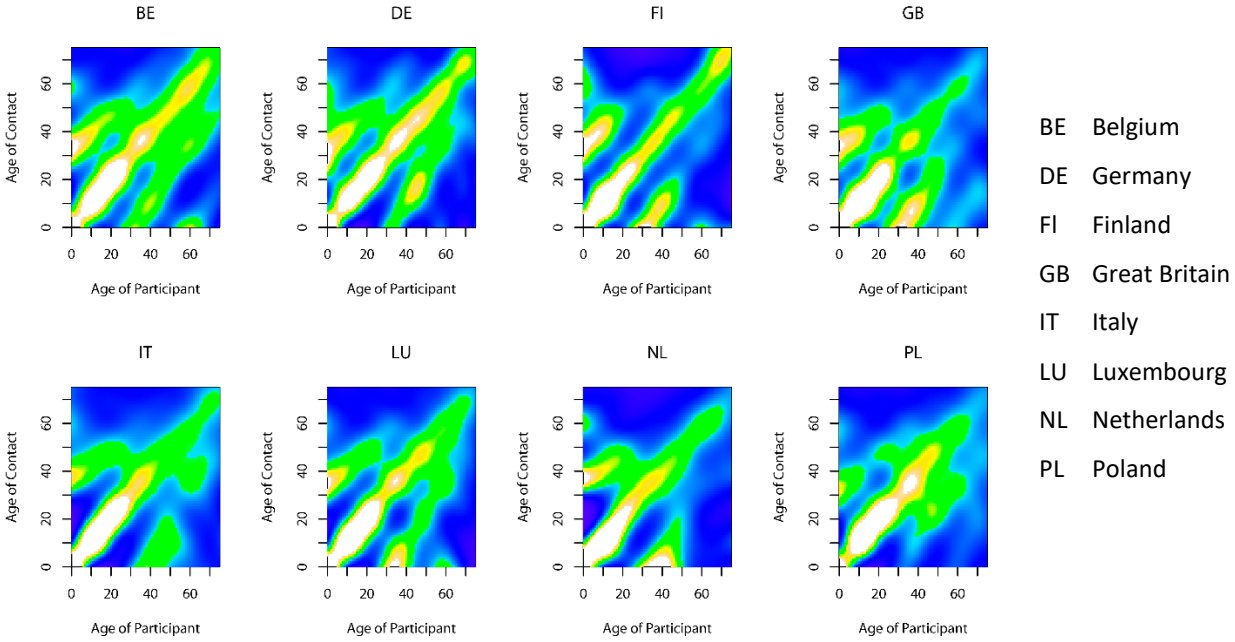
products.[53,68] *Campylobacter* can be found in the intestine of poultry, cattle, birds and domestic pets,[68] with poultry and dairy being the primary sources of *Campylobacter* food poisoning.[50,69,70] *Campylobacter* can also be transmitted by contaminated water, direct contact with animals and, only rarely, person-to-person.[69,70] *Salmonella* is also found in poultry and infection is linked to the consumption of undercooked meat and contaminated egg and dairy products. *Salmonella* can be more readily transmitted person-to-person and directly from animals, with exotic pets identified as a source of infection.[68] *E.coli* O157 is found in cattle but, unlike *Campylobacter* and *Salmonella*, it has a much lower infective dose and can spread more rapidly person-to-person.[68] Bacteria with human reservoirs, such as non-O157 *E.coli* and *Shigella*, are predominately spread by person-to-person contact or by food contaminated with human faeces.[68] Unlike non-O157 *E.coli* strains, *Shigella* has a very low infective dose, with as few as 10 organisms required to cause infection.[71] It can spread easily via the hands of children and consequently can readily cause outbreaks, especially in childcare settings.[68]

Social networks as pathways for disease transmission

For a disease to spread person-to-person through a community, contact is needed between an infectious person and other susceptible people. The level of contact required to transmit an infection is dependent on the organism, its infectivity and mode of transmission.[72] Social networks are formed from a variety of different contacts, including close physical contact, conversation with or without physical contact, and sexual contact.[73] Importantly, contacts within a social network are considered relatively stable contacts which are likely to recur over time. These are distinguished from transient, random contacts which may or may not occur again in the future.[73] Contacts may occur in different environments such as the workplace, school, community settings and at home. Understanding social networks and the probability of disease transmission associated with different levels of contact can help predict and control the spread of infection.

Social networks within communities are highly assortative with age, meaning that most people spend time with people a similar age to themselves.[74,75] There are also high contact rates between people with an approximate 30 year age difference, which represents transgenerational mixing between children, parents and grandparents (Figure 1.3).

Figure 1.3: Smoothed contact matrices, by country, for physical contacts



Reproduced under a CC-BY licence from Mossong et al (2008) [75]

The most contacts are reported in schools and at home, with fewer contacts in workplaces, public places and transport.[74] The nature of contacts also vary between settings. Contacts made at home, school and in leisure time are more likely to be physical contacts which occur regularly.[75,76] In contrast workplace contacts tend to be conversational contacts which occur more irregularly.[76]

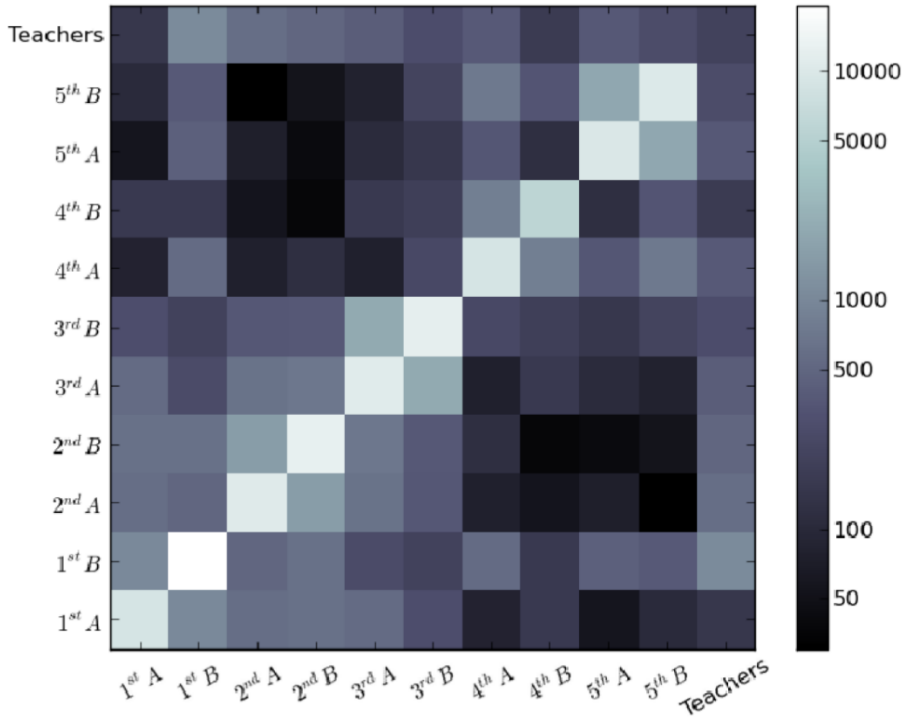
Even within school environments, contact networks are not homogenous and the mixing of children in school is also highly assortative by age. In both primary and secondary schools, the greatest number of contacts occur between children within the same class,[77,78] accounting for as many as 92% of all school contacts.[78] Contacts outside of a given class are more common with others in the same year group and less likely to occur between year groups (Figure 1.4).[77,79] The majority of contacts made in a day are brief, lasting less than 5 minutes, however contact duration is highly heterogeneous with considerably longer times also observed.[77,78,80] Within a class, there is evidence of social clustering, with children forming close links with a smaller sub-group of their peers.[79] Consequently, students will not necessarily have close contact with all of their classmates (defined as being face-to-face within a 1.5m proximity).[78]

Within schools, there are a high number of random contacts made when children pass each other in the corridors or outside during breaks.[81] These contacts are numerous but of short duration and are

estimated to account for less than 3% of the total contact time of a school-child per day.[81] Whilst some studies have shown contacts peak in-between classes and during lunch breaks,[78,80] others have shown a reduction in contacts during the lunch break, driven by students leaving the school grounds for lunch.[77] Consequently, the number of random contacts and peak times of contacts are likely to vary between schools, depending on the structure and timetabling of the school day.

Amongst younger children, the number of close contacts involving teachers appears to be relatively small (Figure 1.4), although in part this is due to the smaller number of teachers compared to students, meaning they contribute less to a given individual's contact network.[77] Amongst older children, there is evidence of role assortativity between students and teachers, which was more pronounced as the duration of the contact increased.[80] This indicates that longer, high risk contacts are more likely between peers, than between students and teachers.

Figure 1.4: The number of contacts made between individuals of two classes (A and B), in year groups 1-5, shown on a logarithmic grayscale



Reproduced under a CC-BY licence from Stehlé et al (2011) [77]

The structure of contact networks within schools appear to be relatively stable from day-to-day and across years,[78] with the number of contacts reducing dramatically at weekends and during school holidays.[74] Whilst the number of contact-hours with classmates remains relatively stable across year groups, the contribution of friendship groups and sports clubs to children's social networks

significantly increase as children get older.[81] Furthermore, students who are involved in a large number of different groups and activities have the potential to act as "super-spreaders", enabling infectious diseases to transmit between different subgroups of children.[81]

Children as community transmitters

Whilst schools account for the highest number of contacts made by children, contact-hours remain greatest at home.[81] For diseases which spread through close contact, the key groups for transmission are therefore households and school classes, with friendship groups and sports clubs of increasing importance for older children.[81] Contacts outside of school are primarily with adults of the parents' age group, who are estimated to account for 40% of all private contacts.[82] Consequently, diseases which affect children can readily spread into adult age groups and out into the wider community.

However, not all diseases affect children and adults equally and the importance of children as wider transmitters of infection is dependent on the organism and its mode of transmission. Children are thought to play a major role in the spread of influenza and studies have shown an increased risk of influenza infection in households with children, compared to adult households.[83–86] However, identifying and quantifying the impact of schools and children in community outbreaks is complex and routine data may be insufficient to characterise transmission networks. Studies during the 2009 H1N1 influenza pandemic have provided evidence of outbreaks which started in schools and subsequently spread into households and the community,[87–89] with reconstructed transmission networks illustrating the back-and-forth transmission which occurs between these settings.[87,88] In one study, young school-aged children (6-10 year olds) were 84% more likely to introduce infection into the household compared to over 18 year olds.[87]

For influenza, the role of children in community transmission has formed part of the rationale for childhood influenza vaccination programmes.[90,91] As well as having a direct protective effect for children, vaccinating children against influenza has been shown to provide indirect protection to unvaccinated age groups.[92,93] This indirect protection has also been observed for other vaccine-preventable diseases, such as rubella.[94] However, despite the recognised role of children in the spread of respiratory infections such as influenza, the COVID-19 pandemic has demonstrated a far smaller role for children in disease transmission than would have otherwise been expected from a respiratory virus in an immune naïve population.[95,96]

The role of children in the transmission of infectious intestinal diseases is also likely to vary depending on the organism, its mode of transmission and its impact on different age groups. Furthermore, for a given organism, strain type may also affect the relative impact on different age groups.[97,98] Norovirus is known to affect all age groups, but is particularly problematic in healthcare settings where outbreaks are common and the vulnerability of patients can result in more severe infection.[99] However, the incidence is highest in young children and a study of routine surveillance data in Germany has suggested that cases in children may be notified earlier in the norovirus season than cases in adults.[100] This raises the question as to the role of children in wider norovirus transmission and whether reducing infection in children could have protective benefits for other, more vulnerable, age groups. Such benefits have been shown for rotavirus following the introduction of the vaccine. Unlike norovirus, the burden of severe rotavirus disease is borne by the very young, but vaccination of this age group has still been found to provide indirect protection to older children and adults, highlighting the importance of young children in the wider transmission of rotavirus.[48]

Further research is needed to explore the role of children in the spread of IID pathogens and whether cases in children start increasing earlier than cases in adults and could, therefore, provide early warning of circulating infections before they enter healthcare settings. This evidence gap forms the basis for one of the research objectives of this thesis.

1.4 Prevention and control of IID within schools

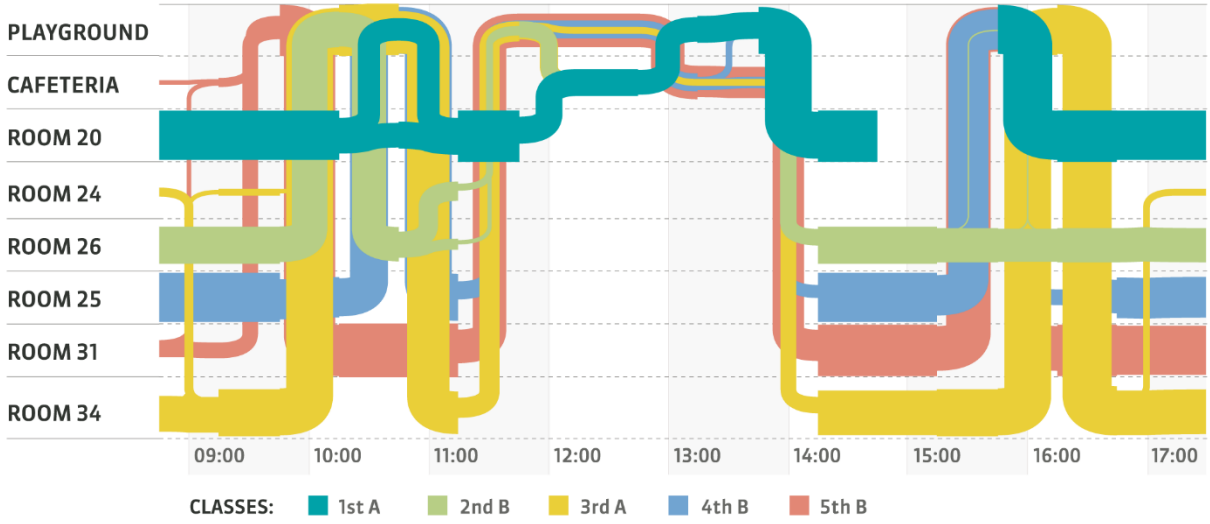
Schools as outbreak settings

Schools are at high risk of experiencing disease outbreaks. This is due to multiple factors such as the number and proximity of people, joint exposure to risk factors, and the challenges of implementing effective infection control procedures within such settings. The large number of close contacts with other children increases the risk of direct person-to-person transmission. Pathogens, particularly viruses, can survive for days or even weeks on surfaces and can contaminate toilet areas, work spaces and equipment, leading to indirect transmission between pupils.[56] The shared use of canteens presents the possibility of large-scale foodborne outbreaks from failings in food hygiene practices. Outbreaks affecting multiple schools have occurred from outside food providers, who cater for a number of schools in an area.[56]

The movement of pupils around the school environment can impact on indirect disease transmission through the shared use of equipment, classrooms, and toilet facilities. One study followed the movements of classes around a primary school, as shown in Figure 1.5.[77] The timetabling of classes

and breaks clearly contributes to the mixing of classes in communal spaces, such as the playground and the canteen. In this study, classes were predominately based within a single classroom, which would reduce the risk of indirect transmission between different classes. However, the structure and layout of individual schools will vary and, in some schools, students may move between classrooms for different subject lessons. In such cases, the risk of indirect disease transmission between separate classes and year groups is likely to increase.

Figure 1.5: Movement of classes through different school areas during a school day



Reproduced under a CC-BY licence from Stehlé et al (2011) [77]

Rashes, IID and respiratory infections are common causes of outbreaks in schools,[101–104] although the relative burden of each of these infections in schools has not been identified from the literature. In 2009-2010, outbreaks of IID in schools accounted for 10% of all reported IID outbreaks in the United States.[62] Schools were responsible for 74% of *Shigella* outbreaks and 5-10% of norovirus outbreaks.[62,105] Norovirus was the predominant cause of school IID outbreaks, accounting for between 21-34% of those reported,[56,62] whilst *Salmonella* was the most commonly identified cause of foodborne outbreaks.[106] The predominant modes of transmission in school outbreaks were person-to-person and foodborne, with waterborne pathogens and contact with animals also reported.[56,107] The scale of school outbreaks is difficult to establish from the literature, as there is likely to be a significant bias towards reporting and publishing large-scale outbreaks. Studies suggest a mean attack rate of 28-30% in school outbreaks, with a median of 40-42 people affected per outbreak.[105–107] Of those infected during a reported school outbreak, 3-5% were hospitalised and 0.02% died as the result of their illness.[56,106] However, not all schools will be affected equally by

outbreaks and there is evidence that outbreaks and subsequent attack rates are influenced by the age of pupils and school size.[56,102,108]

Whilst schools represent a significant proportion of IID outbreaks, the relative burden of IID in schools compared to other pathogens has not been investigated in the literature. Furthermore, IID outbreaks documented within the literature are likely to be biased towards large-scale outbreaks, which makes the impact of IID within schools difficult to assess. There is a need to capture and explore a broader range of outbreak data to assess the relative burden of IID in school settings and this evidence gap forms the basis for one of the research objectives of this thesis.

Infection prevention and control in schools

In contrast to clinical settings, there are some unique challenges to implementing infection prevention and control measures within schools. At the most basic level, some schools may lack the facilities required to support adequate infection control, such as a suitable number and location of handbasins, soap and paper towels.[109] Schools may also have inadequate cleaning programmes, especially with regards to cleaning soft and hard toys.[109] Whilst hand hygiene may be directly taught to children in schools, the focus is less on infection control and more on children's general development and learning of appropriate self-care.[110] Furthermore, hygiene and infection control has to compete with other educational priorities in schools and even within health education other issues may be considered of greater importance.[110] However, increased knowledge of germs and hand washing has been associated with higher levels of hand hygiene amongst children,[111] suggesting that even purely educational interventions are likely to have some benefit.

Different approaches to enhancing infection control in school settings have been identified in the literature (Box 1.2), with most focused on improving hand hygiene through the use of hand sanitisers, supervised hand washing, hand hygiene schedules or educational interventions.[112–115] The role of environmental cleaning, including wiping down toys, desk spaces and equipment has also been investigated.[112]

Some studies have shown hand hygiene interventions, such as the use of hand sanitisers or scheduled handwashing, can lead to a reduction in both the incidence of IID and illness absenteeism within schools.[114,115] However, across studies the results have been inconsistent and many studies are of poor methodological quality.[112,113] Furthermore, studies may be multi-faceted, making it difficult to assess the effectiveness of individual components of infection control interventions. Environmental

cleaning has only been incorporated into a minority of studies and primarily alongside other interventions such as hand hygiene practices and educational programmes.[112] Cleaning approaches in schools included wiping down desks once daily with a disinfectant wipe.[116] Similar interventions in day care settings have included regular cleaning of toys and changes to cleaning protocols, such as alternative cleaning products, cleaning toilet areas last, changing mop water regularly and disinfecting cleaning equipment after use.[117–119] Whilst the multi-faceted interventions may be effective at reducing illness amongst children, the contribution of environmental cleaning to infection control cannot be assessed from these studies.

Box 1.2: Infection prevention and control interventions in schools

- Scheduled hand washing
- Hand washing with soap followed by hand sanitiser
- Hand sanitisers within classrooms
- Individual hand sanitisers
- Hand washing prompts and/or encouragement
- Direct teacher supervision
- Education and training for staff
- Educational programmes for children
- Parental information
- Student peer mentors
- Cleaning toys, desks or equipment
- Reinforcement by school nurses

Alongside hand hygiene and environmental cleaning, public health guidance recommends exclusion of symptomatic cases from school to control the spread of infection.[120] Individuals with acute onset diarrhoea and/or vomiting are advised to remain off school until 48 hours after their last episode of diarrhoea or vomiting.[121] However, there is a scarcity of evidence on the incubation period and period of infectiousness of many IID pathogens.[122] Where rates of infection are high, whole class or school closure may be used to interrupt transmission and terminate outbreaks,[123] and this has been used effectively in the management of influenza outbreaks.[124,125] However, it is less clear whether such interventions reduce rates of illness in schools once they re-open, especially for seasonal or pandemic illnesses where rates in the community may also be high.[125–127]

Where available, vaccination is an important infection control strategy and can reduce illness absenteeism and outbreaks in schools.[128–131] This can be via both direct and indirect protection.[130,131] This has particular relevance to norovirus, which is a major cause of outbreaks in schools.[56,62] With a vaccine currently in development, studies have explored the potential impact of different vaccination strategies, with a focus on young children and the elderly who have the highest burden of norovirus disease.[132,133] Modelling suggests that vaccinating young children could

provide indirect protection to school-aged children and reduce cases in this age group.[133] Such protection has also been observed following the introduction of the rotavirus vaccine.[48]

Finally, infection prevention and control in schools is dependent on the prompt recognition of cases and clusters of illness, which allows schools to put additional control measures in place to prevent further transmission.[56] Whilst monitoring attendance is routine practice in UK schools,[134] schools also need to ascertain key symptoms of illness in order to detect outbreaks. Some schools may have developed their own internal systems for capturing such data but it is likely that outbreaks are under-recognised and under-reported by schools.[135–137] In Japan, these surveillance systems are formally established and school data is routinely monitored to identify possible outbreaks early.[138] Such systems could improve infection control in schools and reduce illness and outbreaks amongst school-aged children.

1.5 Surveillance of infectious intestinal disease

National surveillance systems for IID

In the UK, the surveillance of IID is based on statutory notifications, outbreak reports and syndromic surveillance from primary, secondary and remote health services.[139–141] Statutory notifications obligate clinicians to report suspected cases of certain infectious diseases and laboratories must inform Public Health England when they confirm a notifiable organism within a specimen sample.[139] Not all causes of IID are notifiable and only suspected food poisoning and infectious bloody diarrhoea are formally notifiable in the UK. However, there are also voluntary reporting systems established with most laboratories, who submit weekly electronic reports to Public Health England.

Syndromic surveillance systems in the UK collect and augment data from remote telehealth services (NHS 111 and its predecessor NHS Direct), sentinel general practices (GP), emergency department attendances, ambulance calls and GP out-of-hours activity.[140] Data are extracted on the presenting symptoms and/or the suspected diagnosis. The age and geography of the patient are also collected. These data can be used to monitor trends in seasonal diseases, such as norovirus, and to detect changes in the epidemiology of these infections. However, the geographical information available is not sufficient to allow detection of outbreaks within local areas or outbreaks within settings such as schools.

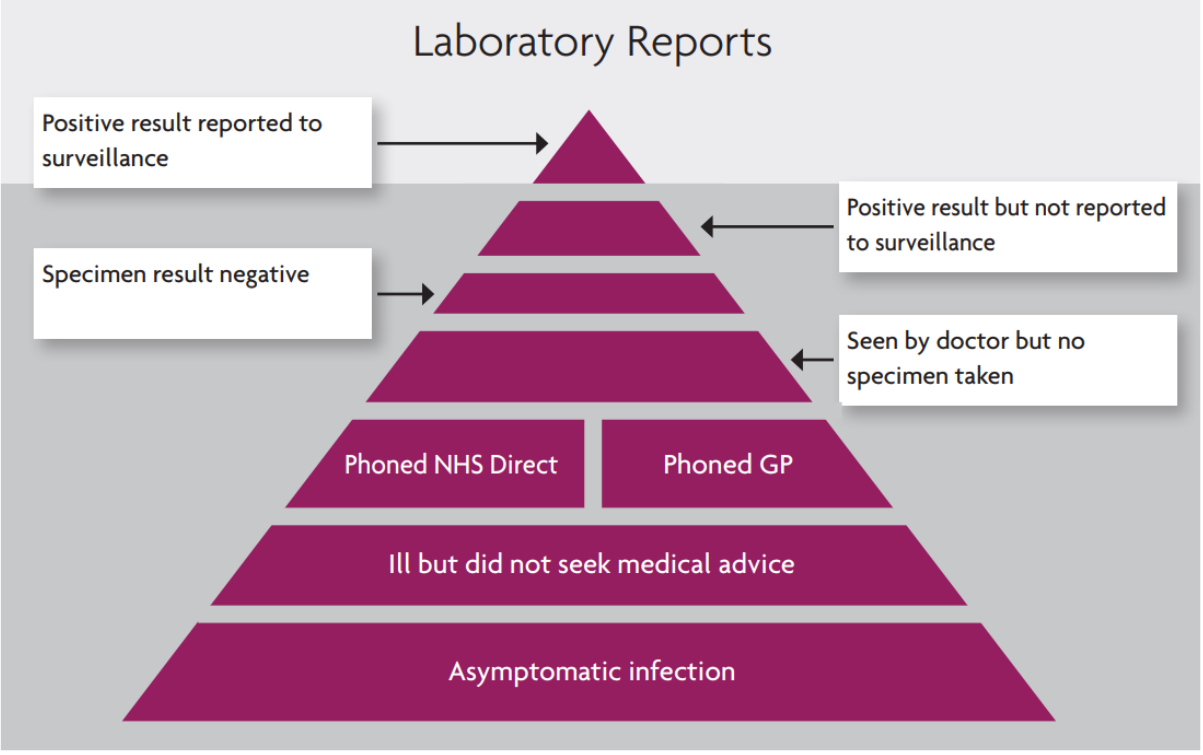
Outbreak surveillance of IID has been in existence in the UK since 1992, and initially consisted of submitting outbreak forms to a national gastrointestinal disease surveillance team.[141] Such systems

have since been replaced by electronic reporting. Hospital norovirus outbreaks are currently reported nationally via the web-based Hospital Norovirus Outbreak Reporting System (HNORS), although participation and reporting is voluntary.[142] Outbreaks in other settings are reported to local Public Health teams, who record details of the outbreak and the subsequent management on a national web-based system.[143]

Case and outbreak definitions

Surveillance of IID is dependent on case definitions to allow accurate detection of cases, clusters and outbreaks of disease. Identification of cases can be based on symptoms, laboratory testing or both. Microbiological testing could be considered the gold standard for case detection as it provides definitive confirmation of the diagnosis through direct detection of the causative organism. However, given IID is typically mild and self-limiting, most symptomatic people will not consult healthcare services and, of those that do, fewer still submit a stool sample (Figure 1.6).

Figure 1.6: The surveillance pyramid: positive cases of IID captured through laboratory reports compared to the true burden of disease.



Reproduced under the Open Government Licence from The Second Study of Infectious Intestinal Disease in the Community [22]

Of the stool samples submitted, microbiological testing fails to detect a pathogen in over half of the samples.[22,27,144] The timing of the sample, volume and laboratory testing policies can also limit case detection.[22,145,146] Case definitions based on microbiological confirmation will, therefore, only capture a minority of the total number of IID cases and in the UK there are an estimated 147 cases in the community for every one confirmed case reported to national surveillance.[29] For some organisms this number is even higher, with 288 norovirus cases for every confirmed case reported.[29]

In contrast, syndromic surveillance systems are based on reported symptoms and/or suspected diagnosis only, with case definitions varying depending on the data source. In general practice (GP), syndromic indicators have been developed based on the Read Code system, which is the recommended national diagnostic classification system for general practice.[147] These syndromic indicators include gastroenteritis, vomiting, and diarrhoea although each indicator may be triggered by a variety of different Read Codes. Similarly, emergency department (ED) surveillance is based on developing syndromic indicators by aggregating relevant diagnostic codes, although a number of different coding systems are used by emergency departments across the UK.[148] Whilst both these systems are based on clinical diagnoses, the diagnosis is not confirmed or validated prior to the data being submitted to the surveillance system.

Ambulance call outs have also been utilised in surveillance, with syndromic indicators based on a list of chief presenting complaints (CPCs) used by the ambulance service.[149] However, none of the resulting syndromic indicators are directly relevant to IID. Unlike primary and secondary healthcare services, telehealth services do not have a diagnostic coding system but instead use electronic algorithms, which contain a series of questions relating to a reported symptom.[150] Syndromic surveillance is based on monitoring how often these algorithms are triggered and identifying exceedances from the normal background level. Relevant algorithms for IID include both vomiting and diarrhoea.

Whilst symptom-based case definitions have the potential to capture a greater proportion of community cases than microbiological definitions, this can be at the expense of specificity. Diarrhoea and vomiting are broad syndromes which can be due to both infective and non-infective causes and can be acute or chronic. The interpretation of diarrhoea can also be subjective and varies between individuals, although it is typically defined by a change in both stool frequency and consistency. The World Health Organization (WHO) defines diarrhoea as “three or more loose or liquid stools per day…or more frequent passage than is normal for the individual”.[151] The International Collaboration on Enteric Disease Burden of Illness Studies suggested an IID case should be defined as “an individual

with ≥3 loose stools, or any vomiting, in 24 h, but excluding those (a) with cancer of the bowel, irritable bowel syndrome, Crohn's disease, ulcerative colitis, cystic fibrosis, coeliac disease, or another chronic illness with symptoms of diarrhoea or vomiting, or (b) who report their symptoms were due to drugs, alcohol, or pregnancy."[152] Other case definitions of IID have also been proposed, including "loose stools or clinically significant vomiting lasting less than 2 weeks, in the absence of a known non-infectious cause, preceded by a symptom-free period of 3 weeks".[29] Whilst such definitions will improve the specificity of syndromic case detection, this level of information may not be readily available in non-healthcare settings.

Finally, for outbreak surveillance, an outbreak is defined as two or more cases linked in time or place, or a greater than expected rate of infection compared with the usual background rate for a given place and time.[120] Whilst laboratory typing can be used to identify linked cases, epidemiological criteria have also been developed to allow professionals to identify norovirus outbreaks in the absence of laboratory confirmation. These include criteria such as the proportion of people affected by vomiting, the incubation period, the mean duration of illness, the fever-to-vomiting ratio and the diarrhoea-to-vomiting ratio.[153–155] The most notable of these are the Kaplan criteria which were developed to distinguish norovirus outbreaks from bacterial IID at a time when diagnostic tests for norovirus were not widely available.[153] The Kaplan criteria consists of four criteria:

1. Vomiting in greater than 50% of affected persons in the outbreak
2. A mean (or median) incubation period of 24–48 hours
3. A mean (or median) duration of illness of 12–60 hours
4. Lack of identification of a bacterial pathogen in a stool culture

A subsequent re-evaluation of these criteria, once diagnostic tests became available, found them to be highly specific (99%) and moderately sensitive (68%) at distinguishing outbreaks of norovirus from bacterial IID outbreaks.[156] Whilst norovirus laboratory tests are now more widely available, such criteria may still be of value in settings where adequate sampling is difficult to achieve, such as schools or workplaces.

Although epidemiological criteria can be used to distinguish norovirus from bacterial IID in outbreak scenarios, these cannot be applied to individual cases. If similar symptom-based definitions could be developed to discriminate between different organisms for single cases of IID, they could significantly enhance syndromic surveillance systems. This evidence gap forms the basis for one of the research objectives of this thesis.

Surveillance of IID in children and schools

Whilst cases of IID in children may be detected via both healthcare consultation data and laboratory testing services, both rely on contact with healthcare services. As many cases of IID in children will be mild, they are more likely to be managed conservatively at home without healthcare input. Consequently, it is only the more severe or persistent illness which will be captured by current surveillance systems. Similarly, outbreaks in schools are often under-recognised and under-reported to public health teams.[135–137] Given the potential role of children in the community transmission of IID, it is important surveillance systems can detect and monitor milder illness in the community before it spreads into higher risk settings such as care homes and hospitals.

There is increasing interest in the use of novel sources of surveillance data, such as over-the-counter purchases, internet-based health searches and worker absenteeism, to better capture illness in the community.[157–164] For children, school attendance registers offer a novel dataset which could be used to provide more timely information regarding infectious disease cases and outbreaks in this age group.[102] Several studies have explored using school absenteeism data to monitor both seasonal and pandemic influenza, especially during the H1N1 influenza pandemic.[165–168] School attendance registers have been found to provide a lead time of 1-2 weeks ahead of traditional surveillance indicators,[165,166] although the benefits of such systems may be affected by the specificity of the absence data collected, with disease-specific absence data proving a better predictor of outbreaks and peaks of seasonal influenza compared to illness absence and all-cause absence.[165–168] However, no studies have been identified that examine the use of school absence data for IID surveillance. Given the high incidence of IID in children, a school-based surveillance system has the potential to be a significant asset to syndromic surveillance in the UK by improving both case detection and the timely recognition of, and response to, school outbreaks.

However, the development of a school-based surveillance system for IID is also dependent on having the structures in place to support data collection, reporting and analysis. In the UK, it is mandatory for schools to record pupil absence and a reason for that absence, with attendance recorded twice each day. The different causes for absence are coded, so that schools can monitor in a consistent way the reasons that a given child is not at school.[134] Illness is given a single code so it can be distinguished from other forms of absence, including those for hospital or dental appointments. However, no other information regarding the illness is recorded. The presence of daily electronic records of attendance presents an opportunity for this dataset to be used for syndromic disease surveillance.

Further study is required to assess the potential role of school attendance registers in the surveillance of IID and the feasibility of implementing a syndromic surveillance system in school settings. Such a system could have direct benefits for schools, allowing prompt identification of clusters and outbreaks so that effective interventions can be put into place to reduce the spread of infection. This would not only reduce the burden of disease within this age group but would also reduce disease transmission into the community. Furthermore, monitoring illness in children could provide an early warning of circulating IID infections. This could allow pre-emptive infection control measures to be put into place in other settings, such as hospitals and nursing homes, thereby reducing the impact of these pathogens on the most vulnerable in society. Exploring the role of school attendance registers in the surveillance of IID forms the basis for one of the research objectives of this thesis.

1.6 Research aims and objectives

This thesis aims to improve knowledge and understanding of the burden of infectious intestinal disease in schools and school-aged children, and the potential role and utility of school attendance registers in the surveillance of IID in children. The following research objectives address gaps identified in the literature.

Research objectives:

1. To describe the burden of IID on illness absence and outbreaks in school settings.
2. To consider the role and utility of school attendance registers in the surveillance of IID.
3. To explore whether children provide an early warning of IID infections in the community.
4. To investigate whether symptoms alone can distinguish between different IID organisms, to enhance case definitions of IID.

1.7 Thesis outline

This thesis consists of six results chapters and a final discussion chapter, which brings together the key findings and conclusions of the research. Five results chapters detail research papers which have been peer-reviewed and published. One results chapter presents a published study protocol alongside a discussion of the subsequent findings and outcomes. Each results chapter contains an introductory section which outlines how the paper fits within the thesis and addresses the above research aims and objectives. This is in accordance with the University of Liverpool Postgraduate Research Code of Practice. The structure of the thesis is as follows:

Chapter 2 utilises data from Public Health England to describe the epidemiology of school-based outbreaks. The relative impact of IID, compared to other causes, is considered and regression analysis is used to explore factors associated with both outbreak occurrence and attack rate. This provides an assessment of the impact of IID in schools in terms of the number of outbreaks, outbreak size and attack rate, meeting the first research objective.

Chapter 3 addresses the second research objective and details a systematic review exploring the utility of school-based surveillance systems in monitoring infectious disease cases and outbreaks in school-aged children. The systematic review considers how well school absenteeism correlates with healthcare surveillance data during seasonal and pandemic disease and whether it provides a lead time ahead of other surveillance indicators.

Chapter 4 is an analytical study exploring whether school illness absences in Merseyside, UK can be used to capture seasonal trends of infectious intestinal disease amongst school-aged children. Trends in school illness absence are compared with routine health surveillance data from primary care, laboratory, and telehealth services. This study also seeks to estimate the burden of IID absenteeism in schools, addressing both research objectives one and two.

Chapter 5 uses national surveillance data for England, UK to explore whether cases of norovirus in children and outbreaks of IID in schools occur earlier in the season than cases and outbreaks amongst adult age groups. This chapter helps improve our understanding of the potential for children and school-based surveillance data to provide an early warning of seasonal norovirus infection and meets the third research objective.

Chapter 6 addresses the fourth research objective and is a secondary analysis of data from a large community cohort and general practice study. This chapter uses multivariable logistic regression analysis to investigate whether symptoms alone can be used to make inferences about the causative organism for individual cases of IID.

Chapter 7 builds on the findings of the previous chapter and considers whether symptom profiles vary across different age groups. A descriptive analysis compares the prevalence of different symptoms in children, adults, and elderly, whilst a multivariable regression analysis explores which symptoms can be used to distinguish bacterial and viral IID within the different age groups. This study also contributes towards the fourth research objective.

Chapter 8 is a discussion of the main findings of this thesis. The results are discussed in terms of their contribution to our knowledge and understanding of the burden of IID in schools and the potential role and utility of school attendance registers in the surveillance of infectious intestinal disease in children. The strengths and limitations of this work are considered as well as the implications of this thesis on future public health research, policy, and practice.

Chapter 2. Infectious disease outbreaks in school settings

Risk factors associated with outbreaks of seasonal infectious disease in school settings, England UK

Anna Donaldson, John Harris, Roberto Vivancos, Sarah O'Brien

Published: Epidemiology & Infection. 2020; 148: e287

doi:10.1017/S0950268820002824

How this publication fits into my thesis

In this chapter, I explore the burden of IID in schools by analysing data from Public Health England on reported school outbreaks over a two-year period. I use descriptive techniques to assess the impact of IID on outbreak occurrence, outbreak size and attack rate, compared to other infectious illnesses. I then use analytical methods to explore school-level factors which impact on the risk of outbreaks and the subsequent attack rate. This chapter contextualises the impact of IID in schools compared to other infectious diseases. It adds to the background provided in the introductory chapter by identifying factors which contribute to a higher risk of outbreaks in different school settings. Recognising which schools are at higher risk of outbreaks is important to help target health protection interventions to reduce transmission between children and prevent future outbreaks.

My contribution

I designed the study with co-authors. I undertook the data cleaning, statistical analysis and wrote the first draft of the manuscript.

2.1 Summary

Children are important transmitters of infection. Within schools they encounter large numbers of contacts and infections can spread easily causing outbreaks. However, not all schools are affected equally. We conducted a retrospective analysis of school outbreaks to identify factors associated with the risk of gastroenteritis, influenza, rash or other outbreaks. Data on reported school outbreaks in England were obtained from Public Health England and linked with data from the Department for Education and the Office for Standards in Education, Children's Services and Skills (Ofsted). Primary and all-through schools were found to be at increased risk of outbreaks, compared with secondary schools (OR 5.82, 95% CI 4.50-7.58 and OR 4.66, 95% CI 3.27-6.61 respectively). School size was also significantly associated with the risk of outbreaks, with higher odds associated with larger schools. Attack rates were higher in gastroenteritis and influenza outbreaks, with lower attack rates associated with rashes (RR 0.17, 95% CI 0.15-0.20). Deprivation and Ofsted rating were not associated with either outbreak occurrence or the subsequent attack rate. This study identifies primary and all-through schools as key settings for health protection interventions. Public health teams need to work closely with these schools to encourage early identification and reporting of outbreaks.

2.2 Introduction

Children are recognised as important transmitters of seasonal and pandemic infectious disease.[84,85,92] They have more naïve immune systems which increase susceptibility to infection and are commonly acknowledged to have poorer levels of hand and respiratory hygiene. Within schools, they experience a large number of contacts with peers,[85] and infections can spread easily through direct and indirect transmission.[56,65] Unlike healthcare settings, schools do not have standard infection control practices and may have inadequate cleaning programmes or lack the facilities to support proper handwashing.[109] Furthermore, infection control is not a key focus for schools or teachers and implementing hygiene interventions can be challenging.[110] These factors result in an increased risk of outbreaks within school settings.

Rashes, gastrointestinal and respiratory infections are common causes of outbreaks within schools.[101–104] Outbreaks need to be managed promptly to prevent further spread of infection both within the school environment and outward to households and the wider community.[87] Measures such as exclusion, environmental cleaning, hand washing and promoting good respiratory hygiene can be used to control outbreaks.[56,122] Public health agencies can offer advice and support to schools on the management of outbreaks, but it is likely that many outbreaks go unrecognised and unreported.[136,137]

Not all schools are affected equally by outbreaks and there is evidence that outbreaks and subsequent attack rates are influenced by the age of pupils and school size.[56,102,108] Socioeconomic status may influence disease risk and severity [169,170] as well as affecting vaccine uptake [171] which could contribute towards the risk of outbreaks in school settings. This study seeks to identify factors associated with outbreak occurrence and attack rate in schools. Understanding which factors increase the risk of outbreaks could help identify higher-risk schools and support targeted interventions and training to help prevent future outbreaks.

2.3 Methods

Study population

The study population was schools in England, UK. There are just under 24 000 registered schools in England, which include state-schools, academies, independent/private schools, special schools for children with special educational needs, and pupil referral units for children who aren't able to attend a mainstream school.[172] Together, these schools cover a pupil population of 8.8 million. Schools teach primary education (ages 4-11yrs), secondary education (ages 11-18yrs) or both (all-through schools).

Data sources

Data on reported school outbreaks in England are held by Public Health England (PHE). Outbreaks in school settings are self-reported to PHE by schools, in line with national guidance for health protection in schools.[120] An outbreak is defined as two or more cases linked in time or place, or a greater than expected rate of infection compared with the usual background rate for a given place and time.[120] Schools are advised to contact PHE as soon as an outbreak is suspected, although reporting is not mandatory. The decision to declare an outbreak is made by PHE and the school is advised on appropriate infection control measures to manage the outbreak. Such measures include hand hygiene, respiratory hygiene, environmental cleaning, exclusion, and letters to parents.[120] Data on reported school outbreaks across England were extracted from the Public Health England database for the 2016/2017 and 2017/2018 academic years. The academic year was defined as running from 1st September to the 31st August. Outbreaks linked to nurseries, universities or colleges for those over 18 years of age, care homes for children, households, community settings or visitor attractions were excluded from the dataset. Special schools and pupil referral units were included.

Outbreak data were combined with nationally available data from the Department for Education and the Office for Standards in Education, Children's Services and Skills (Ofsted).[173,174] The Department

for Education provides routine data on the demographics and performance of registered schools, and Ofsted publishes data on school inspections. Ofsted routinely inspect all state-registered schools and assesses them according to pupil outcomes, quality of teaching, learning and assessment, effectiveness of leadership and management, and personal development, behaviour and welfare. Schools are then given an overall effectiveness score which ranges from 1-4 (outstanding to inadequate). Ofsted also provide a deprivation quintile for each school, which is calculated from the Index of Multiple Deprivation (IMD) for the postcode of residence of each pupil.[175] The IMD scores are then averaged for the school to give an overall score and schools are placed into quintiles; quintile one representing the least deprived and quintile five representing the most deprived.

Statistical methods

The unique reference number (URN) was identified for each school within the outbreak dataset and used to link Public Health England data with data from the Department for Education and Ofsted, for the corresponding academic year. Descriptive statistics were used to explore seasonal trends in outbreaks and variations in the number and proportion of outbreaks, broken down by different explanatory variables. Variables were selected based on the experience of the research group and the availability of national school-level data. The explanatory variables included the size of the school, phase of education, Ofsted score, deprivation index and the gender of the school (single sex verses mixed). The size of school was included as a categorical variable, the categories determined by the distribution of the data. Phase of education was categorised as primary, secondary and all-through schools (covering both primary and secondary year groups). Multivariable logistic regression was used to identify factors associated with schools who had experienced an outbreak, compared with schools which had not had an outbreak. Associations were compared for schools with one outbreak and schools with two or more outbreaks over the study period. Modelling was repeated, stratified by outbreak cause, to explore differences in the predictors of outbreak occurrence for the major causes of outbreaks. Cramer's V coefficients were used to identify any significant correlations between the explanatory variables which could affect the regression model and Variance Inflation Factors (VIF) were used to check the model for multicollinearity.

Attack rates were calculated by dividing the number of symptomatic pupils by the total number in the school, presented as a rate per 100 pupils/year. Descriptive statistics were used to explore variations in attack rate across the different explanatory variables. Explanatory variables included the size of the school, phase of education, Ofsted score, and deprivation index, as well as the cause of the outbreak and the delay in reporting. Size of school and phase of education were categorised as described above.

The cause of the outbreak was broken down into gastroenteritis, influenza, rash or other. Delay in reporting was calculated as the difference in days between the date of onset of the index case and the date the outbreak was reported to Public Health England. As delay in reporting was skewed, a sensitivity analysis was undertaken, removing major and minor outliers to explore the impact of skewed data on model performance. Major outliers were defined as datapoints falling more than three times the interquartile range above the third quartile, and minor outliers as falling one and a half times the interquartile range above the third quartile. Quasi-Poisson regression was used to identify factors associated with attack rate for all outbreaks and outbreaks stratified by cause. This was chosen instead of a Poisson regression due to the high level of variance within the count data. Quasi-Poisson methods relax the assumption that the variance is equal to the mean and allows for more robust calculation of confidence intervals.

Odds ratios (OR) and 95% confidence intervals (CI) were calculated from the logistic regression models. Relative risk (RR) with corresponding confidence intervals were generated from the Quasi-Poisson regression modelling. All statistical analysis was undertaken in R 3.3.2.[176]

2.4 Results

From 1st September 2016 to 31st August 2018, there were 2 207 outbreaks in schools reported to Public Health England. Of these, 90 were excluded as they did not meet the inclusion criteria. Gastroenteritis was the most common cause of reported outbreaks, accounting for 47% (n=998). This was followed by rash (44%, n=935) and influenza (6%, n=126). Other causes of outbreaks accounted for less than one per cent each and included respiratory tract infections, conjunctivitis, hepatitis, impetigo, infestations, and worms. Outbreaks ranged in size from 2 to 300 cases, with a median of 10 cases per outbreak. There were clear seasonal trends in reported outbreaks, as shown in Figure 2.1. Across both academic years, gastroenteritis outbreaks peaked between November and January, followed by a peak in rashes between January and March. In 2017/2018, the peak in rash outbreaks was particularly dominant, driving a higher number of reported outbreaks that year.

Outbreaks occurred in primary, secondary and all-through schools. Schools ranged in size from 11 pupils to 2 170, with a median size of 299 pupils. Only 1% of outbreak schools were single sex schools, so this was dropped as a covariate in the regression analysis. One outbreak occurred in a pupil referral unit (PRU), which was subsequently excluded due to a lack of available national data on PRUs. Ofsted ratings were available for 1 731 (94%) of the schools. The scores in the sub-categories tended to be consistent with the overall score, therefore only the overall Ofsted rating was included in the analysis.

Figure 2.1: School outbreaks reported to Public Health England by week and cause, 1st September 2016 – 31st August 2018

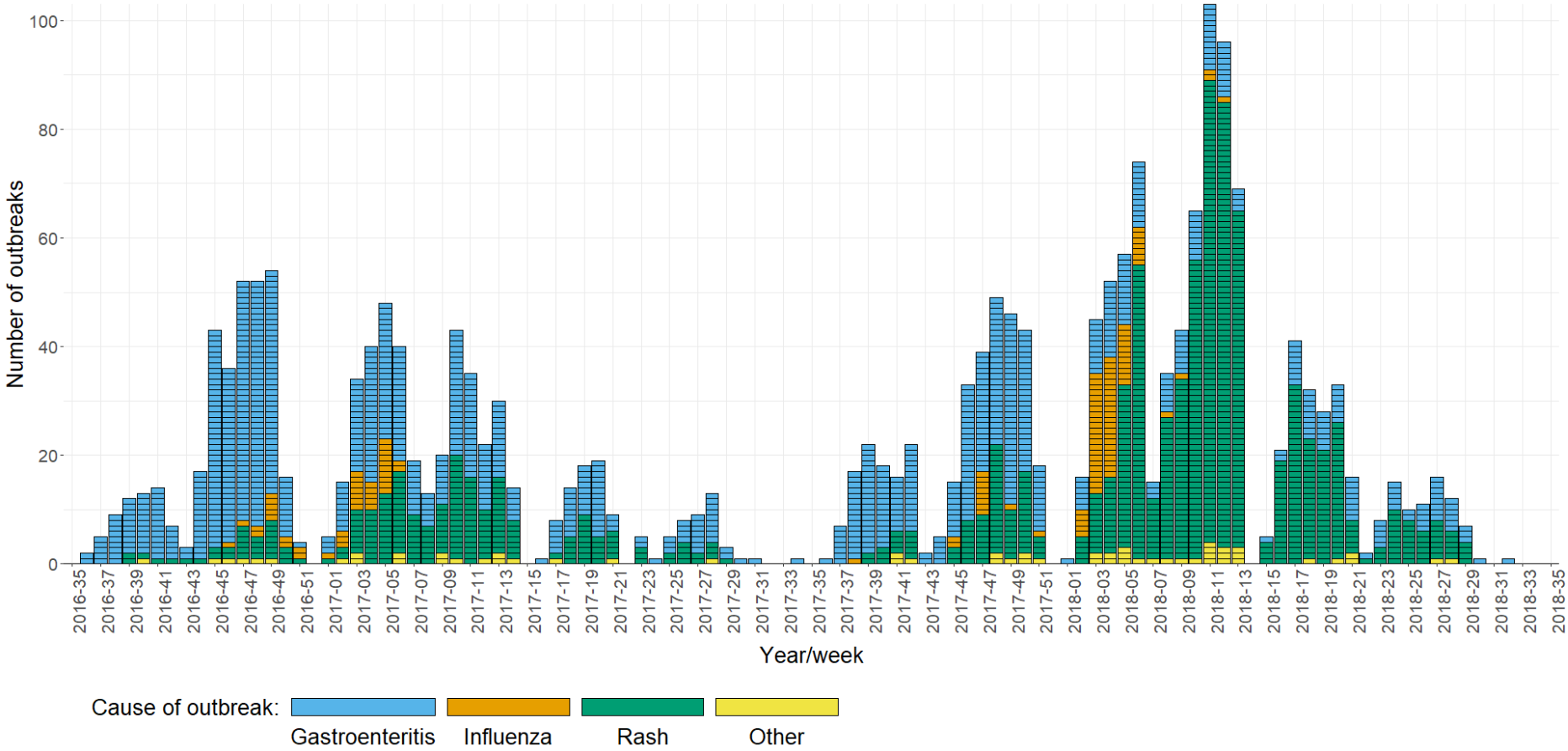


Table 2.1: Summary of the number and proportion of outbreaks, by cause and explanatory variable

Explanatory variables	All schools n (%)	All outbreaks n (%)	Gastroenteritis n (%)	Rash n (%)	Influenza n (%)	Other n (%)
Total number of outbreaks	/	2 116	997	935	126	58
Total number of schools	26 985	1 841	923	859	124	57
School size						
≤50	1 339 (5.0%)	20 (1.1%)	15 (1.6%)	3 (0.3%)	2 (1.6%)	0 (0%)
51 – 200	6 878 (25.5%)	413 (22.4%)	236 (25.6%)	162 (18.9%)	17 (13.7%)	10 (17.5%)
201 – 400	8 578 (31.8%)	727 (39.5%)	359 (38.9%)	360 (41.9%)	40 (32.3%)	20 (35.1%)
401 – 600	4 047 (15.0%)	396 (21.5%)	185 (20.0%)	212 (24.7%)	30 (24.2%)	8 (14.0%)
>600	3 879 (14.4%)	248 (13.5%)	113 (12.2%)	106 (12.3%)	32 (25.8%)	15 (26.3%)
Not known	2 264 (8.4%)	37 (2.0%)	15 (1.6%)	16 (1.9%)	3 (2.4%)	4 (7.0%)
Phase of education						
Primary	20 151 (74.8%)	1 591 (86.4%)	784 (84.9%)	804 (93.6%)	72 (58.0%)	37 (64.9%)
Secondary	4 498 (16.7%)	114 (6.2%)	61 (6.6%)	18 (2.1%)	28 (22.6%)	14 (24.6%)
All-through	2 275 (8.4%)	136 (7.4%)	78 (8.5%)	37 (4.3%)	24 (19.4%)	6 (10.5%)
Not known	61 (0.2%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Ofsted overall rating						
1 – Outstanding	4 598 (17.0%)	372 (20.2%)	179 (19.4%)	175 (20.4%)	25 (20.2%)	13 (22.8%)
2 – Good	15 867 (58.8%)	1 130 (61.4%)	564 (61.1%)	550 (64.0%)	59 (47.6%)	30 (52.6%)
3 – Requires Improvement	2 567 (9.5%)	180 (9.8%)	100 (10.8%)	77 (9.0%)	15 (12.1%)	3 (5.3%)
4 – Inadequate	1 085 (4.0%)	49 (2.7%)	27 (2.9%)	17 (2.0%)	5 (4.0%)	2 (3.5%)

NA	2 868 (10.6%)	110 (6.0%)	53 (5.7%)	40 (4.7%)	20 (16.1%)	9 (15.8%)
Deprivation						
1 – Least deprived	4 370 (16.2%)	316 (17.2%)	149 (16.1%)	147 (17.1%)	29 (23.4%)	13 (22.8%)
2	4 483 (16.6%)	338 (18.4%)	158 (17.1%)	176 (20.5%)	18 (14.5%)	7 (12.3%)
3	4 543 (16.8%)	365 (19.8%)	198 (21.5%)	168 (19.6%)	16 (12.9%)	7 (12.3%)
4	4 670 (17.3%)	334 (18.1%)	173 (18.7%)	157 (18.3%)	19 (15.3%)	10 (17.5%)
5 – Most deprived	4 539 (16.8%)	357 (19.4%)	182 (19.7%)	164 (19.1%)	16 (12.9%)	10 (17.5%)
Not known	4 380 (16.2%)	131 (7.1%)	63 (6.8%)	47 (5.5%)	26 (21.0%)	10 (17.5%)

Outbreak occurrence

A total of 1 841 schools experienced at least one outbreak across the study period and 232 (12.6%) reported more than one outbreak. Primary schools accounted for 86% of outbreak schools, whilst 7% were all-through schools and 6% were secondary schools (Table 2.1). There was evidence of correlation between school size and phase of education (Cramer V = 0.52), but no other correlations were found between the explanatory variables. There was no evidence of multicollinearity in the regression models (VIF<2).

Table 2.2: Factors associated with the odds of an outbreak in a school, by cause; a multivariable logistic regression model

Explanatory variables	All outbreaks OR (95% CI)	Gastroenteritis OR (95% CI)	Rash OR (95% CI)
Phase of education			
Secondary	1	1	1
Primary	5.82 (4.50-7.58)	4.53 (3.19-6.54)	18.31 (11.17-31.94)
All-through	4.66 (3.27-6.61)	4.76 (3.01-7.49)	6.56 (3.13-13.54)
School size			
≤ 50	0.26 (0.14-0.42)	0.43 (0.22-0.76)	0.08 (0.01-0.24)
51-200	0.77 (0.68-0.88)	0.94 (0.79-1.12)	0.62 (0.51-0.75)
201-400	1	1	1
401-600	1.23 (1.07-1.40)	1.07 (0.88-1.29)	1.44 (1.20 -1.72)
>600	1.80 (1.48-2.18)	1.46 (1.10-1.91)	2.26 (1.73-2.91)
Ofsted rating			
1 - Outstanding	1	1	1
2 - Good	0.89 (0.79-1.02)	0.91 (0.76-1.08)	0.93 (0.78-1.12)
3 – Requires improvement	1.06 (0.87-1.29)	1.13 (0.87-1.47)	1.09 (0.82-1.43)
4 – Inadequate	0.92 (0.64-1.29)	1.19 (0.75-1.80)	0.64 (0.32-1.13)
Deprivation Index			
1 – Least deprived	1	1	1
2	1.05 (0.89-1.23)	1.01 (0.80-1.27)	1.18 (0.94-1.49)
3	1.08 (0.92-1.27)	1.27 (1.02-1.58)	1.01 (0.80-1.28)
4	0.89 (0.75-1.05)	1.05 (0.83-1.32)	0.82 (0.64-1.05)
5 – Most deprived	0.95 (0.81-1.13)	1.09 (0.87-1.38)	0.86 (0.67-1.09)

OR, odds ratio; CI, confidence interval.

Primary and all-through schools were found to be at increased risk of an outbreak, compared with secondary schools (OR 5.82, 95% CI 4.50-7.58 and OR 4.66, 95% CI 3.27-6.61 respectively). Outbreak occurrence was also significantly associated with school size, with the odds of an outbreak increasing as school size increased (Table 2.2). The occurrence of an outbreak was not associated with Ofsted rating or the level of deprivation in a school. Similar associations were found for the occurrence of multiple outbreaks and across both gastroenteritis and rash outbreaks. The number of influenza outbreaks were insufficient to allow separate analysis.

Attack rates

Of 2 116 outbreaks, 1 813 (86%) had data on attack rates. Attack rates ranged from 0.1 per 100 pupils to 74.2 per 100, with a median attack rate of 3.6 per 100 pupils. Attack rates varied depending on the cause of the outbreak, with the highest median attack rate occurring for gastroenteritis outbreaks (8.3 per 100) and the lowest for rash (1.1 per 100) (Table 2.3). Median attack rate decreased as school size increased but the number of cases did not vary accordingly, with all but the smallest school size having a median of 9-10 cases per outbreak. Delays in reporting varied from zero (outbreak reported same day as onset) to 105 days. The median delay in reporting was 3 days. This was similar for gastroenteritis outbreaks (3 days), rash (4 days) and influenza (4 days). In addition to the correlation identified between phase of education and school size (Cramer's V = 0.42), there was some evidence of association between the cause of outbreak and delay in reporting (Cramer's V = 0.34). No other correlations were identified between the explanatory variables. There was no evidence of multicollinearity in the regression model (VIF<2.5).

Attack rates were significantly lower in rash and 'other' outbreaks compared with gastroenteritis (RR 0.17, 95% CI 0.15-0.20 and RR 0.62, 95% CI 0.37-0.99 respectively), but there was no difference between influenza and gastroenteritis outbreaks (Table 2.4). School size was also associated with attack rate, with attack rates decreasing as school size increased. Primary schools had higher attack rates compared with secondary schools for gastroenteritis and rash outbreaks, but not for all-cause outbreaks. Attack rate was found to increase marginally with each additional day delay in reporting (RR 1.01, 95% CI 1.00-1.02). This association remained after removing major and minor outliers from the delay in reporting variable but was not significant for gastroenteritis outbreaks. Neither deprivation nor Ofsted score were associated with attack rate. The number of influenza outbreaks were too few to allow separate analysis.

Table 2.3: Summary of the number of cases, number at risk and attack rate, by explanatory variable

Explanatory variable	Number of symptomatic cases Median (IQR)	Number at risk Median (IQR)	Attack rate Median (IQR)
School size			
≤50	4 (3-7)	36 (27-48)	11.8 (8.6-22.5)
51 – 200	9 (3-20)	134 (96-175)	7.2 (3.0-16.7)
201 – 400	10 (3-23)	262 (218-335)	3.5 (1.2-9.3)
401 – 600	10 (3-24)	442 (419-481)	2.4 (0.7-5.8)
>600	10 (3-30)	717 (630-949)	1.4 (0.4-5.8)
Phase of education			
Primary	10 (3-22)	280 (197-420)	3.6 (1.2-9.6)
Secondary	12 (4-40)	917 (580-1200)	2.6 (0.4-10.3)
All-through	8 (3-20)	285 (122-471)	4.2 (1.4-10.1)
Cause			
Gastroenteritis	20 (11-33)	248 (173-419)	8.3 (4.2-15.2)
Influenza	23 (13-44)	392 (230-579)	7.1 (3.7-11.7)
Rash	3 (2-5)	330 (209-450)	1.1 (0.6-2.3)
Other	7 (3-12)	232 (144-610)	2.5 (0.9-6.7)
Ofsted overall rating			
1 – Outstanding	9 (3-22)	289 (179-472)	3.4 (1.1-9.7)
2 – Good	10 (3-22)	277 (190-420)	3.7 (1.2-10.0)
3 – Requires Improvement	10 (3-25)	331 (216-447)	3.7 (1.2-8.6)
4 – Inadequate	10 (4-24)	310 (221-650)	2.8 (1.1-8.2)
Deprivation			
1 – Least deprived	10 (3-20)	213 (120-401)	4.0 (1.6-11.2)
2	7 (3-22)	251 (175-408)	3.7 (1.2-10.5)
3	10 (3-23)	270 (186-421)	3.5 (1.2-9.8)
4	10 (4-25)	349 (213-455)	3.4 (1.1-9.0)
5 – Most deprived	11 (4-24)	360 (230-471)	3.5 (1.0-8.5)

Attack rate per 100 pupils/year; IQR, interquartile range.

Table 2.4: Factors associated with outbreak attack rate in schools, by cause; a multivariable Quasi-Poisson regression model

Explanatory variables	All outbreaks RR (95% CI)	Gastroenteritis RR (95% CI)	Rash RR (95% CI)
Phase of education			
Secondary	1	1	1
Primary	1.15 (0.93-1.41)	1.30 (1.00-1.71)	2.56 (1.08-7.44)
All-through	0.92 (0.66-1.27)	0.87 (0.56-1.31)	1.44 (0.32-5.74)
School size			
≤ 50	2.82 (1.22-5.45)	3.15 (1.18-6.69)	5.62 (0.32-25.33)
51-200	1.58 (1.37-1.82)	1.53 (1.29-1.82)	1.82 (1.37-2.40)
201-400	1	1	1
401-600	0.74 (0.65-0.85)	0.74 (0.62-0.87)	0.70 (0.55-0.90)
>600	0.69 (0.58-0.83)	0.76 (0.61-0.95)	0.38 (0.25-0.56)
Cause			
Gastroenteritis	1	-	-
Influenza	1.00 (0.85-1.17)	-	-
Rash	0.17 (0.15-0.20)	-	-
Other	0.62 (0.37-0.99)	-	-
Ofsted rating			
1 - Outstanding	1	1	1
2 - Good	0.97 (0.85-1.11)	0.89 (0.76-1.05)	0.91 (0.71-1.20)
3 – Requires improvement	0.99 (0.82-1.19)	0.90 (0.70-1.14)	0.90 (0.60-1.34)
4 – Inadequate	0.73 (0.49-1.05)	0.80 (0.49-1.23)	0.94 (0.41-1.88)
Deprivation Index			
1 – Least deprived	1	1	1
2	1.00 (0.86-1.19)	1.05 (0.85-1.30)	1.16 (0.84-1.61)
3	0.98 (0.83-1.16)	0.98 (0.80-1.21)	1.15 (0.83-1.60)
4	1.09 (0.92-1.29)	1.12 (0.91-1.39)	1.31 (0.93-1.85)
5 – Most deprived	1.12 (0.95-1.33)	1.09 (0.88-1.35)	1.51 (1.04-2.18)
Delay in reporting			
Per additional day	1.01 (1.00-1.02)	1.00 (0.99-1.01)	1.03 (1.02-1.03)

RR, relative risk; CI, confidence interval.

2.5 Discussion

In this study, across a two-year period, outbreaks occurred in almost 1 in 10 schools. The most common causes of reported outbreaks were gastroenteritis and rashes, with the majority of outbreaks occurring in primary schools. Primary and all-through schools, as well as larger school size were associated with an increased risk of outbreaks, whilst Ofsted rating and deprivation were not associated with outbreak occurrence. Attack rates were higher in smaller schools and with each additional day delay in reporting. Lower attack rates were associated with larger schools and outbreaks caused by rashes.

The finding that primary schools were disproportionately affected by outbreaks is consistent with existing literature.[56,102,166,177] Younger children are known to have a greater vulnerability to infection, higher virus shedding and poorer levels of hand and respiratory hygiene which increase the risk of illness.[85] The effect of school size on outbreak occurrence mirrors the risk of infection by household size and care home size.[178–180] The increased outbreak risk in larger schools could be attributed to more pupils entering the school environment and therefore more opportunity for infection to be introduced into the school. Larger schools also experienced lower attack rates, a phenomenon observed in other settings.[108,180] In this study, this finding is most likely driven by changes in the attack rate denominator. The median number of cases did not vary significantly between schools of different sizes, but the large differences in the number of children at risk result in larger schools having significantly lower attack rates for the same number of cases. It may be that class size, rather than school size, is a more important variable for disease transmission in schools and studies have shown the majority of children's close contacts in school are within their immediate class.[77,81] These contacts have a high potential for disease transmission, as children are in close proximity to their classmates for prolonged periods of time. Unfortunately, data on class sizes were not available to include in this study, but in England class size is limited to approximately thirty children,[1] so for the majority of schools the class size will be similar. Even small schools may combine year groups to increase otherwise small class sizes. This offers a potential explanation for why similar numbers of cases were observed regardless of the overall size of the school.

Early detection and reporting of outbreaks is crucial for the implementation of control measures and the finding that attack rate increases with delays in reporting is consistent with findings from care home settings.[180] However, the association was borderline and for schools this relationship is likely to be affected by school holidays, which create a natural break in transmission and help control and terminate outbreaks without additional intervention. The effect of school holidays on the size and timing of seasonal outbreaks is a topic for further investigation.

Of note in this study was the relatively small proportion of influenza outbreaks reported over a two-year period (n=126). This is in strong contrast to the numbers reported in other years within the UK.[102] Influenza strains vary year on year,[181] and so the years included in this study may have captured milder influenza seasons and consequently underestimate the role of influenza in school outbreaks. However, in 2013 the UK started the phased introduction of the universal childhood influenza vaccination,[182] and as the numbers of children receiving the vaccine increase, influenza may become a less frequent cause of school outbreaks. Vaccine uptake, and factors which influence vaccine uptake, may then have an increasing role in the occurrence of school outbreaks. This could alter the importance of deprivation, which was not found to be associated with either outbreak occurrence or attack rate in this study, but has been linked to the uptake of influenza vaccination.[171,183]

Limitations

Schools do not have a statutory duty to report outbreaks to Public Health England and therefore, this study may underestimate the total number of outbreaks occurring. As outbreaks are self-reported, these data may be subject to reporting bias which could impact on the strength of associations within the analysis. Unreported outbreaks are most likely to have small numbers of cases, which do not cause significant disruption to the school. This could impact on the association between delay in reporting and attack rate, as such outbreaks are likely to have low attack rates and yet unmeasurable delays in reporting. Furthermore, the distribution of delays in reporting was skewed, which could affect the regression model performance. However, removing major and minor outliers from this variable did not alter the association between delay in reporting and attack rate, suggesting the skewed data had minimal impact on the results.

A further limitation of the data is the variation in how outbreaks in schools are recorded and followed up. The initial documentation of an outbreak includes the number of symptomatic cases at the time the outbreak is declared, and this will be updated each time contact is made with the school. However, smaller outbreaks which resolve quickly may not necessitate further contact with the school beyond the initial report. Therefore, it is possible that additional cases occurring towards the end of outbreaks may have been missed. Whilst the number of additional cases is likely to be small, the case numbers within the PHE dataset may underestimate the total number of children affected and consequently the attack rates represent a conservative estimate.

Finally, each year different organisms may have a greater or lesser role in driving outbreaks in schools. Organism strains vary and some years will have a greater impact on children and schools than others. Consequently, this analysis is not a definitive assessment of the impact of infectious diseases on schools and ongoing timely data on outbreaks is required to detect the key organisms affecting children.

2.6 Conclusion

This study has identified primary and all-through schools as being at increased risk of outbreaks and therefore health protection interventions need to focus on these settings. Larger schools were also at increased risk and need to ensure they are aware of the importance of infection prevention measures such as handwashing and environmental cleaning. Gastroenteritis and influenza outbreaks were associated with higher attack rates and public health teams need to work closely with schools to encourage early identification and reporting of these outbreaks.

Chapter 3. School-based surveillance of infectious disease in children

School-based surveillance of acute infectious disease in children: a systematic review

Anna Donaldson, Jo Hardstaff, John Harris, Roberto Vivancos, Sarah O'Brien

Protocol published: PROSPERO (2019) CRD42019119737

Manuscript published: BMC Infectious Diseases. 2021; 21(1): 744

doi.org/10.1186/s12879-021-06444-6

How this publication fits into my thesis

In the previous chapter, I identified primary schools and all-through schools as being at increased risk of outbreaks and consequently key settings for health protection interventions. I identified IID as a major cause of reported school outbreaks and the higher attack rate associated with IID outbreaks increases the importance of ensuring schools identify and report these outbreaks early. However, as schools do not have a statutory duty to report outbreaks, it is likely that many outbreaks go unrecognised and unreported. This raises the question as to whether a school-based surveillance system could improve the detection of cases and outbreaks of illness amongst children, removing the reporting bias inherent in the current surveillance data and improving the timely detection and control of school outbreaks.

This chapter is a systematic review exploring the role of school attendance registers in the surveillance of infectious disease in schools. Given the differing methodologies of the studies in this area, including the specificity of data collected and the end points examined, there was a need to assess the literature in a systematic and methodical way to better understand the utility of school attendance registers as a form of syndromic surveillance. The systematic review describes the different school surveillance systems in terms of their purpose, specificity of the absence data and frequency of data collection. Estimates are presented for baseline and peak rates of absenteeism. The level of correlation of school absenteeism with other surveillance systems is assessed along with the lead and lag times of the different surveillance indicators.

My contribution

I designed the study with co-authors and wrote the study protocol. I undertook the search and article screening with JLH. I undertook the full text review and data extraction with JPH. I analysed the findings and drafted the manuscript.

3.1 Summary

Background

Syndromic surveillance systems are an essential component of public health surveillance and can provide timely detection of infectious disease cases and outbreaks. Whilst surveillance systems are generally embedded within healthcare, there is increasing interest in novel data sources for monitoring trends in illness, such as over-the-counter purchases, internet-based health searches and worker absenteeism. This systematic review considers the utility of school attendance registers in the surveillance of infectious disease outbreaks and occurrences amongst children.

Methods

We searched eight databases using key words related to school absence, infectious disease and syndromic surveillance. Studies were limited to those published after 1st January 1995. Studies based in nursery schools or higher education settings were excluded. Article screening was undertaken by two independent reviewers using agreed eligibility criteria. Data extraction was performed using a standardised data extraction form. Outcomes included estimates of absenteeism, correlation with existing surveillance systems and associated lead or lag times.

Results

Fifteen studies met the inclusion criteria, all of which were concerned with the surveillance of influenza. The specificity of absence data varied between all-cause absence, illness absence and syndrome-specific absence. Systems differed in terms of the frequency of data submissions from schools and the level of aggregation of the data. Baseline rates of illness absence varied between 2.3%-3.7%, with peak absences ranging between 4.1%-9.8%. Syndrome-specific absenteeism had the strongest correlation with other surveillance systems ($r = 0.92$), with illness absenteeism generating mixed results and all-cause absenteeism performing the least well. A similar pattern of results emerged in terms of lead and lag times, with influenza-like illness (ILI)-specific absence providing a 1-2 week lead time, compared to lag times reported for all-cause absence data and inconsistent results for illness absence data.

Conclusion

Syndrome-specific school absences have potential utility in the syndromic surveillance of influenza, demonstrating good correlation with healthcare surveillance data and a lead time of 1-2 weeks ahead of existing surveillance indicators. Further research should consider the utility of school attendance registers for conditions other than influenza, to broaden our understanding of the potential application of these data for infectious disease surveillance in children.

3.2 Introduction

Public health surveillance is the "continuous, systematic collection, analysis and interpretation of health-related data needed for the planning, implementation, and evaluation of public health practice."[184] For infectious disease, timely surveillance systems are fundamental to providing early detection of cases and outbreaks, allowing measures to be put in place to protect others and reduce transmission.[185] Historically, public health surveillance was disease-specific, relying on clinical diagnoses and laboratory reports.[186] However, such surveillance systems can be subject to significant delays and over recent years there has been increasing recognition of the value of syndromic surveillance in providing more timely detection of infectious illness.[187–190]

Syndromic surveillance can be based on either the identification of clinical syndromes that are indicative of a given disease, or the clustering of non-specific symptoms and changes in patterns of health behaviours which could indicate an outbreak or unusual event.[191] Syndromic surveillance systems have been developed using multiple sources of data, many of which are embedded within healthcare, such as emergency department attendances, ambulance dispatches or calls to remote telehealth services.[148,162,192–195] However, there is increasing interest in the use of novel sources of data, such as over-the-counter purchases, internet-based health searches and worker absenteeism, which have been found to correlate well with traditional surveillance indicators.[157–164]

School attendance registers offer a novel dataset which could be used to provide more timely information regarding infectious disease and outbreaks amongst children.[102] Children are commonly affected by gastrointestinal illness and respiratory illness, both of which are key causes of illness absence from school.[115,116,196,197] Children are recognised as important transmitters of infection,[65,85,198,199] and schools are principal settings in the spread of infections between children.[65,81] Close household contact with parents and grandparents facilitates the spread of illness from schools into the wider community.[74,87] School absence data could support the early identification of outbreaks within schools, enabling timely intervention to reduce the transmission of infections both within and outside of the school setting. Furthermore, as school absence may occur from the first day of illness, this novel dataset has the potential to offer more timely data than healthcare-based surveillance. There is evidence that children may be the first affected by seasonal and pandemic illnesses,[200–202] and by enhancing the detection of disease in children such data could provide early warning of infections before they start circulating in the wider community.

This systematic review considers the utility of school attendance registers in the surveillance of infectious disease outbreaks and occurrences amongst children. The value of a school-based surveillance system will be considered in terms of its correlation and lead time compared to traditional surveillance indicators. A secondary objective of this review is to describe the burden of illness absenteeism and outbreaks in school-aged children.

3.3 Methods

Protocol and registration

The systematic review protocol was registered on PROSPERO in January 2019 (PROSPERO 2019 CRD42019119737).[203] The protocol and article follow the PRISMA checklist for the reporting of systematic reviews (Appendix A).

Eligibility criteria

The population of interest for this review was children aged between 4 and 18 years, attending school. Only studies published on or after 1st January 1995 and available in English were included. As this review considers what school attendance data adds to existing health surveillance systems, studies were limited to those from OECD countries,[204] which are likely to have established health surveillance systems in place for comparison. No limitation was put on school type, but studies based in nursery schools or higher education settings were excluded, as these settings are not components of compulsory education and may be subject to different requirements for attendance and absence reporting. Review papers, editorials, book chapters, conference abstracts or proceedings, randomised controlled trials and case reports were also excluded. Following the full text review of articles, qualitative studies and statistical papers exploring novel mathematical techniques to modelling disease surveillance data were added to the excluded study types.

Information sources

The following electronic databases were searched: Medline, Web of Science, Pubmed, Scopus, Science Direct, Biosis Previews, Open Grey, Proquest dissertations and theses. The searches were conducted on 23rd October 2018.

Search terms

The following search terms were used: ((population surveillance/ or public health surveillance/ or sentinel surveillance/ or surveillance .mp.) OR (syndromic surveillance.mp.) OR (attend*.mp.) OR (absenteeism/ or absen*.mp.) OR (registers.mp.)) AND ((school.mp. or Schools/) OR (school aged children.mp.) OR (school children.mp.)) AND ((Infectious disease.mp. or Communicable Diseases/) OR (Outbreaks.mp. or Disease Outbreaks/) OR (epidemics.mp. or EPIDEMICS/) OR (pandemics.mp. or PANDEMICS/) OR (bugs.mp.)). The search terms were piloted before use and combined using Boolean operators. The search terms were developed for use in MEDLINE. Where possible, the same terms were used in each database, but some adaptation or simplification was required to meet the search requirements of different databases (Appendix B). The terms were searched for within the title and abstracts of papers and, where possible, the keywords.

Study selection

References from each database were imported into Mendeley Reference Manager. Each reference list was first de-duplicated, before combining all references and conducting a further removal of duplicate references. Additional duplicates were removed by manual searching. Two independent reviewers (AD and JLH) then undertook screening of the article titles and abstracts, applying the agreed exclusion and inclusion criteria. Any discrepancies between the reviewers were discussed and consensus reached. Articles meeting the screening criteria underwent a full text review. This was conducted by two reviewers (AD and JPH) using agreed eligibility criteria. Consensus was reached between the reviewers about the final articles for inclusion. Reference lists of the included articles were searched to identify any additional relevant studies not identified as part of the original search strategy. Papers identified in this way underwent the same screening and full text review outlined above.

Data collection process and data items

Data extraction was performed using a standardised data extraction form (Appendix C). Where available, the following data items were extracted; year of publication, country, prospective or retrospective study, age group, school type, sample size, time period of data collection, organism/syndrome, purpose of surveillance (case ascertainment or outbreak detection), case or outbreak definition, primary outcome measure, description of surveillance system (including the specificity, timeliness and spatial-temporal level of data collected), comparator surveillance systems, absenteeism rates with 95% confidence intervals, correlation measures with p-values, and lead or lag times compared to other surveillance indicators.

Summary measures

The summary measures were descriptive of the school surveillance systems and the methods used within each study. Outcomes included estimates of absenteeism, correlation measures and lead/lag times. Due to a high level of heterogeneity, estimates could not be pooled between studies.

Synthesis of results

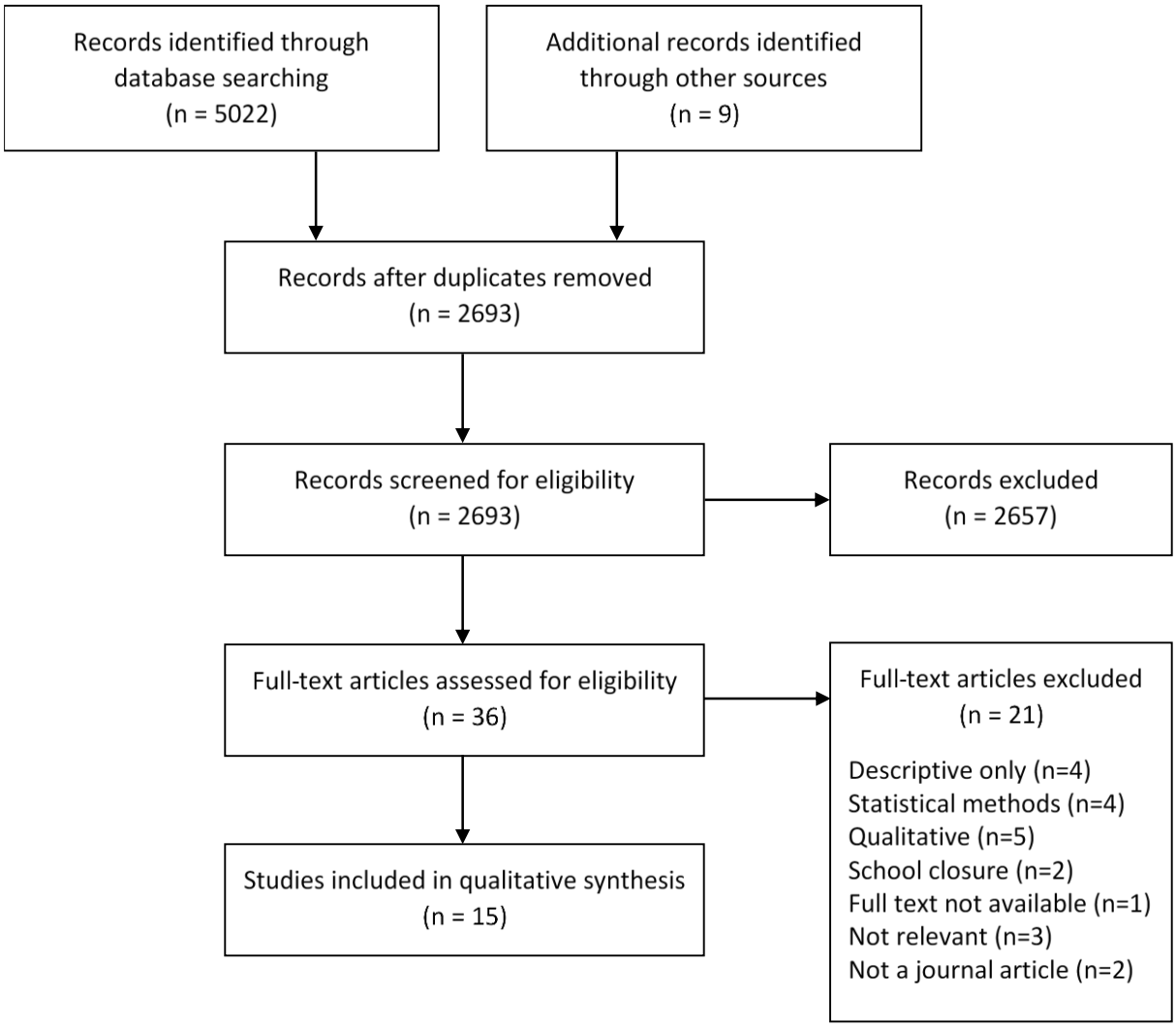
A narrative synthesis approach was adopted, comparing and contrasting the school-based surveillance systems in terms of their design, purpose, population, and performance against existing health surveillance systems.

3.4 Results

Study selection

The initial searches identified 5 022 references, which reduced to 2 684 once duplicates were removed. After screening the abstracts, 33 studies met the eligibility criteria for full text review. Of these, 14 were included in the systematic review. Nine additional studies were identified through searching the references of the papers for inclusion. Following abstract screening, three underwent full text review, one of which was subsequently included in the systematic review, giving a total of 15 studies (Figure 3.1).

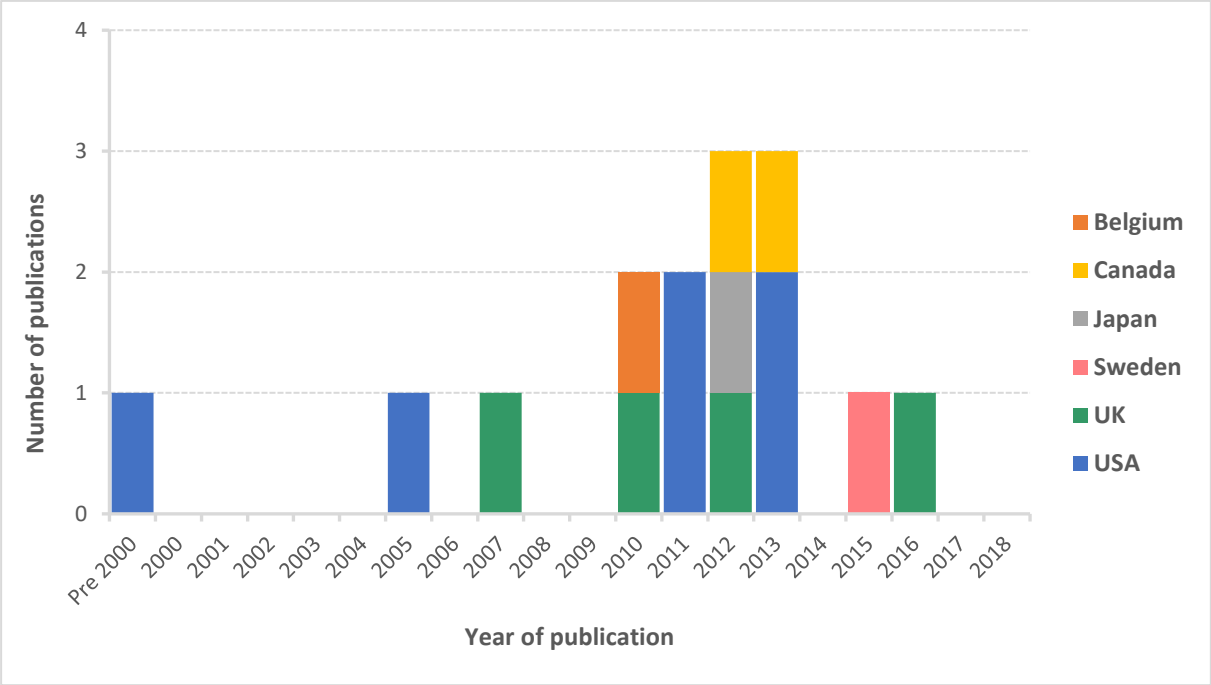
Figure 3.1: Flow diagram of study selection



Characteristics of included studies

All of the studies identified were concerned with the surveillance of influenza and over half were related to pandemic influenza. This is reflected in a peak of studies published between 2010 and 2013, following the H1N1 pandemic in 2009 (Figure 3.2). The greatest number of studies identified originated from the USA (n = 6), with multiple studies also reported from the UK (n=4) and Canada (n=2).

Figure 3.2: Year and country of publication of included studies



**Studies published pre-2000 comprised of one study published in 1995.*

A summary of the included studies is outlined in Table 3.1. Over half (9/15 studies) collected prospective data, the majority of which were during the H1N1 pandemic. Sample size varied from six schools to over 3 000 schools. Most studies included data on all school age groups, ranging from 3 to 18 years of age.

Table 3.1: Description of included studies

Author	Year of Publication	Country	Organism / syndrome	Prospective / retrospective	School ages	Sample size	Specificity of absence	Frequency of data submission
Aldridge et al	2016	UK	Seasonal influenza	Prospective	11-16yrs	27 schools	Medical [a]	Weekly
Besculides et al	2005	USA	Seasonal influenza	Retrospective	5-18yrs	1160 schools	All cause	Daily
Bollaerts et al	2010	Belgium	Pandemic influenza	Prospective	3-18yrs	~ 1 million pupils	Illness	Weekly
Chu et al	2013	Canada	Pandemic influenza	Retrospective	4-14yrs	8 PHUs [b]	Variable	Not specified
Crawford et al	2011	USA	Pandemic influenza	Prospective	5-12yrs	80 schools	All cause	Daily
Kara et al	2012	UK	Pandemic influenza	Retrospective	4-18yrs	373 schools	Illness	Weekly
Kightlinger et al	2013	USA	Pandemic influenza	Prospective	5-18yrs	187 schools	Illness	Weekly
Kom Mogto et al	2012	Canada	Pandemic influenza	Prospective	6-17yrs	3432 schools	Syndrome-specific	Daily
Lenaway et al	1995	USA	Seasonal influenza	Prospective	5-18yrs	44 schools	Illness	Weekly
Ma et al	2015	Sweden	Seasonal influenza	Retrospective	6-16yrs [c]	500 schools	Illness	Not specified
Mann et al	2011	USA	Pandemic influenza	Prospective	5-18yrs	349 schools	All cause	Daily
Mook et al	2007	UK	Seasonal influenza	Prospective	4-16yrs	11 schools	Illness	Daily
Schmidt et al	2010	UK	Seasonal influenza	Retrospective	5-11yrs	6 schools	Illness	Not specified
Suzue et al	2012	Japan	Pandemic influenza	Retrospective	3-18yrs	142 schools	Syndrome specific	Daily
Williams et al	2013	USA	Pandemic influenza	Prospective	5-17yrs	216 schools	All cause + syndrome specific	Weekly

[a] Medical absences include illness absence and absence to attend medical appointments.

[b] Public Health Units (PHUs) varied in size and were broadly divided into large PHUs (population >400 000) and small PHUs (population ≤400 000). Each PHU had a custom surveillance system to measure school absenteeism, collecting data on all cause absenteeism (8 PHUs), illness absenteeism (1 PHU) and respiratory illness absence (1 PHU) from schools within their area.

[c] Not clearly specified. The school ages noted are for compulsory education in the country of study.

Description of methods used for school-based surveillance

The three most common forms of absence data were all-cause absenteeism,[137,165,205,206] illness absenteeism,[164,167,168,207–210] and syndrome-specific absenteeism, which in these studies corresponded to influenza-like-illness (ILI) absences.[165,166,211] One paper reported medical absences, which combined both illness and planned medical appointments.[212] Another reported data from across multiple health authorities, each of which had a different system in place, varying between all-cause absence, illness absence and respiratory absence.[213]

The frequency of data submissions from schools varied between daily [137,166,205,206,210,211] and weekly reports.[164,165,167,207,208,212] Weekly reports often contained details of daily absences, so the frequency of reporting did not necessarily affect the level to which the data were analysed. Most studies analysed either daily or weekly absence rates but five studies used exceedances over a threshold as an indicator of a suspected outbreak.[137,165–167,213] One additional study used an absence threshold at city-level to determine the beginning and end points of the H1N1 influenza pandemic.[211] Outbreak definitions varied and are detailed in Table 3.2.

Table 3.2: Outbreak definitions used within included studies

First author & year of publication	Outbreak threshold / alert	Time period of breach
Chu 2013	Exceedance based on C2-MEDIUM method[a] OR >5% all-cause absenteeism[b]	Single day Single day
Kom Mogto 2012	≥10% ILI-related absenteeism	Single day
Lenaway 1995	>7.5% illness absence	Single week average
Mann 2011	≥8% all-cause absenteeism AND 1 SD above 30 day mean	Single day
Suzue 2012[c]	>2% ILI-related absenteeism	Single day
Williams 2013	>10% all-cause absenteeism >5% ILI-related absenteeism	2 or more consecutive school days

[a] C2-MEDIUM method calculates the mean and standard deviation (SD) from -9 to -3 days before the day of interest. Threshold is an exceedance of the expected value by three standard deviations.

[b] Not clearly stated, assumed from description of methods.

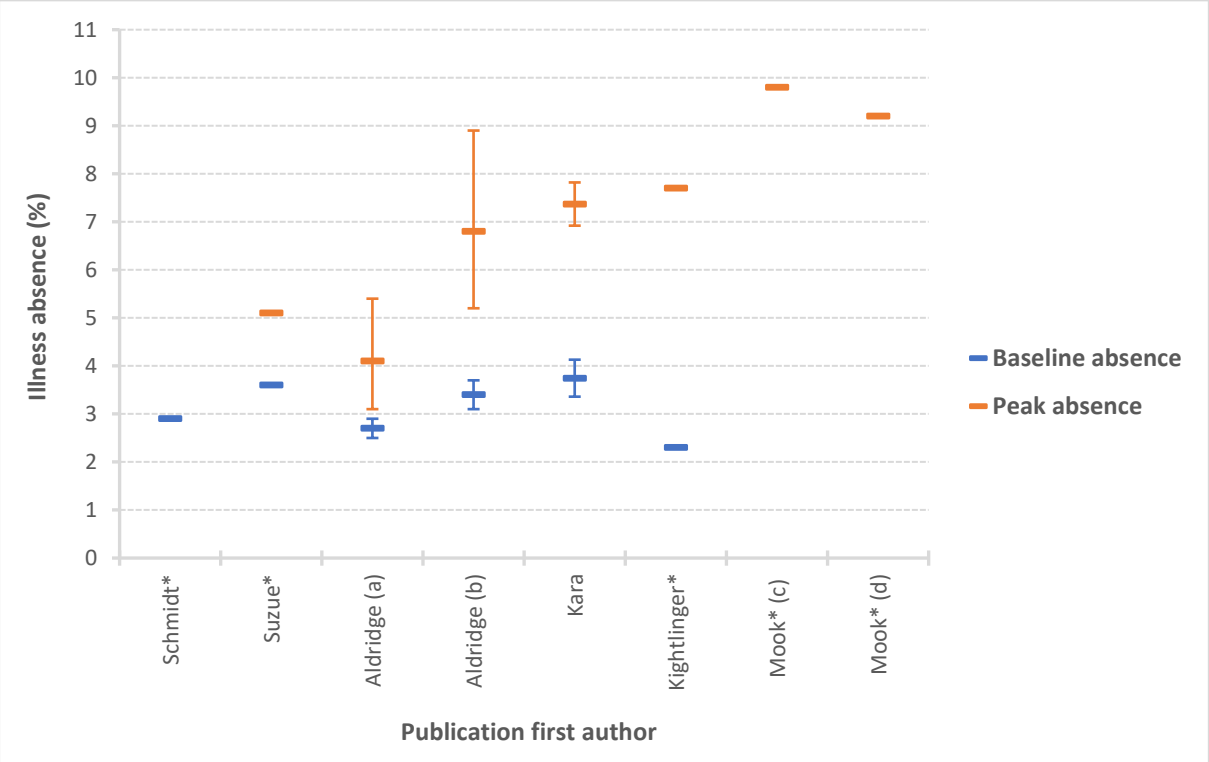
[c] Threshold used to detect start and end of pandemic.

The majority of studies aggregated absences across geographical areas or groups of schools, with only five studies considering absences at the individual school-level.[137,165–168]

Estimates of the burden of absenteeism

No standard measure of absenteeism was used across the included studies. Therefore, we were unable to conduct a pooled estimate of the impact of illness or influenza on school absences. Studies reported a mix of baseline absences, peak absences or both, either aggregated across all school age groups or by school type. Six studies reported rates of illness absenteeism (Figure 3.3). Baseline illness absences varied from 2.3% to 3.7%,[168,207,208,211,212] with peak illness absence ranging from 4.1% to 9.8%.[207,208,210–212] Two studies reported all-cause absenteeism rates, with results varying from 4.4% to 17.8%.[205,206] The higher reported values were for older children aged 14-18yrs. Four papers did not directly report on either the percentage of absenteeism or the number of exceedances identified, but instead reported only trends or correlations.[164,165,167,209]

Figure 3.3: Percentage of illness absenteeism at baseline and peak during influenza season or outbreak, with 95% confidence intervals



**No published confidence interval; (a) 2011/12 estimate; (b) 2012/13 estimate; (c) 4-11 year olds; (d) 11-16 year olds.*

Estimates of correlation with health surveillance indicators

The majority of studies used laboratory isolates as the conclusive marker of influenza activity. Other surveillance indicators used for comparison included primary care consultations, hospitalisations or emergency department attendances, telehealth calls and death certifications. The relationship between school absenteeism and established surveillance indicators was investigated by study authors using visual inspection and correlation coefficients. The measures of correlation varied and as a result it was not possible to generate a pooled estimate. Tests of correlation included Spearman Rank, Pearson's r, cross-correlation analysis and the coefficient of determination.

The correlation between all-cause absenteeism and other surveillance indicators was explored in four studies. Whilst visual inspection suggested that peaks in all-cause absenteeism coincided with community outbreaks,[205] correlations between laboratory reports and both all-cause absences and outbreaks based on >10% all-cause absence were low (r = 0.33* and r = 0.39* respectively, n = 216 schools).[165] In a study of 80 schools, all-cause absenteeism was not correlated with ILI emergency department visits during periods of low influenza activity (r_s = 0.23, p = 0.16), but there was evidence of correlation during periods of high influenza activity (r_s = 0.98, p = 0.05).[206] A study of outbreak alerts, based on all-cause absenteeism at 349 schools, generated a high number of alerts (one quarter of schools over a 6 week period), only 10% of which were subsequently confirmed as influenza.[137]

Two studies explored the correlation between syndrome-specific absenteeism and other surveillance indicators. Based on data from over 3 400 schools, strong correlations were reported between the number of schools who reported >10% ILI-related absence and both laboratory isolates and influenza hospitalisations (r_s = 0.90, p < 0.02 and r_s = 0.83, p = 0.01 respectively).[166] Amongst a smaller number of schools (n = 216), there was evidence of correlation between the lower threshold of >5% ILI-related absence and laboratory isolates, but the correlation coefficient was reduced (r = 0.78*).[165] The highest reported correlations were between laboratory isolates and ILI-absence rates (r = 0.92*), which increased when ILI absences were shifted back by one week (r = 0.97*), suggesting that trends in ILI absences preceded laboratory reports by one week.[165]

Studies exploring illness absenteeism presented mixed results. Based on visual inspection, study authors concluded that the peaks of illness absenteeism preceded or were concurrent with peaks in other surveillance systems across influenza seasons.[164,167,210] However, correlation with laboratory data varied between no correlation (n = 373 schools),[207] mild to moderate correlation (r

* P-value not reported within study paper.

= 0.11-0.45* and cross-correlation = 0.52, $p < 0.001$, n = 500 and 6 schools respectively),[168,209] and strong correlation (r = 0.9, $p < 0.01$, n = 187 schools).[208] The study of 187 schools also reported correlations with ILI hospitalisations (r = 0.9, $p < 0.01$) and ILI-related deaths (r = 0.7, $p < 0.01$).[208] Associations with primary care data ranged from moderate positive correlations to negative correlations (r = -0.19-0.47*, n = 373-500 schools),[207,209] and no association was found with telehealth calls.[207] In a study exploring absences at 27 schools, linear regression modelling identified statistically significant associations between medical absences, which include planned appointments, and both primary care data and laboratory reports ($r^2 = 0.42$, $p < 0.001$ and $r^2 = 0.27$, $p < 0.001$ respectively).[212] The association with primary care ILI reports was strengthened when this surveillance indicator was limited to children aged 5-14 ($r^2 = 0.62$, $p < 0.001$).

Lead and lag times

Thirteen studies considered the lead or lag time of school absence data compared to other surveillance indicators. All-cause absenteeism was not found to contribute significantly in terms of timeliness, with the majority of peaks occurring after other surveillance systems,[205,213] and multiple peaks observed which were unrelated to influenza activity.[165]

Illness absence presented a mixed picture, with the timeliness of peaks varying between no lead or lag time,[167,168,207,210] a 1-4 week lead time,[164,167,209,210,212] and a lag time of 1-11 weeks.[208,209] Syndrome-specific absences peaked concurrently or 1-2 weeks before other surveillance indicators,[165,166] and provided lead time on the start, peak and end point of the H1N1 pandemic (5 day, 10 day and 17 day lead time respectively).[211]

3.5 Discussion

This systematic review identified fifteen papers which explored the utility of school attendance registers in the syndromic surveillance of infectious disease. All of the papers identified were concerned with influenza, either pandemic or seasonal. There was a particular cluster of papers published following the 2009 H1N1 pandemic, indicating the heightened need for community-based surveillance systems during the pandemic. None of the papers we identified considered other common infectious diseases, such as diarrhoea and vomiting.

* P-value not reported within study paper.

The specificity of the data collected varied between all-cause absenteeism, illness absenteeism and syndrome-specific (in this case ILI) absenteeism. Syndrome-specific absenteeism had the strongest correlation with other surveillance systems, with illness absenteeism generating mixed results and all-cause absenteeism performing the least well. A similar pattern of results emerged in terms of lead and lag times, with ILI-specific absence providing a 1-2 week lead time, compared to lag times reported for all-cause absence data and inconsistent results for illness absence data. These results would indicate a potential role for syndrome-specific absences in the surveillance of influenza. However, all three studies which utilised syndrome-specific absence were conducted during the H1N1 pandemic and therefore the results presented may not reflect the performance of these data in non-pandemic situations. It should also be considered whether a two week lead time is sufficient warning to allow additional protective measures to be put in place.

The three studies which used syndrome-specific data also utilised absence thresholds, which were used to trigger alerts at school or city level. The thresholds used were >2%, >5% and ≥10% ILI-related absenteeism. Whilst the ≥10% threshold provided the strongest correlation with other surveillance indicators, it provided less lead time than the >5% and >2% thresholds. The scarcity of papers in this area makes it difficult to explore this further, but such thresholds inevitably result in a trade-off between accuracy and timeliness. Absence thresholds may also be influenced by health protection strategies targeted at children, such as vaccination schemes. Such interventions would be expected to reduce peak absence rates and consequently lower thresholds may be required to trigger alerts.

The development of absence thresholds requires an understanding of baseline rates of absence and these have been found to vary by age group. All-cause absenteeism was highest in older children, which could represent absences from causes other than illness. In contrast, both illness absence and syndrome-specific absence appeared higher in younger children.[166,207,210] There was some indication that influenza started and peaked earlier in younger children,[207,210] with high schools being affected later.[211] This is consistent with evidence that young children may be the first affected by seasonal and pandemic diseases,[200–202] and highlights the potential value in monitoring infectious illness in elementary/primary school children as an early warning of circulating infections.

As the potential lead time of school absence data was 1-2 weeks, the frequency of data submissions from schools is important in ensuring the early warning is optimised. Whilst the frequency of data reports from schools did not appear to affect correlation with other surveillance systems, the reported 5 day lead time on the start of the H1N1 pandemic may not have provided advanced warning had the data been transferred weekly as oppose to daily. If absence data were utilised to detect and manage

outbreaks at the individual school level, daily data submissions would confer additional benefit over weekly reports and aid in the more timely management of localised outbreaks.

Limitations of school absence data

In the studies identified there was variation in the type of school data used, both between countries and across different health authorities within countries. This makes aggregation of absence data across large areas difficult,[214] and could limit the utility of such data at a national level. School holidays result in a natural break in school attendance data, which is problematic for its use in tracking ongoing community outbreaks. There are also multiple factors which can affect school attendance, making its use in surveillance challenging. All-cause absences will not only capture illness but also unauthorised absences, and has been shown to increase around school holidays.[205] Illness absence will be affected by other infections, such as diarrhoea and vomiting, and has also been found to vary by day of the week.[168] This may contribute to the lack of correlation observed with all-cause and illness absence data, especially during periods of low influenza activity. Increases in school absences may also be affected by media coverage of pandemics or high profile deaths in children,[207,213] potentially driven by parental concerns of children catching illnesses at school, or lowering their threshold for keeping a child at home if they are unwell. Whilst this is more likely to be an issue in pandemic influenza, which receives significant media coverage, any high-profile outbreak is likely to create the same effect, regardless of the underlying organism.

3.6 Conclusion

The evidence of the utility of school attendance registers in the surveillance of infectious illness in children is limited to studies concerned with influenza. Therefore, the findings of this review may not be applicable to other conditions, such as diarrhoea and vomiting. There is a high level of heterogeneity between studies, making it impractical to pool results and generate a meaningful estimate of either burden of illness absenteeism or its correlation with other surveillance indicators. However, the studies identified suggest good correlation between syndrome-specific absences and healthcare surveillance data, with a potential lead time especially from absences in younger school age groups. Further research should consider the utility of school attendance registers for conditions other than influenza, to broaden our understanding of the potential application of these data for infectious disease surveillance in children.

Chapter 4. School attendance registers for the surveillance of IID in children

School attendance registers for the syndromic surveillance of infectious intestinal disease in UK children: protocol for a retrospective analysis

Anna Donaldson, John Harris, Roberto Vivancos, Dan Hungerford, Ian Hall, Sarah O'Brien

Protocol published: JMIR Research Protocols. 2022; 11(1): e30078

doi: 10.2196/30078

How this publication fits into my thesis

The systematic review in the previous chapter identified the potential utility of school attendance registers in the syndromic surveillance of influenza, with syndrome-specific absences correlating well with traditional surveillance indicators and providing a lead time of up to two weeks. There was also some evidence that younger children were affected before older children. This builds on the findings of Chapter 2, highlighting the importance of monitoring infectious disease in primary-aged school children not only in terms of outbreak risk but also as a potential early warning of infections before they start circulating in older age groups. However, the systematic review only identified studies relating to the surveillance of influenza and yet Chapter 2 found that IID accounted for nearly half of all reported school outbreaks. Consequently, it should be considered whether school attendance registers could also have a role in the surveillance of IID in schools and school-aged children.

This chapter explores the role and utility of school attendance registers in the surveillance of IID, using retrospective data on school absences in Merseyside, UK as well as routine health surveillance data from primary care, laboratories and telehealth services. The analysis was intended to model spatial and temporal variations in the incidence of IID within three Local Government areas in Merseyside and to apportion likely cause to changes in school absenteeism trends. This would have allowed an assessment to be made of the burden of IID absenteeism in schools and the potential utility and lead time of school attendance data in the surveillance of IID. The impact of the rotavirus vaccine, which was introduced in the UK in 2013, was also going to be examined, using an interrupted time series analysis to explore changes in school absenteeism rates pre- and post- the introduction of the vaccine. As none of the school-aged children included in this study would have received the rotavirus vaccine, this would capture the impact of vaccinating infants on herd immunity and reducing illness absenteeism amongst older, unvaccinated children. Unfortunately, school attendance data were only provided by one Local Government area within Merseyside and for a single academic year which fell outside of the study period. Consequently, the rotavirus-specific modelling could not be undertaken in this study. In this chapter, school absenteeism data are analysed using descriptive methods and univariate Quasi-Poisson regression models are used to explore the association between seasonal trends in illness absenteeism and surveillance data, based on aggregated weekly data from across the study period.

Due to the limited data received from Local Government, an amendment was made to the original study protocol and ethics application to allow data collection directly from schools. A survey was sent out to teachers to explore what information they collected regarding illness absence and how they used that data to manage illness within their school. A descriptive analysis of the survey results will

improve understanding of what level of data is collected informally by schools and the potential application of these data in managing and improving child health in school settings. At the end of the survey, schools were invited to take part in further research. Those which consented to participate were approached to share anonymised attendance data, with a particular focus on identifying schools which collected and recorded symptom-specific absence data. The survey, together with Local Government data, inform an assessment of the potential utility of school attendance data in the surveillance of IID and the burden of IID on illness absence in schools.

Due to the necessary changes made to the methodology, the original published study protocol paper has been amended prior to inclusion in this chapter. The data analysis section has been altered to reflect the descriptive and statistical analyses which were undertaken, and an additional section has been included which details the methodological approach for the school survey. The results and discussion sections of the protocol manuscript have been removed to allow the results of the study to be presented along with a discussion of the findings. The original published protocol paper can be found in Appendix D.

My contribution

All authors contributed to the study design and IH contributed to the statistical methods outlined within the original protocol paper. I led on the ethics application and wrote the first draft of the protocol manuscript. I subsequently designed the survey for schools and sought ethics approval for the amendment to the protocol. Prior to ethics approval, the survey was submitted to a Patient and Public Involvement (PPI) panel for review and comment. A Public Engagement and Science Communications Officer provided the link with the STEM Learning network, who then distributed the survey to schools. Following data collection, I undertook all aspects of the data cleaning, analysis, and final write-up of the results.

4.1 Summary

Background

Infectious intestinal disease (IID) is common and children are more likely than adults both to suffer from IID and to transmit infection onto others. Public health surveillance of IID is primarily based on healthcare data and as such illness which is managed within the community will often go undetected. This study explores the role and utility of school attendance registers in the detection and surveillance of IID in children.

Methods

Weekly school illness absenteeism data were retrospectively requested from Local Government areas within Merseyside, UK from July 2007 to June 2016. A survey was then distributed to teachers in Merseyside to identify schools which recorded symptoms of illness absence. These schools were invited to share anonymised symptom-specific illness absence data to further inform the analysis. Weekly surveillance data from primary care, laboratory and telehealth services were used to capture temporal variations in the incidence of IID. Rates of illness absence were compared to routine health surveillance data to explore whether school absenteeism data captured temporal trends in confirmed and probable cases of IID in children. Quasi-Poisson modelling was used to explore the association between weekly illness absenteeism counts and trends in the other surveillance datasets.

Results

School illness absenteeism data were received from one Local Government area for a single academic year, which fell outside of the study period. Consequently, aggregated data were used to compare seasonal trends between the datasets. The median weekly rate of illness absence was 2.49%, with a peak illness absence of 3.78% which occurred in calendar week 49. There was a positive association between school absences and viral IID (RR 1.083, p=0.015) but negative associations with bacterial and protozoal IID. School illness absence was also positively associated with general practice consultations and telehealth calls for IID (RR 1.029 and RR 1.033 respectively, p<0.005). Ten schools responded to the survey, most of whom collected their own data on symptoms of illness. However, only two schools recorded these data electronically. No schools consented to share symptom-specific absence data.

Conclusion

School illness absenteeism demonstrated a positive association with trends in general practice consultations, telehealth calls and laboratory reports for viral IID pathogens. Future work should consider the feasibility of collecting syndrome-specific absence data from schools to distinguish between absences caused by viral IID and respiratory pathogens, which circulate concurrently over winter months.

4.2 Introduction

Infectious intestinal diseases (IID) are common in both high and low income countries, causing an estimated 2 billion cases globally each year.[215] Norovirus is the leading cause of IID, with *Campylobacter* the most common bacterial cause.[35,37,215] In children, rotavirus has been a major cause of severe IID until the licensing of the vaccine in 2006.[46] The high incidence of IID infection results in significant disease burden and economic costs due to work and school absenteeism, lost earnings, reduced workforce productivity and increased healthcare usage.[29,45,216] In the UK alone, IID has been estimated to result in one million additional general practice consultations each year,[29] and norovirus, rotavirus and *Campylobacter* combined cost the UK economy an estimated £150 million per annum.[45] Over 80% of total costs are borne by patients, driven by lost income and out-of-pocket expenses.[45]

Children are disproportionately affected by IID, with those under 5 years accounting for 38% of foodborne cases globally.[215] Children are thought to be important transmitters of IID infection and experience prolonged symptoms and viral shedding, reduced immunity and higher levels of infectiousness.[23,217–220] The majority of a child's close contacts are based at school and home [74,81] and infections, especially viruses, can spread easily through these semi-enclosed environments.[65] This not only increases the risk of outbreaks within school settings, but also provides a pathway through which infections can spread from schools into the wider community.[74,75,87] There is evidence that children may be the first affected by seasonal and pandemic disease,[100,200–202] and hence enhancing infectious disease surveillance in schools could not only improve the health of children, but could also provide advanced warning before infections start to circulate in the wider community.

Public health surveillance of IID is primarily based on healthcare data such as laboratory reports, statutory notifications, hospital admissions, primary care consultations and calls to remote telehealth services.[140,221] The majority of IID cases, however, will be managed in the community without involvement from healthcare services. As a result, current surveillance is likely to be significantly underestimating the impact of IID. Furthermore, there is an inherent bias in the surveillance of IID, as certain groups are more susceptible to complications and therefore more likely to present to healthcare, such as the very young, the comorbid and the elderly.[35,37,222,223] Laboratory testing policies can also be targeted towards detecting pathogens in these high risk groups,[145] further increasing the surveillance bias. Enhancing the surveillance of IID and improving detection of community cases of disease would provide important information on the epidemiology of these

infections. Such data would be of value to support the evaluation of public health interventions, such as rotavirus vaccination and, in time, norovirus vaccination. As vaccinations can alter the epidemiology of infection,[224] it is crucial we are able to accurately monitor the long-term impact and effectiveness of these interventions, not just on healthcare services but also prevalence in the community.

School attendance registers offer a novel dataset which could be used to identify community cases of IID which might not otherwise be detected. School absenteeism data have shown potential in the surveillance of both seasonal and pandemic influenza,[166–168,205,207–209,213] but no studies have been identified which consider their role in monitoring IID. Whilst mild cases of diarrhoea and vomiting will not necessitate contact with healthcare services, they are still likely to result in an absence from school for the duration of the illness and, in line with public health guidance, an additional 48hrs after symptoms have resolved.[120] This provides a routine dataset which has the potential to capture illness from the day of onset.

This study explores the role and utility of school attendance registers in the detection and surveillance of IID in children. The secondary aims are to estimate the burden of IID on school absenteeism and to assess the impact of the rotavirus vaccine on illness absence amongst school-aged children.

4.3 Methods

Study setting

The study took place in Local Government areas within Merseyside in the North West of England. Merseyside is a predominately urban, metropolitan county with a population of 1.38 million, over 240 000 of whom are school-aged children.[225] It comprises five Local Government areas, which range in size from 145 000 residents to over 450 000 residents.[225] For this study, the population of interest was children aged 4 to 16 years who were registered at a school within Merseyside.

Study design

This study was a retrospective analysis of school absenteeism data to investigate whether school attendance registers could be used to capture seasonal trends and outbreaks of infectious intestinal disease amongst school-aged children. Whilst these data were routinely collected by Local Government for school attendance management,[134] this was a novel application of this dataset. In the United Kingdom (UK), all absences due to illness are given a single code, which distinguishes them from absences due to other causes, including those to attend medical appointments. However, the nature of the illness is not reported. Routine health surveillance data from primary care, laboratories

and telehealth services were used to capture temporal variations in the incidence of IID and to apportion likely cause to changes in school absenteeism trends.

Data sources

School absenteeism data were available at individual school level. Attendance data for schools providing primary (4-11 year olds) and secondary (11-16 year olds) education, regardless of type of school, were sought from Local Government in Merseyside, with data broken down by school and year group. Total absences and absences due to illness were requested. Details of the number of children in each school and year group were also requested to allow corresponding rates to be calculated.

Laboratory data reported to Public Health England (PHE) North West were used to obtain organism-specific rates of IID. These data are routinely collected and reported to Public Health England from diagnostic and reference laboratories.[226] Public Health England also holds data from NHS 111, which is a telehealth service that operates across England.[227] Calls to NHS 111 (and its precursor, NHS Direct) for diarrhoea and/or vomiting were used to indicate probable cases of IID. NHS 111 and NHS Direct data are held securely by the Public Health England Real-time Syndromic Surveillance team (ReSST) and were accessed with permission via PHE.

Primary care consultations for diarrhoea and/or vomiting were used as another indicator of probable IID. These data were previously collected from clinical commissioning groups and general practices across Merseyside to inform an evaluation of the rotavirus vaccine.[228] Read Codes were used to distinguish acute cases of IID from cases linked to chronic conditions or non-infective causes.[228] These data were accessed from the University of Liverpool in an anonymised format as a secondary dataset to further inform the evaluation of the rotavirus vaccine.

Study period

Data were requested retrospectively from July 2007 to June 2016, capturing nine IID seasons. Each season was considered to start in calendar week 27 and end in calendar week 26 of the following year.

Population sample

This study focused on three of the five Local Government areas within Merseyside in order to reflect the coverage of primary care data collected to inform an evaluation of the rotavirus vaccine.[229] Prior to data collection, the population sample was estimated using data from the Department for Education, which holds a record of all Local Government registered schools.[173] Data were based on

the 2017/2018 academic year, limited to schools providing primary and secondary education (ages 4-16 years).

The total number of schools across the three Local Government areas was 372, consisting of 299 primary schools and 103 secondary schools. Thirty of the schools delivered both primary and secondary education. The total pupil population across all included schools was 140 164. Assuming that each year one in four pupils are affected by IID,[22] in each academic year we estimated there would be approximately 35 000 cases of IID in schools within the study area. As data were requested over a 9-year period, the total number of cases across the study period was estimated to be 315 000.

Case definitions

The case definitions used within each dataset are outlined in Box 4.1.

Box 4.1: Case definitions

School attendance registers

- Absence with registration code 'I' (Illness, not medical or dental appointments)

Telehealth calls (NHS 111 and NHS Direct)

- Calls for vomiting
- Calls for diarrhoea

General Practice consultations (Read Codes in parenthesis)

- Diarrhoea and vomiting (19G)
- Diarrhoea symptom NOS (19F6)
- Viral gastroenteritis (A07y0)
- Diarrhoea (19F2)
- Gastroenteritis—presumed infectious origin (A0812)
- Diarrhoea of presumed infectious origin (A083)
- Infantile viral gastroenteritis (A07y1)
- Infectious gastroenteritis (A0803)
- Enteritis due to rotavirus (A0762)
- Infectious diarrhoea (A082)

Laboratory detections

- Detection of bacterial, viral or protozoal IID organisms in a faecal specimen

Recruitment and consent

Recruitment was conducted at Local Government level. Local Government were approached via their public health departments and invited to participate in the study. Consent for use of aggregated school attendance data was sought from the Local Government, who carry the legal responsibility for the data and its usage. As the data were aggregated and anonymised, consent was not sought from individual schools or parents.

Data analysis

Weekly-level data were analysed according to the norovirus season week, with the year considered to start in calendar week 27 and end in calendar week 26 of the following year. Each dataset was described in terms of the number and proportion of absences/IID cases across three different school age groups: preschool, primary school, and secondary school. As the age range of schools overlap, the year group of each case at the time of illness was calculated. Each case was then assigned to the appropriate school category with primary schools capturing reception to year 6 (ages 4-11 years) and secondary schools capturing year groups 7 to 12 (ages 11-16 years). Children too young to attend compulsory education were assigned to the preschool group.

A descriptive analysis was undertaken of each dataset to examine and describe the temporal trends and seasonality of illness absenteeism rates, and of confirmed and probable cases of IID. School illness absenteeism rates were broken down into primary, secondary, and all-schools to identify differences in illness absence trends between these settings. The weekly rate of illness absence was calculated at Local Government level. Attendance and absence numbers were provided for each school session missed, with two sessions per day and ten sessions across the school week. Therefore, a child who missed a single day of school would contribute two school absences. The maximum possible attendance for each week varied depending on the number of school sessions provided. The rate of illness absence was calculated by aggregating the total number of illness absences across all schools for each given week. The denominator used was the maximum possible attendance for that week.

Surveillance data were aggregated by norovirus season week. Weekly rates of general practice (GP) consultations, telehealth calls and laboratory reports were calculated per person-years using the Office for National Statistics (ONS) mid-year population estimates for 2012,[230] which represented the mid-point of the study period. These were plotted against the weekly rates of illness absences to explore whether school illness absenteeism data captured temporal trends in confirmed and probable cases of IID. The number of individual pathogens across the study period were too few to allow separate

analysis, so laboratory data were aggregated into bacterial, viral, and protozoal IID. National rates of norovirus laboratory reports for all age groups were plotted alongside viral IID rates. Adenovirus was excluded from the analysis as, in the absence of typing, it was not possible to distinguish IID from respiratory adenoviruses.

A univariable Quasi-Poisson regression model was used to explore the association between weekly illness absenteeism counts and trends in the other surveillance datasets. Quasi-Poisson was chosen instead of Poisson regression due to the high variance of the count data. The outcome variable within each model was school illness absenteeism. Data were analysed for all schools and for primary and secondary schools separately. Explanatory variables included GP consultations and telehealth calls for the corresponding age groups. Due to low case numbers, laboratory data were grouped into broad classes of pathogen (bacterial, protozoal and viral) before inclusion in the regression model and could not be broken down by age group. To investigate the impact of school holidays on the association between illness absences and IID surveillance data, the Quasi-Poisson modelling was repeated controlling for school holidays. Weeks of school holiday were based on existing literature.[231] All statistical analysis was undertaken in R 4.0.2.[176] Relative risks (RR) were calculated from the Quasi-Poisson regression models. Upper and lower 95% confidence intervals (CI) were calculated around each estimate.

Ethics approval

This study received ethical approval from the University of Liverpool Research Ethics Committee (Reference number 1819, Appendix E). Use of general practice data had been approved for the evaluation of rotavirus vaccination in Merseyside by NHS Research Ethics Committee, South Central-Berkshire REC Reference: 14/SC/1140.

Amendments to methodology

Due to unforeseen challenges in accessing and extracting weekly school attendance data from Local Government, an amendment was made to the original protocol and ethics application to pursue data collection directly from schools. The amended methodology is outlined in Table 4.1 and comprised of a school survey exploring whether schools collected their own symptom-specific absence data. The survey investigated how parents could inform school of a sickness absence, whether additional questions were asked regarding the cause of illness and if so, how that information was recorded and used by the school. Schools which recorded data on the symptoms of illness absence were invited to share anonymised symptom-specific absence data to further inform this study.

Table 4.1: Adapted methodology and rationale

Methodology section	Amendments	Rationale
Study design	A survey was sent out to schools in Merseyside to explore what information was collected regarding illness absence and how they used that information to manage illness within their school. At the end of the survey, schools were invited to take part in further research. Those which consented to participate were approached to share anonymised symptom-specific absence data to inform this study.	The survey was to improve our understanding of what level of data are collected informally by schools and the potential application of these data in managing and improving child health in school settings. Symptom-specific absence data, combined with data received from Local Government, were to inform an assessment of the overall burden of IID on illness absenteeism and the potential utility of school absenteeism data in the surveillance of IID.
Study setting	The survey was sent out to all schools in the Merseyside area and was not restricted to the three Local Government areas outlined in the original protocol.	Collecting data on the symptoms of illness removed the need for statistical modelling to identify the likely cause of illness-absence and therefore the school data did not need to cover the same geographical areas as the GP data collected to inform an evaluation of the rotavirus vaccine.
Data sources*	The survey was sent out to schools in Merseyside through STEM (Science, Technology, Engineering and Mathematics) Learning networks. The school survey is shown in Appendix F and included questions regarding what information, if any, was collected on the cause of illness absence and how these data were recorded and subsequently used by the school*. Schools which consent to share anonymised symptom-specific absence data will be asked to provide data broken down by week and year group, with details of the length of absence and symptoms associated with the illness.*	STEM Learning networks provide schools with teaching resources in Science, Technology, Engineering and Mathematics, and are linked with many schools in the area. Distributing the survey through these networks enabled a larger number of schools to be reached than would otherwise have been achievable in the time available. Prior to being distributed, the school survey was reviewed by a Public and Patient Involvement (PPI) panel to ensure readability, language, and clarity of the questions.

Study period*	*As a minimum, data will be requested for the previous complete academic year (1st September 2018 – 31st August 2019). More academic years will be requested if available, dating back to 2007.*	It was not known whether schools kept historical data on illness absences and therefore it may only have been possible for them to provide data for the previous academic year. If available, schools were asked to share data as far back as 2007, to mirror the data requested from Local Government, NHS 111, GPs, and laboratories.
Population sample	Not calculated.	As it was not known whether any schools collected information on cause of illness absence, it was not possible to estimate a sample size.
Case definitions	Absence with registration code 'I' (Illness, not medical or dental appointments).	The case definition remained unchanged, but any additional information recorded for each illness absence was also requested.
Recruitment and consent*	Schools were recruited via the survey and invited to leave contact details if they wished to take part in further research. *Schools that agreed to share symptom-specific absence data will be required to complete an Information Sharing Agreement. Prior to data transfer, cases will be assigned a unique case ID and all personal identifiable information (name, date of birth, age, postcode) will be removed by the school. In return for their participation, schools will be offered a workshop for pupils around the spread of infections and how to prevent them, along with lesson plans and work sheets.*	The Information Sharing Agreement was to detail the method and frequency of data extraction, the data fields to be extracted, and how the data would be transferred and managed. It was to include an agreement on data governance and security, outlining the responsibilities of both the researchers and the school in the handling of these data.
Data analysis*	The survey results were analysed quantitatively in terms of the number and proportion of responses for each question. This was used to describe what data was being collected by schools and how it was being utilised.	The analysis of school survey responses was to improve our understanding of how many schools collect additional data around illness absences and the potential benefit of these data in managing illness within school settings.

	Symptom-specific absence data will be categorised into broad groups (diarrhoea and vomiting, influenza-like illness, rash, other). These data will be analysed descriptively to identify the most common causes of illness absence in schools and the burden of IID on school absenteeism. Temporal variations in the different causes of illness absence will be examined. If possible, these will be compared to trends in surveillance datasets, to test the ability of illness absenteeism data to accurately detect seasonal variations in IID.	A descriptive analysis of symptom-specific illness absence data was to support an assessment of the burden of IID within schools. Further comparison with existing surveillance data was dependent on the timespan of the data received from each individual school.
Ethics approval	This amendment to the original protocol received ethical approval from the University of Liverpool Research Ethics Committee (Reference number 1819, Appendix G).	N/A

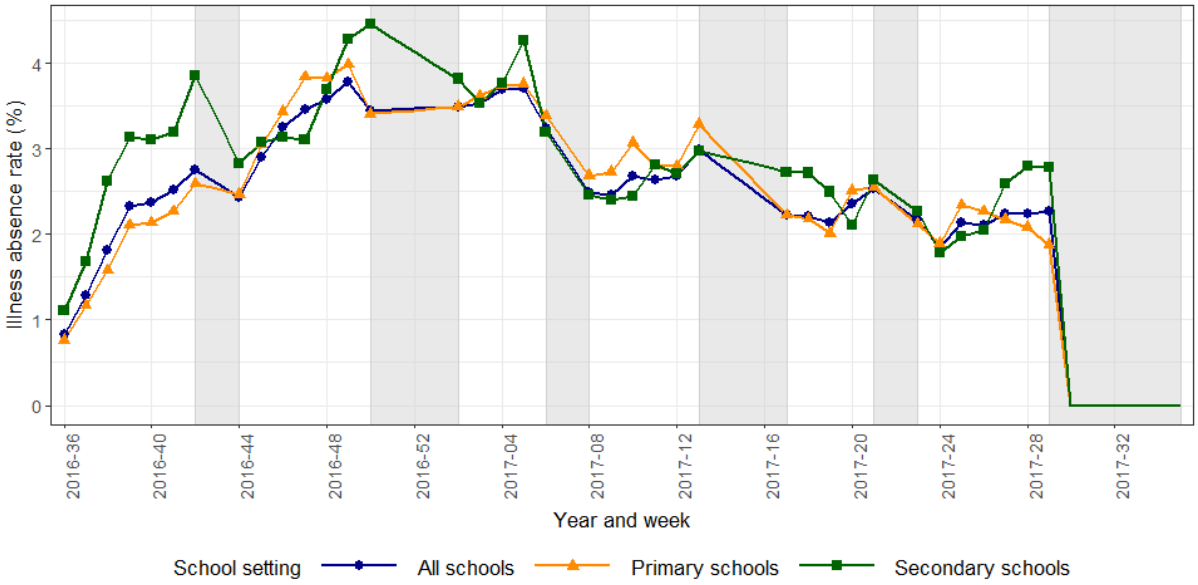
*Aspects of the methodology which were planned but could not subsequently be undertaken are shown in grey italics.

4.4 Results

The three Local Government areas approached all consented to participate in the study, however, attendance data was only received from one Local Government area for two academic years (2016/17 and 2017/18). Both years fell outside of the study period and as data for the 2017/18 academic year were incomplete, it was subsequently excluded from the analysis. School attendance data were provided for the 2016/17 academic year for 69 schools, one of which provided higher education for 16-18 year old pupils and was therefore excluded from the study. The 68 included schools covered a pupil population of 25 825 pupils and included 52 primary schools and 5 secondary schools. Eleven schools did not specify whether they provided primary or secondary education and included academies (n=6), special schools (n=3) and pupil referral units (PRUs) (n=2). These schools were included in the all-schools category, but excluded when data were broken down into primary and secondary age groups. Absence and illness absence numbers were provided for each school session missed, with two sessions per day and 10 sessions across the school week. A further breakdown of data by year group and sex could not be provided by the Local Government.

Across the 68 included schools there were a total of 253 153 missed sessions due to illness in the 2016/17 academic year. This equated to over 126 000 missed school days and 4.9 days of illness absence per pupil per year. Illness absenteeism demonstrated a seasonal trend, with absences increasing from September onwards and staying high during the winter months (Figure 4.1). Illness absence rates started declining again in the spring and plateaued throughout the summer term. Excluding school holidays, the median weekly rate of illness absence across the Local Government area for the 2016/17 academic year was 2.49%, with a peak rate of illness absence of 3.78% which occurred in calendar week 49. Secondary schools had a higher median weekly rate of illness absence compared to primary schools (2.79% and 2.54% respectively) and a higher peak absence (4.46% and 3.98% respectively) which occurred one week later than the peak in primary schools.

Figure 4.1: Weekly rates of illness absence by school setting (2016/17)



*Shaded areas represent school holidays. Missing data over the school holidays were fitted using linear interpolation. Absence rates for 'all schools' included schools where the phase of education was not known.

The number of IID cases reported, by surveillance indicator and age group, is shown in Table 4.2. GP consultations, telehealth calls, and laboratory reports were included for all three Local Government areas across nine surveillance years, with the exception of telehealth calls which captured six surveillance years. Amongst children aged 0-16 years, school children (4-16 years) represented 27% of GP consultations for IID and 19% of telehealth calls, with the largest burden of cases in preschool children (0-4 years). Cases in school children disproportionality affected primary school age groups (4-11 years) with significantly fewer cases reported in secondary-aged children (11-16 years). Whilst most laboratory cases (>70%) were in the preschool age group, there was variation between organisms. Almost 90% of all viral laboratory reports were in preschool children with those aged 11-16 years accounting for only 3%. In contrast, secondary-aged children represented nearly a third of C*lostridium* reports and over a fifth of *Cryptosporidium* and C*ampylobacter* confirmed cases.

Table 4.2: Number and proportion of cases reported across the study period (2007/08 – 2015/16 surveillance years), by surveillance indicator and school age group*

Surveillance dataset	All children	Preschool	Primary	Secondary
School absences[$]	253 153	-	147 508 (58%)	49 058 (19%)
GP consultations	30 402	22 059 (73%)	6 268 (21%)	2 075 (7%)
Telehealth calls^	14 193	11 510 (81%)	2 030 (14%)	653 (5%)
Laboratory samples (all)	2756	1966 (71%)	505 (18%)	285 (10%)
All bacteria	1304	769 (59%)	321 (25%)	214 (16%)
Campylobacter	624	313 (50%)	174 (28%)	137 (22%)
Clostridium	32	18 (56%)	4 (13%)	10 (31%)
E.coli	360	261 (73%)	64 (18%)	35 (10%)
Salmonella	288	177 (61%)	79 (27%)	32 (11%)
All protozoa	187	74 (40%)	80 (43%)	33 (18%)
Cryptosporidium	150	50 (33%)	69 (46%)	31 (21%)
Giardia	37	24 (65%)	11 (30%)	2 (5%)
All viruses	1265	1123 (89%)	104 (8%)	38 (3%)
Adenovirus	495	419 (85%)	51 (10%)	25 (5%)
Astrovirus	44	39 (89%)	5 (11%)	-
Norovirus	98	79 (81%)	13 (13%)	6 (6%)
Rotavirus	599	559 (93%)	33 (6%)	7 (1%)
Sapovirus	29	27 (93%)	2 (7%)	-

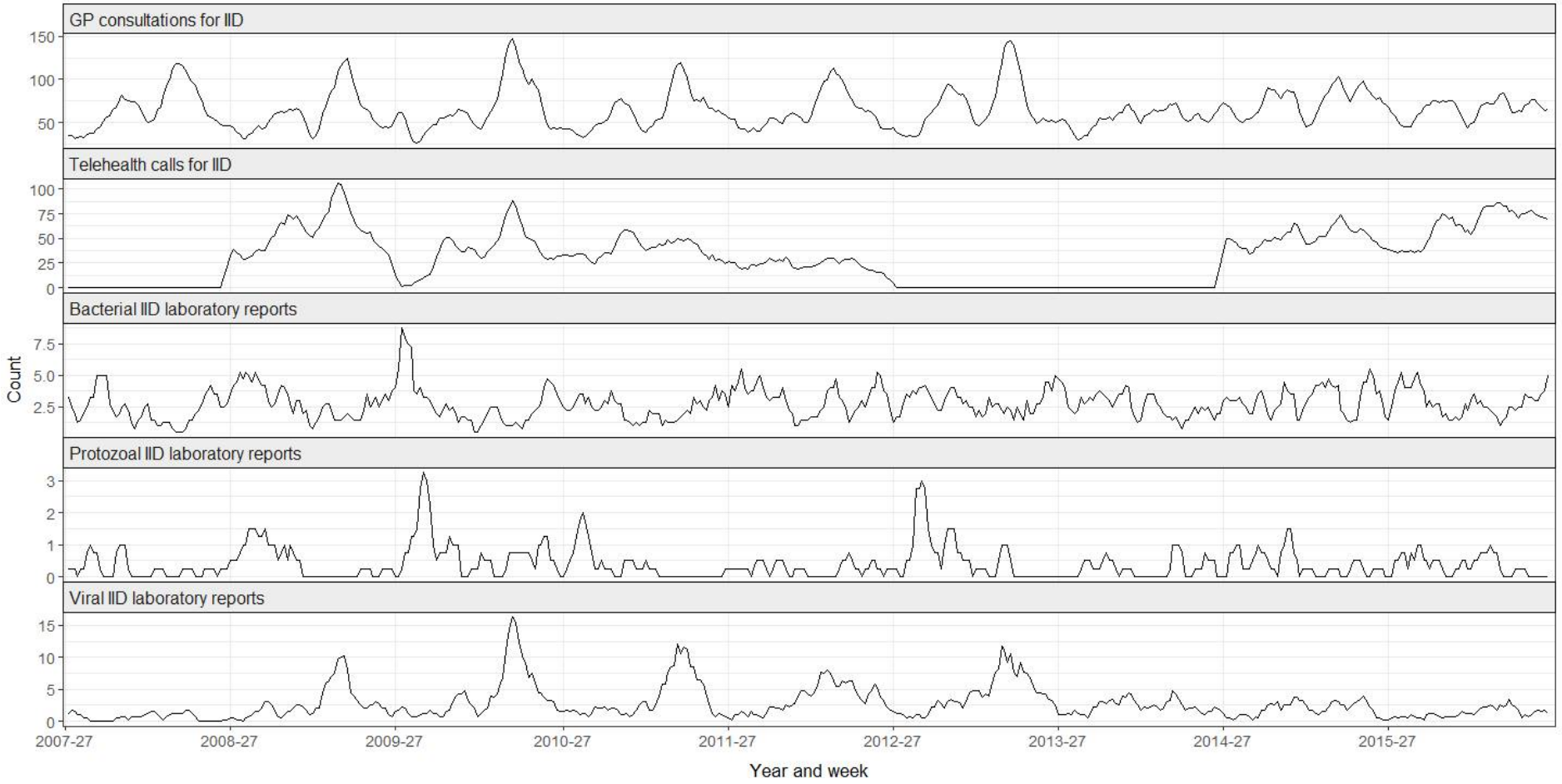
* 'All children' represent ages 0-16yrs; preschool included children younger than compulsory school age (0-4yrs); primary school included reception to year 6 (4-11yrs); secondary school included year groups 7-12 (11-16yrs).

[$] Data for 2016/17 academic year only. Absences for 'all children' included schools where the phase of education was not known.

^ Data for telehealth calls spanned six surveillance years: NHS Direct (2008/09-2011/12) and NHS 111 (2014/15–2015/16).

Trends in the surveillance datasets are shown in Figure 4.2. GP consultations for 0-16 year old children demonstrated a clear seasonal trend, with a small autumn peak followed by a larger spring peak occurring each year. The spring peaks coincided with peaks in viral laboratory reports and were less visible following the introduction of the rotavirus vaccine in 2013. Telehealth calls for IID also demonstrated a spring peak in two surveillance years, but thereafter no clear trends can be seen in the data. Bacterial and protozoal IID demonstrated some seasonality, with summer peaks in the early surveillance years, but the trend disappeared in the latter half of the study period.

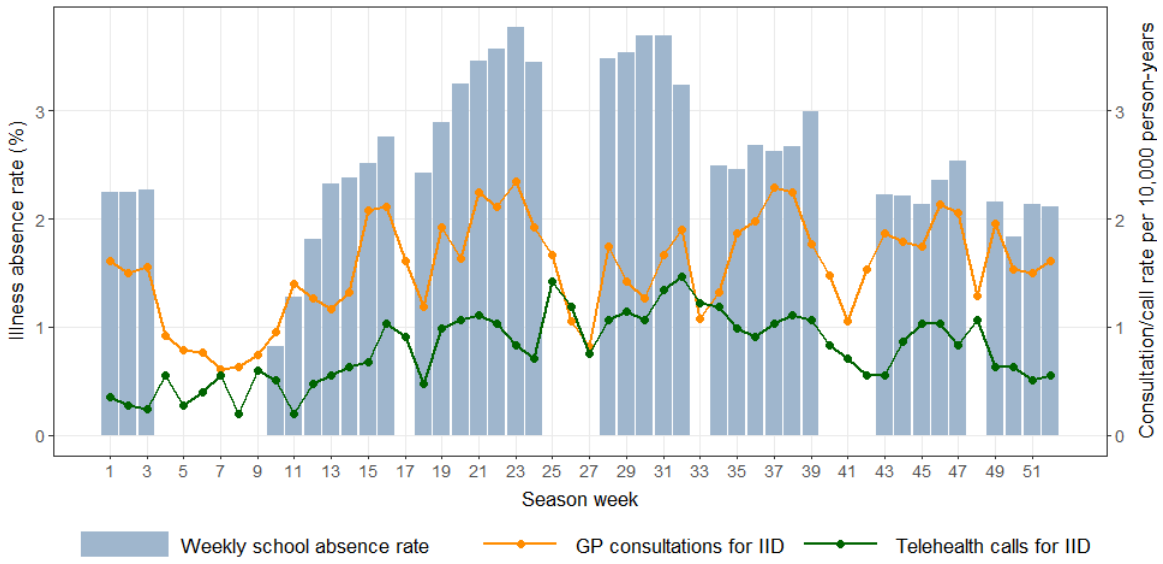
Figure 4.2: Time trends in IID surveillance datasets, based on a 4-week rolling average, for children aged 0-16 years



* Telehealth data were incomplete during the transition from NHS Direct to NHS 111, therefore no data were included for the 2012/13 and 2013/14 surveillance years.

Figure 4.3 shows the trend in the rate of illness absenteeism plotted against the rate of GP consultations and telehealth calls for school-aged children. Due to the changes in trend of the surveillance indicators following the introduction of the rotavirus vaccine, GP and telehealth data were only included from the 2013/14 surveillance year onwards, so that all three datasets captured illness trends in the post-rotavirus vaccine era only. GP consultations for school-aged children appeared to closely follow the trend in illness absenteeism, with peaks during school term time and notable troughs during school holidays. Like school illness absenteeism, GP consultations also peaked in season week 23, although the peak was not significantly larger than other peaks throughout the surveillance year. Telehealth calls peaked later, in week 32, although a spike in telehealth calls also occurred in week 25.

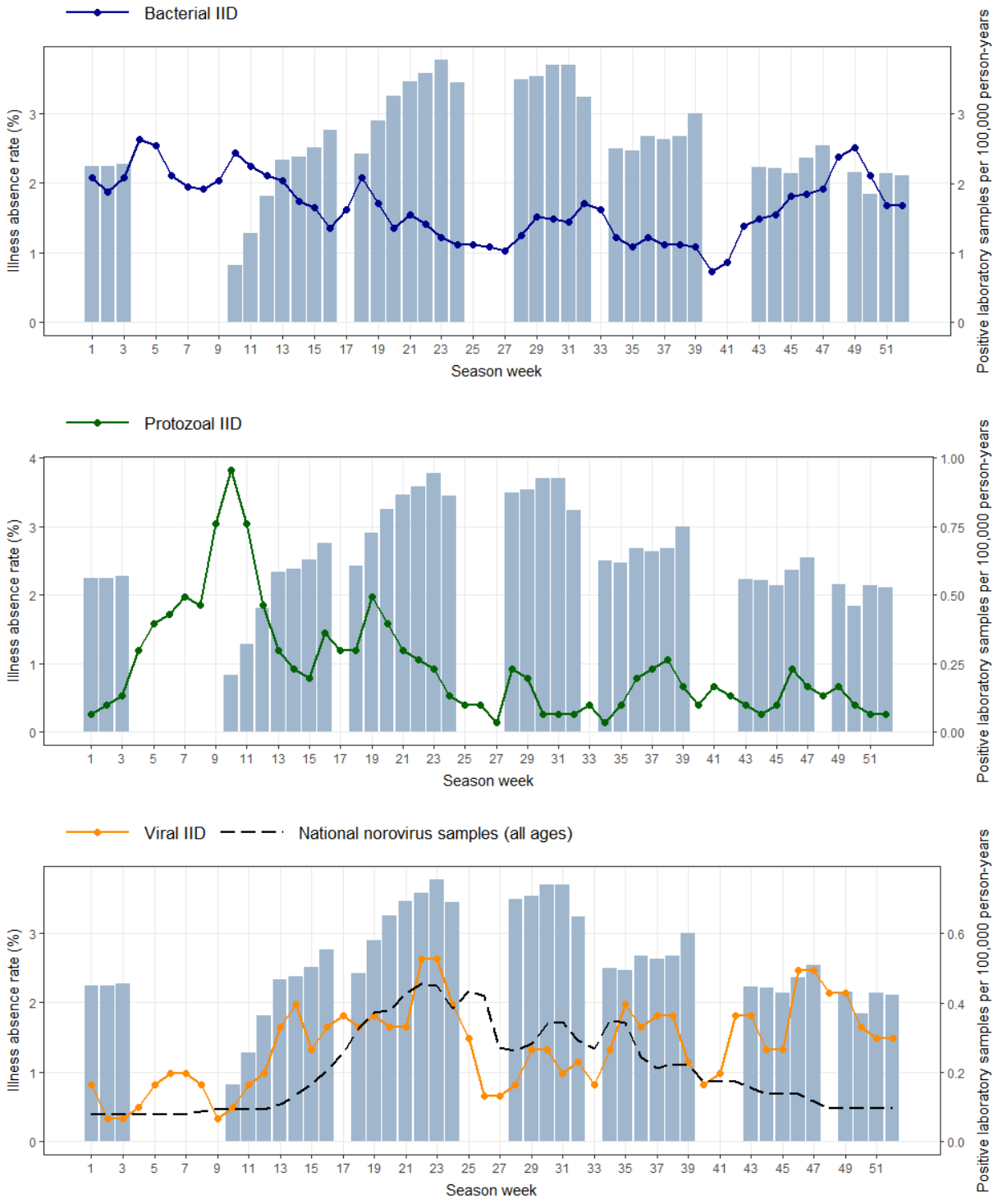
Figure 4.3: Weekly rates of school illness absence, general practice consultations and telehealth calls for IID, for children aged 4-16 years*



*School absence data was included for the 2016/17 academic year only. GP consultation data spanned three surveillance years (2013/14-2015/16) and data for telehealth calls (NHS 111) spanned two surveillance years (2014/15–2015/16).

When school illness absence rates were plotted against laboratory reports (Figure 4.4), there was no clear association between the trends in bacterial and protozoal IID and the rate of illness absence, with laboratory cases peaking as illness absence decreased over the summer months. In contrast, viral IID cases mirrored the rise in illness absenteeism throughout the autumn, but a later summer peak in viral IID was not reflected in the school data. Trends in school illness absenteeism appeared to correlate well with national rates of norovirus for all age groups for the 2016/17 surveillance year.

Figure 4.4: Weekly rates of school illness absence plotted against laboratory reports for all children (0-16yrs) and national norovirus reports for all ages, based on a 2-week rolling average*



*Laboratory data were aggregated across all nine surveillance years, except for rotavirus, which was aggregated for post-vaccine years only (2013/14 – 2015/16). National norovirus data were for all age groups for the 2016/17 surveillance year only.

Quasi-Poisson regression

In the Quasi-Poisson regression modelling, viral IID showed a positive association with school absences, with illness absence increasing by 8.3% for each additional positive viral laboratory report (*p*=0.015). In contrast, both bacterial and protozoal IID had negative associations with school absences (Table 4.3).

Table 4.3: Univariable Quasi-Poisson regression examining the association between the number of laboratory reports (0-16 years) and school illness absences, by season week

Organism	Coefficient	RR (95% CI)	p-value
All bacteria	-0.023	0.977 (0.955-1.000)	0.055
All protozoa	-0.052	0.949 (0.890-1.006)	0.103
All viruses	0.080	1.083 (1.016-1.150)	0.015

GP consultations demonstrated a statistically significant, positive association with school absences across all three age groups (Table 4.4), with illness absence increasing by 3.5% for primary-aged children and 7.3% for secondary-aged children, with each additional GP consultation. Similar associations were found for telehealth calls, with illness absence increasing by 3.3% for all school-aged children and 4.4% for primary-aged children, with each additional telehealth call. The association was not significant for secondary-aged children (*p*=0.559).

Table 4.4: Univariable Quasi-Poisson regression examining the association between school illness absences and IID surveillance data for GP consultations and telehealth calls, by school age group

Surveillance data	Coefficient	RR (95% CI)	p-value
All schools (4-16yrs)			
GP consultations	0.029	1.029 (1.019-1.039)	<0.001
Telehealth calls	0.032	1.033 (1.012-1.054)	0.003
Primary schools (4-11yrs)			
GP consultations	0.034	1.035 (1.021-1.048)	<0.001
Telehealth calls	0.044	1.044 (1.020-1.069)	<0.001
Secondary schools (11-16yrs)			
GP consultations	0.070	1.073 (1.043-1.103)	<0.001
Telehealth calls	0.022	1.022 (0.949-1.100)	0.559

*GP consultations included from the 2013/14 surveillance year onwards.

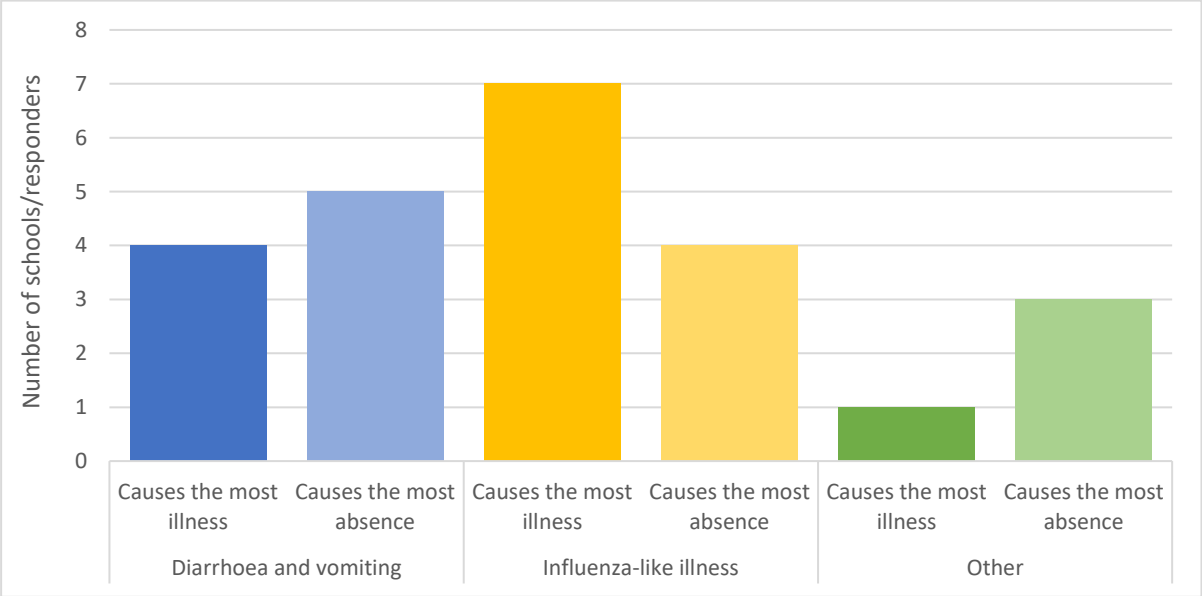
*Telehealth calls included from the 2014/15 surveillance year onwards.

Controlling for school holidays within the regression models did not significantly alter the association between telehealth calls and school illness absence across the three school categories. The coefficients for GP consultations were reduced but remained significant for primary schools and all schools, but not for secondary schools. The positive association between viral laboratory reports and school absences remained but was no longer statistically significant.

School survey

There were 41 survey submissions, only 10 of which were partially or fully complete. Six responses were received from primary schools and four from secondary schools. Schools ranged in size from less than 250 pupils to more than 1 000 pupils. Whilst more responders felt that influenza-like-illness (ILI) caused the most illness within their schools, diarrhoea and vomiting was more commonly highlighted as a cause of absence (Figure 4.5). Other causes of absence highlighted by secondary schools included non-communicable diseases such as mental health, migraines and chronic health conditions.

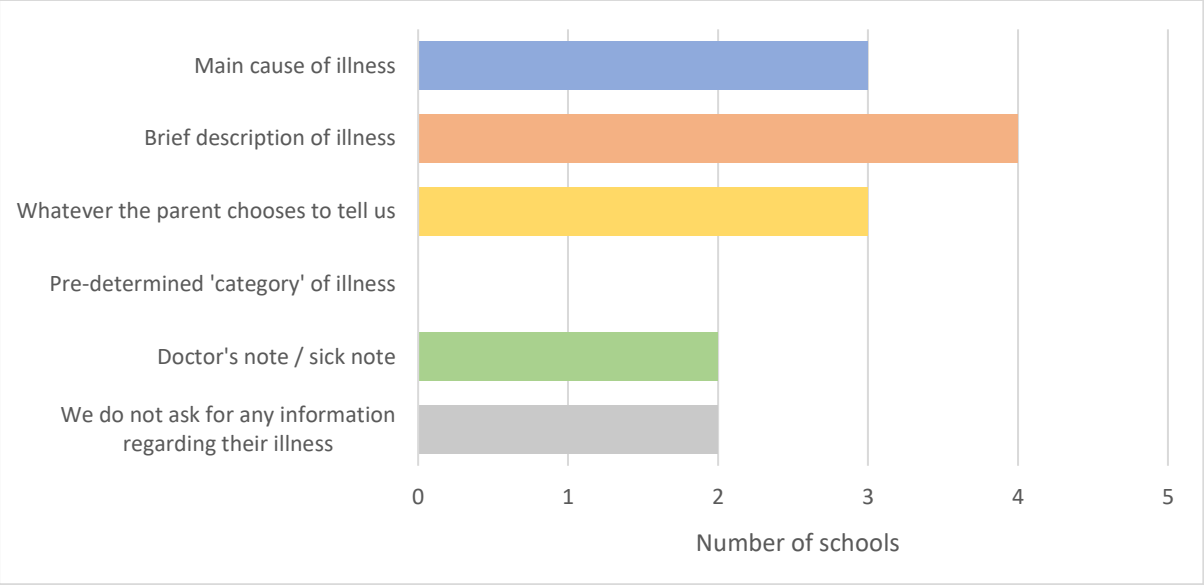
Figure 4.5: Causes of illness and illness absence in schools



Out of seven schools who responded to the question, five collected information from parents on the cause or symptoms of illness (71%). The type of information collected is shown in Figure 4.6. Only two schools said they recorded this information electronically, with other schools either keeping paper records or asking verbally without further documentation. The two schools which kept electronic records utilised the same system used for statutory attendance data. How these data were subsequently used varied between schools. Three schools used it to validate the reason for absence and to monitor illness within the school, including identifying individual pupils who were frequently

unwell. Two schools used symptom data to share relevant information with parents about different illnesses. None of the schools reported symptom data to the Local Government. Of the two schools which asked for information on symptoms but did not record it, one cited a lack of appropriate systems in place and limited teacher time as a reason for this. The other school, whilst they did not record the information, did use it to pick up patterns of illness and responded with hygiene messages to children and a deep clean of relevant areas of the school.

Figure 4.6: Information collected regarding cause or symptoms of illness



Out of seven schools, six felt it would be beneficial for schools to routinely collect information on symptoms of illness absences, citing reasons such as improving preventative measures, identifying trends and outbreaks, and informing parents about illnesses. The school that did not feel it would be beneficial noted that the information received from parents was highly selective. Interestingly, another school felt that collecting data on symptoms would encourage parents to be more honest regarding the cause of absence and ultimately improve attendance. None of the schools who collected information on the cause of illness agreed to take part in further research, so an analysis of symptom-specific absence data was not possible.

4.5 Discussion

Current surveillance of infectious intestinal disease is predominantly based on healthcare data and therefore illness which is managed within the community will often go undetected. This study is unique in considering whether school absenteeism data could be used to enhance the surveillance of IID. Unfortunately, due to the limited school attendance data obtained, not all aspects of the intended

analysis could be undertaken. Therefore, whilst this study has been able to describe and explore the association between seasonal trends in school illness absences and IID surveillance indicators, it has not been able to provide an assessment of the burden of IID on school absenteeism or the impact of the rotavirus vaccine on illness absence.

The descriptive analysis revealed a seasonal trend in illness absenteeism, with absences increasing from September onwards and staying high during the winter months before declining again in the spring. Secondary schools had a marginally higher median weekly rate of illness absence compared to primary schools and a higher peak absence which occurred one week later than the peak in primary schools. This was unexpected, as it was anticipated that younger children would experience more infectious illness than secondary aged children. However, previous studies have found little difference in illness absenteeism between age groups,[168,207] and these data will also capture causes of illness which are non-infectious, such as chronic disease and mental health.

Trends in illness absenteeism appeared to correlate with trends in GP consultations, telehealth calls and viral IID laboratory reports, with Quasi-Poisson regression revealing a positive, statistically significant association between these surveillance indicators and school illness absences. Negative associations were identified for bacterial and protozoal IID, suggesting that viral IID is more likely to be contributing to school illness absence than bacterial or protozoal IID. The cases of individual viral pathogens were too few to allow a more detailed analysis, although given the winter peak of illness absence and the trend in national norovirus reports, norovirus is likely to be an important causative organism. However, as school illness absence is non-specific and both viral IID and respiratory pathogens have a high prevalence over the winter months, it cannot be concluded that IID is driving the trend in school attendance. Attributing cause to trends in illness absence is further complicated by the limited testing undertaken in this age group, which means a significant proportion of both viral IID and respiratory illness in school-aged children will go undetected in surveillance data.

The school survey findings suggest that whilst respiratory pathogens may cause more illness in schools, diarrhoea and vomiting is thought to result in more absences. This is unsurprising given the public health guidance advising exclusion during, and for 48-hours after, diarrhoea and vomiting whilst no exclusion is required for respiratory illness. Of those who responded, most schools collected their own symptom-specific attendance data, although the exact nature of the information collected varied. This suggests it may be feasible to encourage schools to adopt a standardised approach to recording symptoms, which could enhance the utility of school attendance data as a form of surveillance. This could include having pre-determined categories of illness, such as diarrhoea and vomiting, respiratory,

rash, or non-infectious, which could be used to distinguish seasonal trends in different syndromes. Of note, none of the schools who recorded data on symptoms classified the absence by syndrome, which is likely to make it difficult to aggregate symptom-specific absences in their current format both at school level and across larger footprints. Developing a standardised approach to recording symptom data will be crucial if these data are to be used for surveillance purposes. Furthermore, only two schools recorded the data electronically, with other schools keeping paper records or not documenting reported symptoms. Again, the use of electronic record keeping will be essential to allow easy and timely data transfer to relevant public health bodies. The finding that some schools are already recording symptom-specific absences electronically is encouraging and could provide a model for data collection that could be rolled out to schools more widely.

Data on symptom-specific school absences could also have application beyond public health surveillance and most schools who responded to the survey felt that routinely collecting information on symptoms would be beneficial to the school. However, an important critique of these data was the accuracy of information received from parents. Nevertheless, this is likely to be most problematic at an individual pupil level. For wider surveillance purposes, whilst it may lead to an over-estimate of the burden of illness amongst children, if false illness reports occur randomly throughout the year they will not alter the overall trends of illness absenteeism and as such these data could still be used for detecting peaks in seasonal illness. However, consideration needs to be given as to whether false illness reports are patterned, with previous studies noting that all-cause absences increased around school holidays and illness absences varied by weekday and with media coverage of pandemics or high-profile deaths in children.[168,205,207,213] With more years' worth of data these trends could be analysed and taken into account when modelling rates of school illness absenteeism.

Strengths and limitations

This study uses nine years of surveillance data to explore whether school illness absences could be used to capture seasonal trends and outbreaks of infectious intestinal disease amongst school-aged children. Unfortunately, insufficient school data were received to allow the planned statistical modelling and consequently a temporal-spatial comparison between school absenteeism and health surveillance datasets was not possible. This was primarily due to issues accessing school data in the required format. Whilst all three Local Government areas approached consented to participate in the study, only one area was able to provide weekly school attendance data. Of the other Local Government areas, one had only started to collect weekly data for the 2017/18 academic year and was, therefore, unable to provide historical data at this level. The other area used an external IT

provider to handle attendance data and were only able to access data for individual pupils, or summary level data provided in standardised formats, none of which included weekly attendance per school. The researchers requested, but were unable, to directly contact the external IT provider to explore accessing attendance data at the required level for this study and therefore no data were provided for this Local Government area. Whilst the third Local Government area was able to provide weekly illness absence rates per school within their area, a further breakdown of data by year group and sex were not possible due to the limits of their computer systems. The combination of historical variations in data collection and IT system issues made this a difficult dataset to access and variation in data handling between Local Government areas make it complex to access school data across a large geographical footprint. However, the same IT system was used by individual schools to record absences and consequently it may be feasible to explore data collection directly from the schools' IT provider. Whilst this approach would be more practical in terms of data collection, it does raise issues regarding consent as the IT provider does not have ownership of the data. If consent is required from each school who uses the system, this approach would also become unfeasible unless the sharing of these data with health authorities could be incorporated into the statutory reporting of school absences.

There was also poor engagement with the school survey, with the small number of responses making it difficult to make wider inferences from the findings. The reason for the low response rate can only be hypothesized and could be due to the dissemination strategy or the lack of time amongst teachers to engage with external projects and research. Certainly, schools are likely to have significant pressures on their time and being first and foremost educational institutions, may have limited capacity to engage with health-related issues unless they are directly impacting on academic outcomes and targets. The success of a school-based surveillance system is likely to be improved if the collection and dissemination of attendance data can be streamlined using existing electronic attendance systems, minimising the additional workload for schools.

A further limitation was the small number of cases reported within the surveillance datasets when broken down by Local Government area and age group. Even with data aggregated across nine surveillance years, the number of laboratory reports were too small to allow pathogen-specific modelling. For some pathogens these data are unlikely to represent the true number of cases, but rather the lack of testing carried out in this age group. School-aged children are less likely than the very young and old to experience severe disease from organisms such as norovirus [218,232] and can, therefore, be managed in the community without consulting medical services. Consequently, they are likely to be under-represented in existing surveillance data. This limits the potential to model these

data for school children at a local level and investigate the ability of school attendance data to capture local outbreaks of infection. However, it also highlights the need for improving the detection of community cases of disease which do not require healthcare involvement. Future studies should consider modelling data across a regional or national footprint and across a wider age range to allow for the small number of cases identified in school-aged children within existing surveillance data.

4.6 Conclusion

School illness absenteeism peaked during the winter months and demonstrated a positive association with trends in GP consultations and telehealth calls for IID, as well as laboratory reports for viral IID pathogens. Negative associations were identified for bacterial and protozoal IID, suggesting that viral IID is more likely to be contributing to illness absence in schools. However, as illness absenteeism is non-specific and both viral IID and respiratory pathogens have a high prevalence over the winter months, it cannot be concluded that IID is driving the trends in school attendance data. Future work should consider the feasibility of collecting syndrome-specific absence data to enhance the specificity of school data and explore its role in the syndromic surveillance of a broader range of childhood infectious disease.

Chapter 5. Do children provide an early warning of seasonal norovirus infection?

Can cases and outbreaks of norovirus in children provide an early warning of seasonal norovirus infection: an analysis of nine seasons of surveillance data in England UK

Anna Donaldson, John Harris, Roberto Vivancos, Sarah O'Brien

Published: BMC Public Health. 2022; 22(1): 1393
doi.org/10.1186/s12889-022-13771-z

How this publication fits into my thesis

Chapter 2 and Chapter 3 have highlighted the importance of primary-aged school children both in terms of the risk of experiencing school outbreaks and as a potential early warning of circulating infections in the community. In Chapter 4, the role of school attendance data in the surveillance of IID was considered but unfortunately due to insufficient data no firm conclusions could be drawn. The previous chapter also highlighted the challenges of analysing data for school children at a local level as the number of cases captured in routine surveillance at a Local Government level were limited, even across a nine-year period. This was especially pertinent for laboratory data, which captured very few cases in this age group within each area.

This chapter builds on the findings of the systematic review and explores whether cases and outbreaks of IID in children can provide an early warning of seasonal norovirus infection. To address some of the data issues in the previous chapter, national level surveillance data are analysed across nine norovirus seasons. Data on reported IID outbreaks within schools are compared with outbreaks in care homes and hospitals, general practice (GP) consultations for IID, telehealth calls for diarrhoea and/or vomiting and laboratory norovirus reports for England, UK. The analysis focuses on norovirus, as opposed to all IID, as norovirus is the predominant cause of outbreaks in semi-enclosed settings and therefore the outbreak data better represent norovirus than all-cause IID. Lagged correlations and breakpoint analysis are used to explore whether cases in children and outbreaks in schools provide an early warning of seasonal norovirus infection.

My contribution

My co-authors and I conceived of the study, and all contributed to the study design and methodology. I undertook the data cleaning, statistical analysis and wrote the manuscript.

5.1 Summary

Background

Children are important transmitters of norovirus infection and there is evidence that laboratory reports in children increase earlier in the norovirus season than in adults. This raises the question as to whether cases and outbreaks in children could provide an early warning of seasonal norovirus before cases start increasing in older, more vulnerable age groups.

Methods

This study uses weekly national surveillance data on reported outbreaks within schools, care homes and hospitals, general practice (GP) consultations for infectious intestinal disease (IID), telehealth calls for diarrhoea and/or vomiting and laboratory norovirus reports from across England, UK for nine norovirus seasons (2010/11-2018/19). Lagged correlation analysis was undertaken to identify lead or lag times between cases in children and those in adults for each surveillance dataset. A partial correlation analysis explored whether school outbreaks provided a lead time ahead of other surveillance indicators, controlling for breaks in the data due to school holidays. A breakpoint analysis was used to identify which surveillance indicator and age group provided the earliest warning of the norovirus season each year.

Results

School outbreaks occurred 3-weeks before care home and hospital outbreaks, norovirus laboratory reports and NHS 111 calls for diarrhoea, and provided a 2-week lead time ahead of NHS 111 calls for vomiting. Children provided a lead time ahead of adults for norovirus laboratory reports (+1-2 weeks), NHS 111 calls for vomiting (+1 week) and NHS 111 calls for diarrhoea (+1 week) but occurred concurrently with adults for GP consultations. Breakpoint analysis revealed an earlier seasonal increase in cases among children compared to adults for laboratory, GP and NHS 111 data, with school outbreaks increasing earlier than other surveillance indicators in five out of nine surveillance years.

Conclusion

These findings suggest that monitoring cases and outbreaks of norovirus in children could provide an early warning of seasonal norovirus infection. However, both school outbreak data and syndromic surveillance data are not norovirus specific and will also capture other causes of IID. The use of school outbreak data as an early warning indicator may be improved by enhancing sampling in community outbreaks to confirm the causative organism.

5.2 Introduction

Norovirus is the single most common cause of infectious intestinal disease (IID) in high-income countries, accounting for approximately 11-16% of community cases.[25–28] In the UK, it affects nearly 5% of the population every year.[29] Norovirus infection occurs all year round but is more common during the winter months (December to February in the Northern Hemisphere).[30] Norovirus typically causes a mild, self-limiting illness characterised by vomiting, watery diarrhoea, abdominal cramps and fever, with symptoms typically lasting two to three days.[37] However, the severity of disease and duration of symptoms can be affected by factors such as age and co-morbidity, with hospital patients found to experience more prolonged illness.[38–40] Norovirus is highly transmissible due to the low infectious dose and high levels of viral shedding,[41] with as few as ten to one hundred particles sufficient to cause infection.[65] It can spread through faecal-oral transmission as well as being widely dispersed by vomit where it can transmit to others via inhalation, contamination of surfaces or direct contamination of hands.[60,65] Consequently, it is a common cause of outbreaks in semi-enclosed environments, such as hospitals, nursing homes and schools.[42,43] Each year norovirus causes widespread disruption to healthcare services and has been estimated to cost the global economy $4.2 billion in healthcare costs and $60.3 billion in societal costs per annum.[44] Each year in the UK, norovirus is estimated to cause between 6 000 and 18 000 hospital admissions, 30 000 accident and emergency attendances, 160 000 general practice (GP) consultations and 56 000 calls to telehealth services.[45]

Children are thought to be important drivers of norovirus infection and experience prolonged symptoms and viral shedding, reduced immunity and higher levels of infectiousness.[217–220,233] Their high numbers of close social contacts, especially in home and school environments enables the spread of infection to both child and adult age groups.[74,75] Young children have one of the highest incidence of norovirus,[27,100,234] and household contact with a symptomatic child is a risk factor for infection in older children and adults.[235–237] Mathematical modelling has predicted paediatric norovirus vaccination could prevent 18-21 times more cases than elderly vaccination by providing both direct protection to children and indirect protection to adults.[133] In addition, there is evidence that cases in children may start increasing earlier in the norovirus season than cases in adults.[100] This raises the question as to whether cases in children could provide an early warning of seasonal norovirus before cases start increasing in older, more vulnerable age groups. This study uses national surveillance data for England, UK to explore whether cases of norovirus in children and outbreaks of IID in schools occur earlier in the season than cases and outbreaks amongst adult age groups and could, therefore, act as an early warning of seasonal norovirus.

5.3 Methods

Data sources

National surveillance data held by Public Health England (PHE) were requested over a 10-year period (1st January 2010 to 31st December 2019). Data were extracted on reported outbreaks within schools, care homes and hospitals, general practice (GP) consultations for IID, calls for diarrhoea and/or vomiting to remote telehealth services, which provide telephone-based health advice and information, and laboratory norovirus reports from across England, UK.

Outbreak surveillance of IID has been in existence in the UK since 1992 and data on outbreaks are currently collected via two reporting systems. Since 2009, hospital norovirus outbreaks have been reported nationally via the web-based Hospital Norovirus Outbreak Reporting System (HNORS), although participation and reporting are voluntary.[142] IID outbreaks in other settings are voluntarily reported to local public health teams, who record details of the outbreak and the subsequent management on a national web-based system.[143] An outbreak is defined as two or more cases linked in time or place, or a greater than expected rate of infection compared with the usual background rate for a given place and time.[120] Outbreaks are recorded as suspected or laboratory confirmed, depending on whether a causative organism has been isolated.

Data on GP consultations and telehealth calls form part of PHE's National Real-time Syndromic Surveillance Service, which collects and augments data on presenting symptoms and/or suspected diagnoses from different parts of the healthcare system across England.[140] In general practice, syndromic indicators have been developed based on the Read Code system, which is the recommended national diagnostic classification system for GPs.[147] These syndromic indicators include gastroenteritis, vomiting, and diarrhoea although each indicator may be triggered by a variety of different Read Codes. Data on GP in-hours consultations are collected through a sentinel surveillance system which covers approximately 12% of England's population and has been monitored by PHE since 2012. Telehealth services (NHS 111 and its predecessor NHS Direct) utilise electronic clinical algorithms, which contain a series of questions relating to a reported symptom.[150] Syndromic surveillance is based on monitoring how often these algorithms are triggered and identifying exceedances from the normal background level. Relevant algorithms for IID include both vomiting and diarrhoea. NHS Direct was in operation from 2001 until 2013, when the service was replaced by NHS 111. During the piloting and transition to NHS 111 (2012-2013), the coverage of both systems was reduced and therefore NHS 111 data were only included from the 2014/15 norovirus season onwards.

Data on positive laboratory samples are reported to PHE via two mechanisms. The statutory notification system within the UK makes it mandatory for clinicians to report suspected cases of certain infectious diseases and laboratories must inform PHE when they confirm a notifiable organism within a specimen sample.[139] Norovirus is not classed as a notifiable organism, but both suspected food poisoning and infectious bloody diarrhoea are formally notifiable. In addition, there are voluntary reporting systems established with the majority of laboratories across the country, who submit weekly electronic reports of isolated organisms, including norovirus, to Public Health England.

Data analysis

Weekly-level data were analysed according to the norovirus season, with the season considered to start in calendar week 27 and end in calendar week 26 of the following year. Data were only included if they were available for the complete norovirus season. The analysis incorporated outbreak data and laboratory data from nine norovirus seasons (2010/11-2018/19), GP data from seven seasons (2012/13-2018/19) and NHS 111 data from five (2014/15-2018/19). For the analysis, cases were divided into child and adult age groups. Both NHS 111 and GP data contained pre-determined age categories, so the age boundaries for children and adults varied depending on the categories available within each dataset. For laboratory and NHS 111 data, children were defined as 0-15 years and adults ≥16 years. For GP data, the alternative definitions of 0-14 years and ≥15 years were used. Cases with missing or invalid data on age were excluded from the analysis.

A descriptive analysis was undertaken to explore the number of cases and outbreaks reported, time trends, and seasonality within each dataset. Median season week and cumulative proportions were used to identify which surveillance indicator and age group were reported earliest in the norovirus season. A Spearman's rank correlation analysis was used to compare the temporal patterns of cases in children with those in adults and to identify any lead or lag times between the age groups for laboratory, NHS 111 and GP data. For each dataset, data were broken down into child and adult age groups and then aggregated by norovirus season week. A further correlation analysis was undertaken to explore whether school outbreaks provided a lead time ahead of other surveillance indicators. To adjust for the natural breaks in school outbreak data, a Spearman's rank partial correlation was undertaken, controlling for school holidays. To allow data to be combined from across multiple years, school holidays were assumed to fall on the same weeks each year. The selected weeks were based on existing literature.[231] For both analyses, lead or lag time were determined by the week with the highest positive correlation up to ±4 weeks.

Finally, a breakpoint analysis was conducted to identify which surveillance indicator and age group provided the earliest warning of the norovirus season. Each surveillance indicator was analysed as a single timeseries, spanning multiple norovirus seasons, regressed against a constant. A breakpoint represented a structural change in the regression model. A breakpoint function was applied which allowed for multiple breakpoints to be detected across the study period,[238] allowing for one or more norovirus peaks to be identified in each dataset each year. No limits were put on the number of possible breakpoints across the study period. Data were smoothed prior to analysis, using a 4-week rolling average, to mitigate the effects of breaks in data due to school holidays. The minimum number of observations between breakpoints was set to 13 weeks (3 months). This was selected to account for the prolonged break in school outbreak data over the summer months and ensure breakpoints were not triggered when outbreak reporting re-commenced after school holidays. The season week of the first breakpoint in each norovirus season was extracted, alongside 95% confidence intervals (CI), to identify which surveillance indicator and age group provided the earliest warning of the norovirus peak each year. All analysis was undertaken in R 4.0.2.[176]

5.4 Results

For the norovirus seasons 2010/11 to 2018/19, laboratory surveillance detected 65 361 cases of confirmed norovirus infection, 18% of which were in children under the age of 16 years (Table 5.1). Over the same time period, 33 051 IID outbreaks were reported in schools, care homes and hospitals. Care homes accounted for the largest proportion of these (57%), whilst 33% occurred in hospitals, and 10% were reported in schools. From 2012/13 to 2018/19 there were over 6 million reported GP consultations for IID and over the course of five norovirus seasons, NHS 111 received over 1.1 million calls for diarrhoea and 1.7 million calls for vomiting. Whilst children accounted for a third of GP consultations for IID, they were responsible for nearly half of all calls to NHS 111 for vomiting.

Figure 5.1 shows the time trends of each surveillance dataset. Laboratory norovirus reports demonstrated a distinct seasonal trend with a peak during the winter and spring each year, although the exact timing of the peak varied. Hospital and care home outbreaks closely mirrored the seasonality of laboratory norovirus reports, but school outbreaks showed more variability. There were visible peaks in school outbreaks coinciding with laboratory reports in six of the surveillance years, but less defined peaks in the remaining three years. Winter/spring peaks were also captured in NHS 111 data for both vomiting and diarrhoea but GP consultations for IID showed a less clear seasonal trend.

Table 5.1: Characteristics of included surveillance datasets

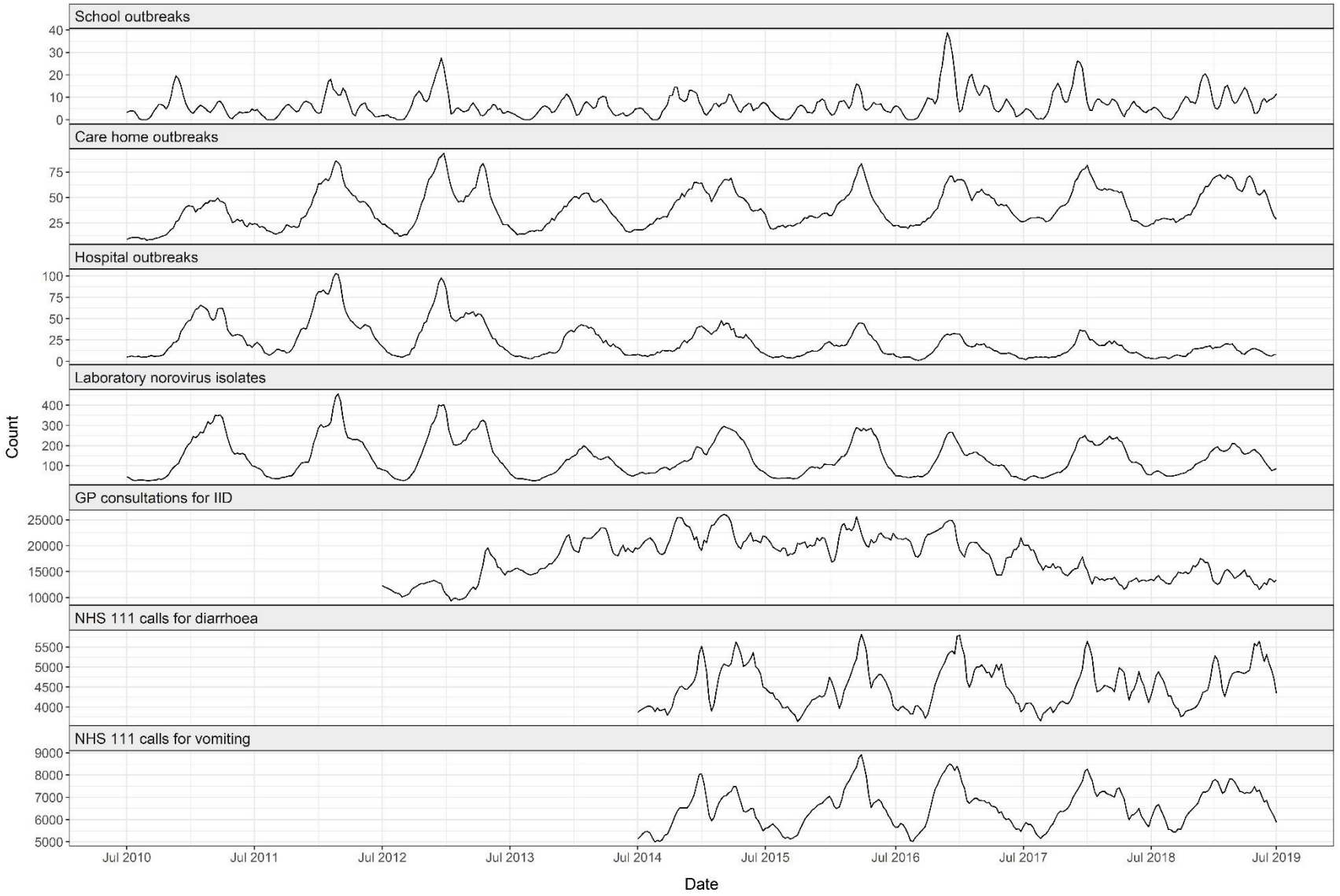
Surveillance dataset (2010/11 – 2018/19)	Total reported	Median number of cases/outbreaks reported per norovirus season* (IQR)	Median season week of reported cases/outbreaks (IQR)
Outbreaks			
Schools	3 168	344 (285-413)	25 (19-36)
Care homes	19 000	2 215 (1 975-2 391)	29 (20-38)
Hospitals	10 883	948 (732-1 569)	30 (23-38)
Laboratory norovirus reports			
Children (0-15yrs)	11 463	1 287 (1 116-1 449)	28 (17-39)
Adults (≥16yrs)	53 898	6 011 (5 308-7 075)	32 (24-39)
All	65 361	7 089 (6 424-7 912)	32 (23-39)
NHS 111 calls for diarrhoea^			
Children (0-15yrs)	470 928	95 159 (94 450-96 568)	28 (16-40)
Adults (≥16yrs)	711 480	142 355 (141 547-142 677)	27 (14-40)
All	1 182 408	238 923 (233 941-239 620)	27 (15-40)
NHS 111 calls for vomiting^			
Children (0-15yrs)	841 587	167 614 (165 405-171 424)	28 (17-39)
Adults (≥16yrs)	868 772	175 200 (170 481-177 051)	27 (14-39)
All	1 710 359	341 905 (340 605-346 398)	27 (16-39)
GP consultations for IID$			
Children (0-14yrs)	2 059 558	312 334 (231 719-341 421)	28 (16-39)
Adults (≥15yrs)	4 362 475	658 813 (518 330-731 522)	26 (13-39)
All	6 422 033	971 147 (750 049-1 072 942)	27 (14-39)

* The norovirus season was considered to start in calendar week 27 and end in calendar week 26 of the following year.

^ NHS 111 data ran from 2014/15 to 2018/19.

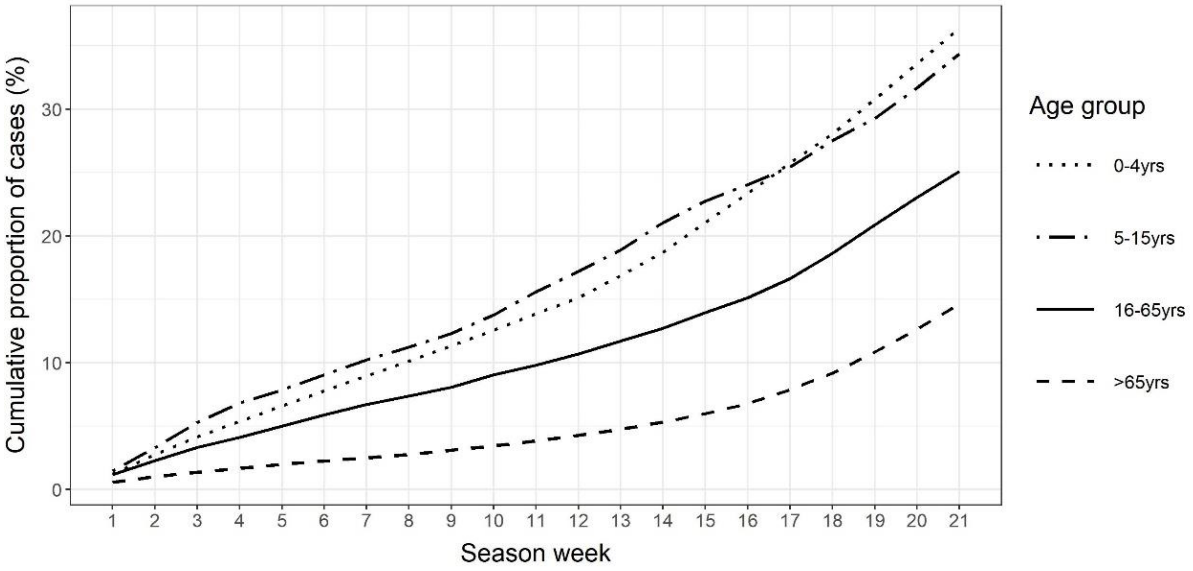
$ GP data ran from 2012/13 to 2018/19.

Figure 5.1: Time trends in surveillance datasets, based on a 4-week rolling average



Based on the median season week of reported cases and outbreaks, school outbreaks occurred earlier in the norovirus season than the other surveillance indicators (week 25), two weeks earlier than NHS 111 calls and GP consultations, and 4-5 weeks earlier than care home and hospital outbreaks (Table 5.1). Laboratory reports had the latest median season week (week 32), seven weeks after school outbreaks. Whilst GP consultations and NHS 111 calls in children did not have an earlier median season week than adults, laboratory reports in children occurred 4 weeks earlier than for adults. Further analysis of laboratory samples by age showed that cases of laboratory-confirmed norovirus in children started increasing earlier in the season than cases in adults (Figure 5.2). In preschool (<5yrs) and school-aged children (5-15yrs), 25% of cases were reported by week 17, compared to week 21 in adults (16-65yrs) and week 25 in elderly (>65yrs).

Figure 5.2: Cumulative proportion of norovirus laboratory reports (2010/11-2018/19), by age group



*Season week 1 corresponds to ISOweek 27.

Correlation analysis

As shown in Table 5.2, laboratory-confirmed cases of norovirus in children showed a positive correlation with cases in adults and provided a 1-2-week lead time across the norovirus season (r_s 0.80, p<0.001). Children provided a 1-week lead time ahead of adults for both NHS 111 vomiting calls (r_s 0.78, p<0.001) and NHS 111 diarrhoea calls (r_s 0.69, p<0.001). GP consultations for children did not appear to be correlated with consultations for adults, with no evidence of significant lead or lag times.

Table 5.2: Spearman's rank correlation, showing the relative temporal position of cases in children (0-15yrs) in relation to adults (≥16yrs) (by week)

Surveillance dataset	Lead time					Lag time			
	+4 weeks	+3 weeks	+2 weeks	+1 week	Concurrent	-1 week	-2 weeks	-3 weeks	-4 weeks
Laboratory norovirus reports	0.75	0.78	0.80	0.80	0.75	0.67	0.58	0.48	0.40
NHS 111 calls for diarrhoea[$]	0.68	0.66	0.64	0.69	0.64	0.39	0.20	0.10	0.02
NHS 111 calls for vomiting[$]	0.62	0.71	0.76	0.78	0.73	0.58	0.48	0.41	0.37
GP consultations[^*]	-0.09	-0.09	-0.30	-0.21	0.14	-0.03	-0.15	-0.01	-0.09

[^] Age groups for GP consultations are 0-14yrs and ≥15yrs.

[*] GP data runs from 2012/13 to 2018/19.

[$] NHS 111 data runs from 2014/15 to 2018/19.

Table 5.3: Spearman's rank partial correlation, comparing outbreaks in schools in relation to listed surveillance indicator, controlled for school holidays

Relative temporal position of school outbreaks in relation to listed dataset	Lead time					Lag time			
	+4 weeks	+3 weeks	+2 weeks	+1 week	Concurrent	-1 week	-2 weeks	-3 weeks	-4 weeks
Care home outbreaks	0.73	0.76	0.73	0.68	0.61	0.46	0.42	0.33	0.25
Hospital outbreaks	0.73	0.77	0.72	0.67	0.61	0.50	0.41	0.34	0.30
Laboratory norovirus reports	0.68	0.69	0.64	0.59	0.52	0.43	0.35	0.23	0.16
NHS 111 calls for diarrhoea[$]	0.54	0.59	0.55	0.34	0.22	0.06	0.00	-0.02	-0.10
NHS 111 calls for vomiting[$]	0.63	0.76	0.80	0.69	0.60	0.48	0.41	0.38	0.30
GP consultations[^*]	0.16	0.24	0.24	0.52	0.55	0.48	0.37	0.21	0.23

[^] Age groups for GP consultations are 0-14yrs and ≥15yrs.

[*] GP data runs from 2012/13 to 2018/19.

[$] NHS 111 data runs from 2014/15 to 2018/19.

When controlled for school holidays, school outbreaks were positively correlated with outbreaks in care home and hospitals and provided a 3-week lead time ahead of outbreaks in both settings (r_s 0.76, $p<0.001$ and r_s 0.77, $p<0.001$ respectively) (Table 5.3). School outbreaks also provided a 3-week lead time ahead of laboratory surveillance data (r_s 0.69, $p<0.001$) and NHS 111 calls for diarrhoea (r_s 0.59, $p<0.001$), as well as a 2-week lead time ahead of NHS 111 calls for vomiting (r_s 0.80, $p<0.001$). GP consultations were concurrent with outbreaks in schools (r_s 0.55, $p<0.001$).

Breakpoint analysis

When laboratory reports, GP consultations and NHS 111 calls were broken down into child and adult age groups, the breakpoint analysis identified an earlier increase in laboratory reports in children in all nine surveillance years, 3-10 weeks ahead of an increase in adult cases (Table 5.4). GP consultations and NHS 111 calls for children also led adults, with breakpoints occurring earlier in all five seasons for NHS 111 calls, and five out of six seasons for GP consultations. There was an earlier increase in school outbreaks compared to other surveillance indicators in five out of nine surveillance years (Table 5.5). No peak was identified for school outbreaks in two of the years and the breakpoint was concurrent or lagged behind other indicators in the remaining two years.

Table 5.4: Season week of first detected breakpoint with 95% confidence intervals, based on 4-week rolling average, by norovirus season and age group

Surveillance dataset	2010/11	2011/12	2012/13	2013/14	2014/15	2015/16	2016/17	2017/18	2018/19
Laboratory norovirus reports									
Children	16 (14-17)	16 (15-17)	13 (11-14)	16 (15-17)	9 (5-13)	15 (14-18)	14 (12-15)	14 (12-15)	14 (5-21)
Adults	24 (23-25)	24 (23-25)	17 (16-18)	22 (20-25)	19 (17-20)	18 (16-23)	18 (17-19)	20 (19-21)	21 (18-23)
GP consultations for IID									
Children	NA	NA	13 (6-14)	19 (18-20)	14 (11-15)	15 (11-18)	13 (11-14)	-	13 (49-18)
Adults	NA	NA	41 (40-42)	21 (19-22)	27 (20-41)	29 (23-35)	48 (46-50)	-	9 (8-11)
NHS 111 calls for diarrhoea									
Children	NA	NA	NA	NA	15 (12-16)	17 (15-20)	15 (14-16)	15 (13-17)	19 (17-20)
Adults	NA	NA	NA	NA	24 (19-25)	34 (31-37)	17 (12-18)	23 (17-24)	24 (21-25)
NHS 111 calls for vomiting									
Children	NA	NA	NA	NA	15 (13-16)	16 (15-18)	15 (14-16)	15 (14-16)	17 (15-19)
Adults	NA	NA	NA	NA	23 (18-24)	25 (22-26)	17 (13-18)	23 (19-24)	24 (21-25)

Table 5.5: Season week of first detected breakpoint with 95% confidence intervals, by norovirus season, based on a 4-week rolling average

Surveillance dataset	2010/11	2011/12	2012/13	2013/14	2014/15	2015/16	2016/17	2017/18	2018/19
School outbreaks	13 (8-14)	27 (23-28)	13 (10-14)	-	13 (11-15)	-	18 (11-19)	12 (9-13)	13 (4-16)
Care home outbreaks	21 (20-22)	24 (23-25)	16 (15-17)	20 (19-21)	15 (14-16)	30 (28-31)	18 (17-19)	19 (17-20)	20 (19-21)
Hospital outbreaks	22 (21-23)	23 (22-24)	17 (15-18)	22 (21-23)	21 (20-22)	21 (20-24)	17 (15-18)	20 (18-21)	17 (13-20)
Laboratory norovirus reports	24 (23-25)	24 (23-25)	17 (15-18)	22 (21-24)	18 (16-19)	18 (16-22)	17 (15-18)	20 (19-21)	20 (17-22)
GP consultations for IID	NA	NA	40 (38-41)	21 (20-22)	14 (11-15)	29 (24-33)	13 (11-16)	-	11 (9-12)
NHS 111 calls for diarrhoea	NA	NA	NA	NA	16 (11-17)	33 (30-34)	16 (15-17)	21 (17-22)	24 (22-25)
NHS 111 calls for vomiting	NA	NA	NA	NA	15 (13-16)	17 (16-18)	15 (14-16)	19 (18-20)	20 (19-21)

5.5 Discussion

Whilst previous studies have demonstrated an important role for children in the transmission of norovirus infection,[133,235–237] it was uncertain whether or not children were affected earlier in the norovirus season than adults. This study found that outbreaks of IID in schools had an earlier median season week than outbreaks in other settings and correlated well with outbreaks in care homes and hospitals, laboratory norovirus reports and NHS 111 calls for vomiting. School outbreaks occurred 3 weeks before care home and hospital outbreaks, norovirus laboratory reports and NHS 111 calls for diarrhoea, and provided a 2-week lead time ahead of NHS 111 calls for vomiting. Children provided a lead time ahead of adults for both norovirus laboratory reports (+1-2 weeks), NHS 111 calls for vomiting (+1 week) and NHS 111 calls for diarrhoea (+1 week) but occurred concurrently with adults for GP consultations. Breakpoint analysis revealed an earlier seasonal increase in cases in children compared to adults for laboratory, GP and NHS 111 data, with school outbreaks increasing earlier than other surveillance indicators in five out of nine surveillance years.

Our study supports the findings of Bernard *et al*. who identified that laboratory-confirmed norovirus cases in Germany started rising in children earlier in the season than adults and elderly.[100] However, in our study laboratory reports still had the latest median season week of all the surveillance datasets. Previous studies had identified that telehealth calls for vomiting provided a lead time ahead of laboratory surveillance data, with vomiting calls for young children providing the earliest indication of the norovirus season.[192,239] In this study, whilst NHS 111 calls for vomiting did have an earlier median season week than laboratory reports, school outbreaks had the earliest median season week, demonstrating a lead time ahead of other surveillance indicators and an earlier seasonal increase in five out of nine surveillance years. This would suggest a potential role for school outbreak data in the surveillance of norovirus which could provide an earlier warning of the start of norovirus season compared to existing indicators. Studies have previously explored the role of other school-based surveillance systems, such as those based on school absenteeism, and have found syndrome-specific absences for influenza provided a lead time ahead of traditional surveillance systems during the H1N1 pandemic.[165,166,211] Whilst school outbreak data were not norovirus specific, high levels of viral shedding and a low infective dose make norovirus a common cause of outbreaks in semi-enclosed settings.[43] In this study, the close mirroring of time trends of outbreaks in care homes and hospitals with laboratory confirmed norovirus cases would suggest that norovirus was driving the majority of outbreaks in these settings. Whilst school outbreaks did not mirror norovirus trends as closely, the correlation with laboratory-confirmed cases would suggest that norovirus was a likely cause of many of the IID outbreaks reported in schools. The utility of outbreak data for norovirus surveillance may be

further improved by enhancing sampling and laboratory testing in community outbreaks to confirm norovirus as the causative organism.

Within individual surveillance datasets, the breakpoint analysis suggested children provided an earlier signal than adults across all datasets and for all norovirus seasons, a finding that was less consistent in the correlation analysis and not reflected in the descriptive analysis. The correlation analysis identified a lead time for children ahead of adults in laboratory data and NHS 111 calls for both vomiting and diarrhoea, consistent with findings that telehealth calls for vomiting in young children provide an earlier signal than vomiting calls for all ages combined.[192] However, no lead time was identified for GP consultations and the correlation coefficients suggested no correlation between trends in children and those in adults. A possible explanation for this finding is that surveillance indicators which are based on broad syndromic definitions will also capture causes of IID other than norovirus. Consequently, children and adults may exhibit different trends in GP consultations, caused by different organisms. This could affect the application of this study's findings to other settings, as seasonal trends in IID may be driven by different organisms in other countries. This is particularly pertinent to rotavirus, where vaccine coverage will impact on the relative importance and burden of this pathogen amongst children and consequently the seasonal trends of IID observed in this age group.

However, some syndromic indicators may be better at capturing certain pathogens than others. As most norovirus infections are mild and short-lived, people with norovirus may be less likely to require a GP consultation and longitudinal data suggest there are 23 norovirus cases in the community for every one which presents to the GP.[29] GP data may, therefore, be better at detecting trends in organisms which cause more severe or prolonged symptoms. This could explain why the GP data in this study did not reflect the seasonal trends seen in the other datasets. For NHS 111 data, whilst both diarrhoea and vomiting are features of norovirus infection, there is evidence that vomiting may be a more prevalent feature amongst children and diarrhoea more common amongst adults.[154,235] This could make calls for vomiting a more sensitive indicator of norovirus infection amongst children and may explain why school outbreaks correlated better with NHS 111 vomiting calls than diarrhoea calls (r_s 0.80 and r_s 0.59 respectively). This highlights the challenge of using syndrome-based surveillance data to monitor specific organisms in the community and it should be considered that the utility of different syndromic surveillance indicators may alter depending on the organism and age group.

Strengths and limitations

This study utilises national surveillance data on over 65 000 laboratory confirmed cases of norovirus, 33 000 outbreaks of IID and over 9 million calls and consultations for IID across nine norovirus seasons. The use of routine surveillance data for this study allows large numbers of cases to be captured across multiple norovirus seasons. However, all surveillance data is subject to reporting bias, as only cases which present to healthcare will be captured in the datasets. This also applies to outbreaks, which are voluntarily reported to Public Health England. Consequently, it cannot be determined whether the lack of a peak in school outbreaks in certain years is the result of fewer outbreaks occurring or lower levels of reporting from schools. Equally, differences in reporting behaviour between children and adults will also be reflected in the data, although as reporting biases are unlikely to change throughout a given norovirus season, it is more likely to affect overall case numbers rather than trend.

An additional limitation of using school-based data are the natural breaks in data collection which occur during school holidays. This could affect the utility of school outbreaks as a surveillance indicator for norovirus. It is well documented that school holidays impact on social mixing patterns [240] and there is evidence that the timing of school holidays can impact on transmission and the size of peaks for other infectious diseases, such as influenza.[231,241,242] In this study, in the years where the breakpoint analysis did not identify a seasonal peak in school outbreaks, norovirus laboratory reports increased later in the season and peaked after the school Christmas break. The same occurred for the year where school outbreaks had a later breakpoint than other surveillance datasets. Consequently, the timing of school holidays relative to the norovirus peak may be affecting the size and timing of peaks in school outbreak data. This could affect the potential of school outbreak data to provide an early warning ahead of other surveillance indicators in any given year.

5.6 Conclusion

Children are recognised as important transmitters of norovirus infection and this study explored whether cases in children and outbreaks in schools occurred earlier in the norovirus season than cases and outbreaks amongst adult age groups. Trends in school outbreaks had a lead time ahead of other surveillance indicators and cases in children provided a lead time ahead of adults for norovirus laboratory reports and NHS 111 calls for both vomiting and diarrhoea. Cases in children started increasing earlier in the season than adults for all surveillance datasets across the study period and school outbreaks increased earlier than other surveillance indicators in five out of nine surveillance years. These findings suggest that monitoring cases and outbreaks of norovirus in children could provide an early warning of seasonal norovirus infection. However, the utility of using school outbreaks

as a surveillance indicator may be affected by the timing of school holidays in relation to the norovirus peak in any given year. Furthermore, both school outbreak data and cases in children from syndromic surveillance are not norovirus specific and hence will also capture other causes of IID. The use of school outbreak data as an early warning surveillance indicator may be improved by enhancing sampling in community outbreaks to confirm the causative organism.

Chapter 6. Symptom profiling for infectious intestinal disease

Symptom profiling for infectious intestinal disease (IID): a secondary data analysis of the IID2 study

Anna Donaldson, Helen Clough, Sarah O'Brien, John Harris

Published: Epidemiology & Infection. 2019; 147: e229

doi.org/10.1017/S0950268819001201

How this publication fits into my thesis

Chapter 5 revealed that IID outbreaks in schools demonstrated a lead time ahead of other surveillance indicators and could provide an early warning of seasonal norovirus infection. Within other routine surveillance datasets, cases in children also started increasing earlier in the season than cases in adults. However, as syndromic surveillance indicators use a broad definition of diarrhoea and vomiting, they are not able to distinguish between different causes of IID. Consequently, it cannot be known whether differences between age groups are driven by different organisms or whether children are truly affected by norovirus earlier in the season than adults.

This chapter considers whether symptoms alone could be used to make inferences as to the causative organism for individual cases of IID. Data from the Second Study of Infectious Intestinal Disease in the Community are analysed to identify if symptoms can differentiate between pathogens. Such symptom profiles could be used to guide case definitions within syndromic surveillance systems and improve the specificity of routine surveillance data. They could also be of value to help guide clinicians and public health professionals in the management of IID, in the absence of microbiological confirmation.

My contribution

I designed the study with co-authors. I undertook the data cleaning, statistical analysis and drafted the manuscript.

6.1 Summary

Less than half of stool samples from people symptomatic with infectious intestinal disease (IID) will identify a causative organism. A secondary data analysis was undertaken to explore whether symptomology alone could be used to make inferences about causative organisms. Data were utilised from the Second Study of Infectious Intestinal Disease in the Community. A total of 844 cases were analysed. Few symptoms differentiated individual pathogens, but grouping pathogens together showed that viral IID was more likely when symptom onset was in winter (odds ratio (OR) 2.08, 95% confidence interval (CI) 1.16-3.75) or spring (OR 1.92, 95% CI 1.11-3.33), the patient was aged under 5 years (OR 3.63, 95% CI 2.24-6.03) and there was loss of appetite (OR 2.19, 95% CI 1.29-3.72). The odds of bacterial IID were higher with diarrhoea in the absence of vomiting (OR 3.54, 95% CI 2.37-5.32), diarrhoea which persisted for greater than 3 days (OR 2.69, 95% CI 1.82-3.99), bloody diarrhoea (OR 4.17, 95% CI 1.63-11.83) and fever (OR 1.67, 95% CI 1.11-2.53). Symptom profiles could be of value to help guide clinicians and public health professionals in the management of IID, in the absence of microbiological confirmation.

6.2 Introduction

Infectious intestinal disease (IID) is characterised by the acute onset of diarrhoea and/or vomiting in otherwise healthy people caused by an infectious, transmissible organism.[22] In the UK, the surveillance of IID is based on statutory notifications, outbreak reports and syndromic surveillance from primary, secondary and remote health services.[140,221] However, as the syndrome of diarrhoea and vomiting can have non-infectious causes, microbiological confirmation remains central to the conclusive diagnosis of IID. Although microbiological testing is the gold standard, in some cohort studies causative organisms have only been identified in 37% to 46% of samples from symptomatic individuals.[22,27,144] The likelihood of identifying a causative organism has been found to be affected by factors such as age, sex, occupation, the absence of specific symptoms such as vomiting, and the timing of the stool sample in relation to symptom onset.[22] Other factors such as the volume of the sample, the performance of the microbiological test and local organism testing policies may also impact on the isolation of organisms.[145,146]

This diagnostic gap means that for over half of symptomatic patients, the cause of illness will not be identified. Whilst the majority of IID cases are self-limiting, being aware of the underlying cause can be of value in both case and outbreak management. For outbreak situations, epidemiological criteria have been developed which utilise, among other factors, the proportion of people affected by given symptoms in order to make inferences as to the underlying organism. The most notable of these are

the Kaplan criteria [153] which were developed in the 1980s in response to the lack of diagnostic tests available for isolating norovirus. Kaplan identified that, where no bacterial organism had been identified in stool cultures, outbreaks were more likely to be caused by norovirus when greater than 50% of people were affected by vomiting; the incubation period was 24 to 48 hours; and the mean duration of illness was 12 to 60 hours. A subsequent re-evaluation of these criteria, once diagnostic tests became available, found them to be highly specific (99%) and moderately sensitive (68%) at distinguishing outbreaks of norovirus from bacterial IID outbreaks.[156] Other epidemiological criteria have also been proposed, including a greater fever-vomiting ratio [154] and a higher diarrhoea-vomiting ratio in bacterial outbreaks, suggesting that fever and diarrhoea are more indicative of a bacterial cause.[155] However, the basis of epidemiological criteria is the relative prevalence of symptoms occurring within a group of affected people and as such they cannot be applied to individual cases. Seasonal outbreaks of IID may present as an increase in reporting of individual cases and therefore being able to ascribe likely cause to single cases of IID has public health and epidemiological value, as well as clinical application. This study uses data from a large community cohort and general practice study to investigate whether symptoms alone can be used to make inferences as to the causative organisms for individual cases of IID.

6.3 Methods

Data sources

A secondary data analysis was undertaken using data from the Second Study of Infectious Intestinal Disease in the Community (IID2 Study), the methodology of which is detailed elsewhere.[22,29] This analysis included data from the two main components of the IID2 study: the general practice (GP) presentation study, which was a 12 month prospective study of people consulting a GP with symptoms of IID; and the prospective population-based cohort study, which involved weekly follow up of healthy volunteers in the community to identify any symptoms of IID. The case definition for IID that was used in the original study was loose stools or clinically significant vomiting lasting less than 2 weeks, in the absence of a known non-infectious cause. Both studies utilised symptom questionnaires and stool sample testing of symptomatic people who met the case definition. Cases were included in this analysis if they had completed a symptom questionnaire and submitted a stool sample. Cases with negative stool samples, where no pathogen was identified, were excluded. Data from dual and triple infections were included multiple times; once for each organism identified, as the primary cause of symptoms could not be determined.

Data analysis

Multivariable logistic regression was used to determine the odds of a case being caused by a given pathogen based on reported symptoms. The explanatory variables included the symptoms outlined in the IID2 study symptom questionnaire, along with the participant's age and date of symptom onset (Table 6.1). Continuous data, namely symptom duration, date of illness onset and age, were categorised before inclusion in the regression models. Given that diarrhoea and vomiting are the predominant symptoms of IID and many people will have both, variables were created to capture cases of diarrhoea in the absence of vomiting, and vomiting in the absence of diarrhoea. These variables were used to explore whether this is a symptom profile which offers discrimination between pathogens. Phi coefficients were used to identify any significant correlations between the explanatory variables which might lead to mathematical problems with model fitting.

The outcome variable was the presence of the infectious organism. Pathogens which accounted for greater than 10% of the total number of cases were analysed independently, to identify symptoms which distinguished them from any other cause of IID. Below this threshold, case numbers were too small to generate meaningful output for a single organism. Grouped organism models were used to capture differences between the broader classes of pathogen; bacteria, viruses and protozoa, sequentially comparing one class against any other cause of IID.

Statistical analysis was undertaken in R 3.3.2.[176] Odds ratios (OR) were calculated using binomial backward stepwise regression. Models were selected based on the Akaike information criterion (AIC). Upper and lower 95% confidence intervals (CI) were calculated around each estimate.

Table 6.1: Explanatory variables included in the multivariate analysis and their coding

Explanatory variables	Coding
Symptoms from IID2 questionnaire [a]	
Diarrhoea days	>3 days = 1 <=3 days = 0
Bloody diarrhoea	Yes = 1 No = 0
Vomiting days	>3 days = 1 <=3 days = 0
Nausea	Yes = 1 No = 0
Nausea days	>3 days = 1 <=3 days = 0
Abdominal pain	Yes = 1 No = 0
Loss of appetite	Yes = 1 No = 0
Fever	Yes = 1 No = 0
Headache	Yes = 1 No = 0
Cough/nose/throat	Yes = 1 No = 0
Combined symptom variables	
Diarrhoea but no vomiting	Yes = 1 No = 0
Vomiting but no diarrhoea	Yes = 1 No = 0
Participant characteristics	
Age [b]	>16yrs = 0, 5-16yrs = 1, <5yrs = 2
Date of onset [b]	Autumn(0), Winter(1), Spring(2), Summer(3) [c]

[a] 'Not sure' responses from the original questionnaires were left blank and treated as missing data.

[b] Coded as factors for analysis.

[c] Seasons defined by meteorological calendar.

6.4 Results

There was a total of 1 657 cases identified from the IID2 study which met the IID2 case definition and had both completed a questionnaire and submitted a stool sample. Of these, 898 cases (54%) were excluded from the analysis as no organism was identified from their stool sample. The total sample size for analysis was 844, including 69 dual infections and eight triple infections.

Norovirus was the most commonly identified cause of IID, and *Campylobacter* was the commonest bacterial cause (Table 6.2). Only four pathogens met the criteria for organism-specific analysis; norovirus, *Campylobacter*, rotavirus and sapovirus. The total number of protozoal infections was less than 10% of the total number of cases and consequently grouped organism models were only

generated for bacterial and viral IID. To capture any important differences in symptoms, protozoa were included in the comparison group for both the bacterial and viral models.

Table 6.2: Organisms, and the associated number of cases, as included in the analysis

Pathogen	No. cases identified
Bacteria	***238***
C. difficile	11
C. perfringens	25
Campylobacter sp.	150
E.coli VTEC	15
Enteroaggregative *E.coli*	27
Salmonella sp.	9
Yersinia sp.	1
Viruses	***576***
Adenovirus	58
Astrovirus	36
Norovirus	237
Rotavirus	96
Sapovirus	149
Protozoa	***30***
Cryptosporidium sp.	15
Giardia sp.	15
Total organisms identified	**844**

The grouped organism models (Table 6.3) showed that the odds of the causative organism being bacterial were higher with diarrhoea in the absence of vomiting (OR 3.54, 95% CI 2.37-5.32), diarrhoea which persisted for greater than 3 days (OR 2.69, 95% CI 1.82-3.99), bloody diarrhoea (OR 4.17, 95% CI 1.63-11.83) and fever (OR 1.67, 95% CI 1.11-2.53). The odds of a viral cause of illness were higher when symptom onset was in winter (OR 2.08, 95% CI 1.16-3.75) or spring (OR 1.92, 95% CI 1.11-3.33), the patient was under 5 years of age (OR 3.63, 95% CI 2.24-6.03) and there was loss of appetite (OR 2.19, 95% CI 1.29-3.72). Given protozoa have a similar aetiology to bacterial IID, as contrasted to viral IID, the analysis was repeated with protozoa assigned to the bacterial group to explore what impact this would have on the symptom profiling. The resulting viral and bacterial/protozoal models did not

differ significantly from the above models; the same explanatory variables were identified, but the significance of winter and spring in the bacteria/protozoa model was increased.

Table 6.3: Grouped organism multivariate model outputs (OR with 95% confidence intervals) for bacterial and viral pathogens, as compared to any other pathogen

Explanatory variable	Odds Ratio (95% confidence intervals)	
	Bacterial	Viral
Aged <5 years	0.25 (0.14-0.41)	3.63 (2.24-6.03)
Onset in winter	0.65 (0.35-1.19)	2.08 (1.16-3.75)
Onset in spring	0.70 (0.40-1.23)	1.92 (1.11-3.33)
Diarrhoea but no vomiting	3.54 (2.37-5.32)	0.27 (0.18-0.39)
Bloody diarrhoea	4.17 (1.63-11.83)	0.30 (0.11-0.79)
Diarrhoea lasting >3 days	2.69 (1.82-3.99)	0.33 (0.22-0.48)
Loss of appetite	0.44 (0.26-0.75)	2.19 (1.29-3.72)
Fever	1.67 (1.11-2.53)	0.59 (0.39-0.88)

The organism-specific modelling generated less meaningful outputs. The *Campylobacter* model largely mirrored the grouped bacterial model and did not provide any further discriminatory information. The virus-specific analysis for norovirus, rotavirus and sapovirus were sensitive to changes in the parameters of the models which led to inconsistent symptom profiles. Phi coefficients were used to identify any significant correlations between the binary explanatory variables which could impact on the model fitting. There was no evidence of significant co-linearity which would affect the modelling, although there were some mild to moderate correlations (phi coefficient <0.5) between some symptoms such as nausea and loss of appetite.

6.5 Discussion

This study has identified that people with IID who reported symptoms of diarrhoea in the absence of vomiting, diarrhoea lasting for more 3 days, bloody diarrhoea and fever were at increased odds of having a bacterial pathogen. Young age (<5 years), onset in spring or winter, and loss of appetite were associated with increased odds of a viral cause. These findings are consistent with other studies which have found associations between bacterial pathogens and symptoms of fever, bloody diarrhoea and prolonged illness, whilst vomiting and a short duration of symptoms have been associated with viral causes.[23,243–245] Epidemiological criteria, utilised in outbreak situations, have similarly highlighted the importance of vomiting and short duration of illness as indicative of norovirus, whilst symptoms

such as fever and diarrhoea have been associated with bacterial outbreaks.[153–156] This study is largely consistent with these criteria, identifying similar associations between these symptoms and the class of the underlying pathogen. However, our analysis would suggest that vomiting and a short duration of symptoms are better ascribed to viral IID than any single viral pathogen. Furthermore, the duration of norovirus symptoms is known to be affected by individual risk factors, such as hospitalisation and age.[39,246] Therefore, the 12-60 hour duration used in the Kaplan criteria may be less applicable when considering individual cases of norovirus illness. This study used less than or equal to 3 days to categorise symptom duration, which provided good discrimination between bacterial and viral causes of IID.

This analysis did not identify symptoms which could be used to adequately differentiate individual IID pathogens. This should act as a caution against making assumptions about the underlying organism on the basis of symptoms alone. However, this dataset did not contain sufficient numbers of some organisms to generate the statistical power necessary to model at the level of individual pathogens. Furthermore, the mild to moderate correlations identified between certain symptoms could make it harder for statistical models to distinguish individual pathogens on the basis of these symptoms alone.

The findings of this study have application for clinicians, public health professionals and epidemiologists, who use symptoms to generate hypotheses regarding causative organisms when managing cases and outbreaks of IID. This analysis would suggest that assumptions should not be made as to the individual pathogen in the absence of microbiological confirmation. However, given the different transmission patterns and natural histories of bacterial and viral IID,[62] using symptom profiles to indicate a likely bacterial or viral cause could assist the early stages of outbreak investigations when microbiology is not yet available. This could help guide infection prevention and control; for example, viral causes are more likely to be spread person-to-person, whereas bacterial IID would rouse suspicion of a food or animal contact. Given the large diagnostic gap for IID, the role of symptoms is still of vital importance to guide clinical and public health action. These findings could have further application for syndromic surveillance systems, enabling symptomatic cases to be categorised as either suspected bacterial or viral IID. However, the benefits of this would have to be weighed against the practical challenges of developing sensitive and specific case definitions that would be compatible with the level of symptom detail gathered and recorded by syndromic surveillance systems.[247]

Strengths and limitations

This study utilised data from a large prospective cohort study,[29] removing some of the reporting biases inherent within national surveillance data.[248] However, given that the severity and duration of illness is known to affect health-seeking behaviour and stool sample submission,[248] mild short-lived illness is still likely to be underrepresented in these data. Despite the large size of the dataset, the total numbers of some of the organisms were too low to allow organism-specific models to be developed for all but the four most common causes. Furthermore, protozoa could not be examined as a separate class of pathogen due to small numbers. In this analysis, protozoa were included in the comparison group for both the bacterial and viral models. To explore the impact this could have had on the modelling, the analysis was repeated with protozoa assigned to the bacterial group. The resulting viral and bacterial/protozoal models did not differ significantly from the original models indicating that the group allocation of the protozoa had little impact on the findings of this analysis.

It should be considered that the grouped organism profiles will be naturally weighted by the relative prevalence of different organisms within each class; *Campylobacter* accounted for almost two thirds of all the bacterial cases and norovirus accounted for over 40% of viral cases. Consequently, the bacterial and viral models will disproportionality reflect the symptoms associated with these pathogens. However, this reflects real-life diagnostics where certain symptoms or organisms are more likely simply because they occur more commonly. Whilst this analysis could not identify symptom profiles which discriminated individual pathogens, this is an area that warrants further exploration. Future studies could also consider the role of co-infections, as co-infections have been found to affect the pathogenicity of organisms.[249]

6.6 Conclusion

Symptom profiles could be used to help dissociate between bacterial and viral causes of IID however, symptoms do not allow further discrimination of individual organisms. Microbiology remains the gold standard and where possible, microbiological confirmation is recommended. However, in situations where microbiology is not available or results are inconclusive, symptom profiling could be of value for clinicians, public health professionals and epidemiologists to distinguish likely bacterial and viral pathogens to guide the management of cases and outbreaks of IID.

Chapter 7. Do symptom profiles alter with age?

Symptom profiling for infectious intestinal disease (IID): do symptom profiles alter with age?

Anna Donaldson, John Harris, Roberto Vivancos, Sarah O'Brien

Published: PLoS ONE. 2022; 17(6): e0269676
doi.org/10.1371/journal.pone.0269676

How this publication fits into my thesis

Chapter 5 highlighted some of the limitations of using syndromic data to monitor IID pathogens and in the previous chapter I identified symptom profiles which could be used to distinguish bacterial IID and viral IID. Such profiles could be of value to guide case definitions within syndromic surveillance systems and could enhance our ability to monitor the epidemiology of these infections. The systematic review in Chapter 3 also highlighted the additional value of using syndrome-specific absences in schools compared to illness absences, with syndrome-specific data correlating better with health surveillance data and providing more lead time ahead of existing indicators. Consequently, symptom profiles could also be used to guide enhanced data collection in schools and to improve the specificity of illness absence data for IID surveillance.

This chapter builds on the findings of the previous study and considers whether symptom profiles for bacterial and viral IID vary across different age groups. Cases of IID captured within the Second Study of Infectious Intestinal Disease in the Community are divided into three age categories: children, adults, and elderly. A descriptive analysis compares the prevalence of different symptoms across the age groups and a multivariable regression analysis explores which symptoms can be used to distinguish bacterial and viral IID within these age bands. This has application to both syndromic surveillance systems and school illness absence data and will help identify what level of symptom data would need to be collected by schools to distinguish between different causes of IID in children.

My contribution

I led on the design, statistical analysis and drafted the manuscript. My co-authors supported the statistical analysis and interpretation of the findings.

7.1 Summary

Symptom profiles have previously been identified for infectious intestinal disease (IID) which distinguish bacterial from viral organisms. However, there is evidence that the seasonality, severity, and duration of IID may differ between children, adults and elderly. A secondary data analysis was undertaken to explore whether symptom profiles for bacterial and viral IID vary across different age groups. Data from 844 cases of IID were divided into three age categories: <16 years, 16-65 years and >65 years. Multivariable logistic regression modelling was used to compare the significance of different symptoms across the three age groups. The odds of bacterial IID in children were increased by onset in the summer, diarrhoea in the absence of vomiting and fever. These symptoms were also associated with lower odds of a viral pathogen. In adults, diarrhoea but no vomiting, bloody diarrhoea and diarrhoea lasting more than 3 days were associated with increased odds of a bacterial organism, whilst onset in the winter or spring and a loss of appetite were associated with viral IID. In the elderly, diarrhoea in the absence of vomiting and diarrhoea lasting more than 3 days were associated with higher odds of bacterial IID and lower odds of a viral cause. Only diarrhoea in the absence of vomiting emerged as a key symptom across all three age groups. Variation in symptom profiles by age has implications for clinicians, public health specialists and epidemiologists who use symptoms to guide presumptive diagnoses in the absence of microbiological confirmation.

7.2 Introduction

Infectious intestinal disease (IID) is characterised by the syndrome of diarrhoea and vomiting, in the absence of a known non-infectious cause.[22] Microbiological testing is key to providing a definitive diagnosis and yet laboratory tests identify a causative organism in less than half of samples.[22,27] Factors such as sample volume, the age and sex of the patient, and the timing of the sample all affect the likelihood of identifying an organism.[22,146] Prior to the widespread availability of laboratory testing, symptoms were used to distinguish between bacterial and viral IID outbreaks.[153–155] These included the diarrhoea-to-vomiting ratio, fever-to-vomiting ratio, and the average duration of illness. The most notable of these, the Kaplan criteria, has subsequently been re-assessed alongside diagnostic testing and was found to be highly specific and moderately sensitive at detecting norovirus outbreaks.[156] Analysis of symptoms from a large community cohort study of IID also demonstrated that symptoms differ between bacterial and viral organisms and that symptoms may be used to dissociate bacterial and viral causes of IID in the absence of microbiological confirmation.[250] The use of IID symptoms to distinguish between organisms has particular relevance to syndromic surveillance systems, many of which use the broad syndrome of diarrhoea and vomiting to monitor and detect seasonal peaks of IID.[148,150] Being able to distinguish between bacterial and viral IID based on

symptoms alone could enhance syndromic surveillance systems and improve the detection and monitoring of these infections.

However, it should be considered whether the symptoms of IID vary with age. The severity of norovirus is known to alter across age groups, with the very young and elderly more likely to experience severe norovirus disease and require hospitalisation.[218,232] Young children and the elderly are also more likely to experience prolonged diarrhoea with norovirus infection, and prolonged viral shedding.[40,217,233,251] Rotavirus infection, prior to vaccination, was more severe amongst young children, with older children and adults typically experiencing milder or asymptomatic infection.[31] However, adults may still be hospitalised with rotavirus infection, especially older adults and the immunosuppressed.[252] There is also evidence that the seasonality of infections may vary between children and adults, with some studies finding that cases of rotavirus and norovirus in adults occurred later in the season than cases in children.[100,253] This study uses data from the Second Study of Infectious Intestinal Disease in the Community to explore whether symptom profiles for bacterial and viral IID vary across age groups.

7.3 Methods

Data sources

Data were used from the Second Study of Infectious Intestinal Disease in the Community (IID2 Study), which is described in detail elsewhere.[22,29] For this analysis, data were included from the two main components of the IID2 study: the general practice (GP) presentation study, which was a 12 month prospective study of people consulting a GP with symptoms of IID; and the prospective population-based cohort study, which involved weekly follow up of healthy volunteers in the community to identify any symptoms of IID. The case definition for IID was loose stools or clinically significant vomiting lasting less than 2 weeks, in the absence of a known non-infectious cause. Both studies utilised symptom questionnaires and stool sample testing of symptomatic people who met the case definition. Cases were included in this analysis if they had completed a symptom questionnaire and submitted a stool sample. Cases with negative stool samples, where no pathogen was identified, were excluded. Data from dual and triple infections were included multiple times; once for each organism identified, as the primary cause of symptoms could not be determined.

The IID2 study received a favourable ethics opinion from the North West Research Ethics Committee (07/MRE08/5), and 37 NHS Research Management and Governance organisations for the 88 included general practices approved the study. All participants of the IID2 study provided written informed

consent, including for the use of their information, suitably anonymised, for future research. All data from the IID2 study were fully anonymised prior to inclusion in this study.

Data analysis

Cases were stratified by age and divided into three categories; children under 16 years, adults aged 16-65 years, and elderly adults aged over 65 years. The outcome variable was the presence of a bacterial or viral organism in the stool sample. The number of protozoal infections were too few to allow separate analysis, but these infections were included in the comparison group for both bacterial and viral models. The explanatory variables were based on the symptoms outlined in the IID2 study symptom questionnaire, as well as the date of symptom onset (Table 7.1). Descriptive statistics were used to examine the proportion of reported symptoms in each age group for both bacterial and viral IID, and medians and interquartile ranges (IQR) were used to investigate differences in the duration of symptoms. Time trends were plotted to explore differences in seasonality based on age and cause. Multivariable logistic regression modelling was used to compare the significance of different symptoms across the three age groups. Continuous data were categorised prior to inclusion in the regression models (Table 7.1).

Statistical analysis was undertaken in R 4.0.4.[176] Odds ratios (OR) were calculated using binomial backward stepwise regression and 95% confidence intervals (CI) were calculated around each estimate. Models were selected based on the Akaike information criterion (AIC). Phi coefficients were used to identify any significant correlations between the explanatory variables which might lead to mathematical problems with model fitting and variance inflation factors (VIF) were used to check the models for multicollinearity.

Table 7.1: Explanatory variables included in the analysis and their coding

Explanatory variables	Coding
Symptoms from IID2 questionnaire [a]	
Diarrhoea days	>3 days = 1 <=3 days = 0
Bloody diarrhoea	Yes = 1 No = 0
Vomiting days	>3 days = 1 <=3 days = 0
Nausea	Yes = 1 No = 0
Nausea days	>3 days = 1 <=3 days = 0
Abdominal pain	Yes = 1 No = 0
Loss of appetite	Yes = 1 No = 0
Fever	Yes = 1 No = 0
Headache	Yes = 1 No = 0
Cough/nose/throat	Yes = 1 No = 0
Combined symptom variables	
Diarrhoea but no vomiting	Yes = 1 No = 0
Vomiting but no diarrhoea	Yes = 1 No = 0
Additional variables	
Date of onset [b]	Autumn(0), Winter(1), Spring(2), Summer(3) [c]

[a] 'Not sure' responses from the original questionnaires were left blank and treated as missing data.

[b] Coded as a factor for analysis.

[c] Seasons defined by meteorological calendar.

7.4 Results

There was a total of 844 cases which met the case definition and inclusion criteria for this analysis. Cases ranged in age from 0 to 91 years. Approximately 35% of cases were children aged 16 or under, 46% were adults aged 16-65 years and 18% were elderly adults aged over 65 years. Over 80% of cases in children were caused by viruses compared to 62% in people aged 16-65 years and 60% of cases aged over 65 years (Table 7.2).

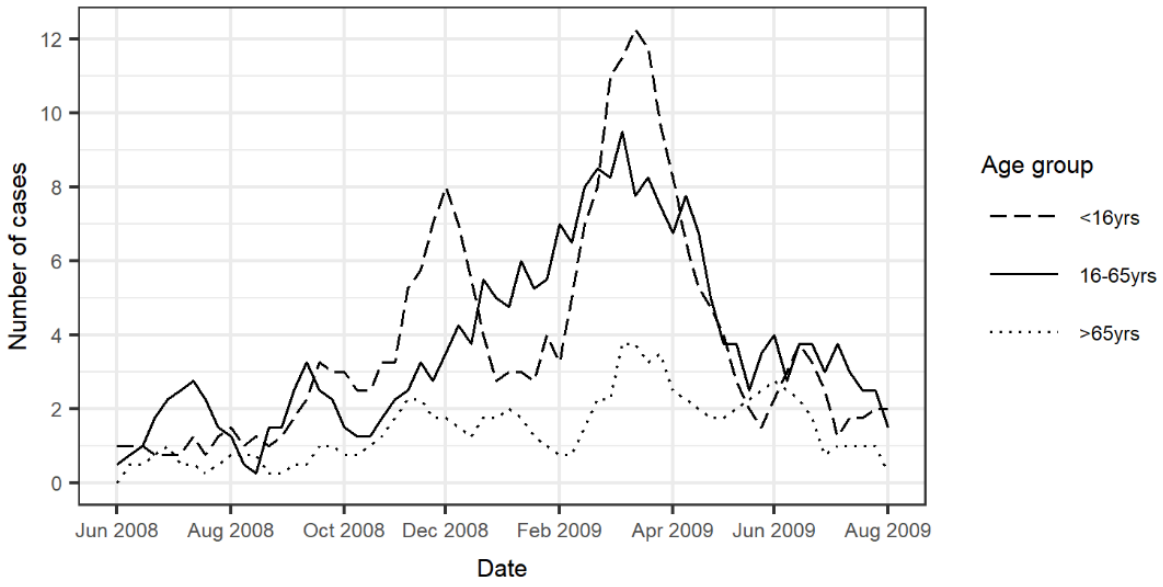
Table 7.2: The number and proportion of isolated organisms, by age group

Pathogen	Age group - number (%)			Total
	<16yrs	16-65yrs	>65yrs	
Bacteria	**42 (14.1%)**	**134 (34.2%)**	**62 (40.3%)**	**238 (28.2%)**
C. difficile	0 (0.0%)	6 (1.5%)	5 (3.2%)	11 (1.3%)
C. perfringens	2 (0.7%)	10 (2.6%)	13 (8.4%)	25 (3.0%)
Campylobacter sp.	25 (8.4%)	93 (23.7%)	32 (20.8%)	150 (17.8%)
E.coli VTEC	1 (0.3%)	7 (1.8%)	7 (4.5%)	15 (1.8%)
Enteroaggregative *E.coli*	11 (3.7%)	13 (3.3%)	3 (1.9%)	27 (3.2%)
Salmonella sp.	2 (0.7%)	5 (1.3%)	2 (1.3%)	9 (1.1%)
Yersinia sp.	1 (0.3%)	0 (0.0%)	0 (0.0%)	1 (0.1%)
Viruses	**241 (80.9%)**	**243 (62.0%)**	**92 (59.7%)**	**576 (68.2%)**
Adenovirus	28 (9.4%)	23 (5.9%)	7 (4.5%)	58 (6.9%)
Astrovirus	23 (7.7%)	8 (2.0%)	5 (3.2%)	36 (4.3%)
Norovirus	78 (26.2%)	119 (30.4%)	40 (26.0%)	237 (28.1%)
Rotavirus	55 (18.5%)	30 (7.7%)	11 (7.1%)	96 (11.4%)
Sapovirus	57 (19.1%)	63 (16.1%)	29 (18.8%)	149 (17.7%)
Protozoa	**15 (5.0%)**	**15 (3.8%)**	**0 (0.0%)**	**30 (3.6%)**
Cryptosporidium sp.	12 (4.0%)	3 (0.8%)	0 (0.0%)	15 (1.8%)
Giardia sp.	3 (1.0%)	12 (3.1%)	0 (0.0%)	15 (1.8%)
Total organisms identified	**298**	**392**	**154**	**844**

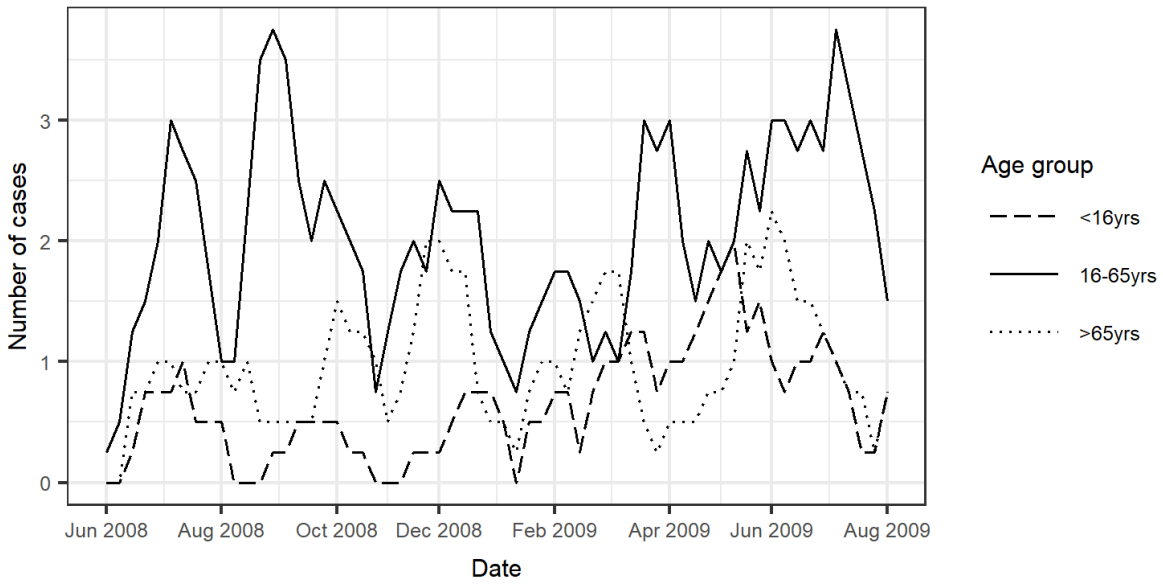
Seasonal trends differed between viral and bacterial IID, and between age groups, as shown in Figure 7.1. Cases of viral IID in children started increasing in September and demonstrated two peaks; one in December and a second, larger peak in March. In contrast, infections in adults increased more steadily and consistently from October to result in a single peak in March. Cases in elderly (>65 years) were lower than children and adults throughout the year and showed only a small peak in March. The number of viral IID infections were lowest for all ages in the summer months. Bacterial IID showed a less clear seasonal trend with lower case numbers than viral IID. Adults aged 16-65 years had the greatest number of bacterial IID infections, with case numbers highest over the summer months and lowest during winter. Children also showed an increase in cases over the warmer months, although case numbers remained lower than adults throughout the year. In contrast to other age groups, a summer peak was not distinguishable amongst elderly adults (>65 years).

Figure 7.1: Time trends for IID, by age, for a) viral IID and b) bacterial IID, based on 4-week rolling average case numbers

a) Viral IID



b) Bacterial IID



The proportion of reported symptoms also differed across age groups (Table 7.3). More children experienced vomiting without diarrhoea and, for viral IID, more reported prolonged diarrhoea and vomiting, compared to adults. Of note, this was not the case for bacterial IID, where prolonged diarrhoea was less prevalent amongst children and more common in adult age groups. Adults aged 16-65 years reported more bloody diarrhoea with bacterial IID than other age groups, with 13.4% experiencing this symptom compared to 9.7% in over 65s, and 0% in children.

Table 7.3: Number and percentage of reported symptoms, by age and cause

Explanatory variables	Children <16yrs		Adults 16-65yrs		Elderly >65yrs	
	Viral n (%)	Bacterial n (%)	Viral n (%)	Bacterial n (%)	Viral n (%)	Bacterial n (%)
Diarrhoea, no vomiting	45 (18.7%)	19 (45.2%)	86 (35.4%)	100 (74.6%)	35 (38.0%)	45 (72.6%)
Diarrhoea >3days	108 (44.8%)	20 (47.6%)	45 (18.5%)	78 (58.2%)	32 (34.8%)	42 (67.7%)
Bloody diarrhoea	2 (0.8%)	0 (0.0%)	4 (1.6%)	18 (13.4%)	1 (1.1%)	6 (9.7%)
Vomiting, no diarrhoea	34 (14.1%)	4 (9.5%)	20 (8.2%)	1 (0.7%)	2 (2.2%)	0 (0.0%)
Vomiting >3days	49 (20.3%)	3 (7.1%)	7 (2.9%)	4 (3.0%)	2 (2.2%)	2 (3.2%)
Nausea	119 (49.4%)	16 (38.1%)	195 (80.2%)	85 (63.4%)	64 (69.6%)	30 (48.4%)
Nausea >3days	38 (15.8%)	6 (14.3%)	46 (18.9%)	36 (26.9%)	21 (22.8%)	9 (14.5%)
Abdominal pain	124 (51.5%)	27 (64.3%)	178 (73.3%)	113 (84.3%)	56 (60.9%)	44 (71.0%)
Loss of appetite	197 (81.7%)	30 (71.4%)	216 (88.9%)	107 (79.9%)	81 (88.0%)	46 (74.2%)
Fever	110 (45.6%)	25 (59.5%)	93 (38.3%)	69 (51.5%)	25 (27.2%)	19 (30.6%)
Headache	39 (16.2%)	16 (38.1%)	133 (54.7%)	76 (56.7%)	40 (43.5%)	25 (40.3%)
Cough, nose and throat	131 (54.4%)	21 (50.0%)	64 (26.3%)	29 (21.6%)	19 (20.7%)	15 (24.2%)
Total number of cases	**241 (80.9%)**	**42 (14.1%)**	**243 (62.0%)**	**134 (34.2%)**	**92 (59.7%)**	**62 (40.3%)**

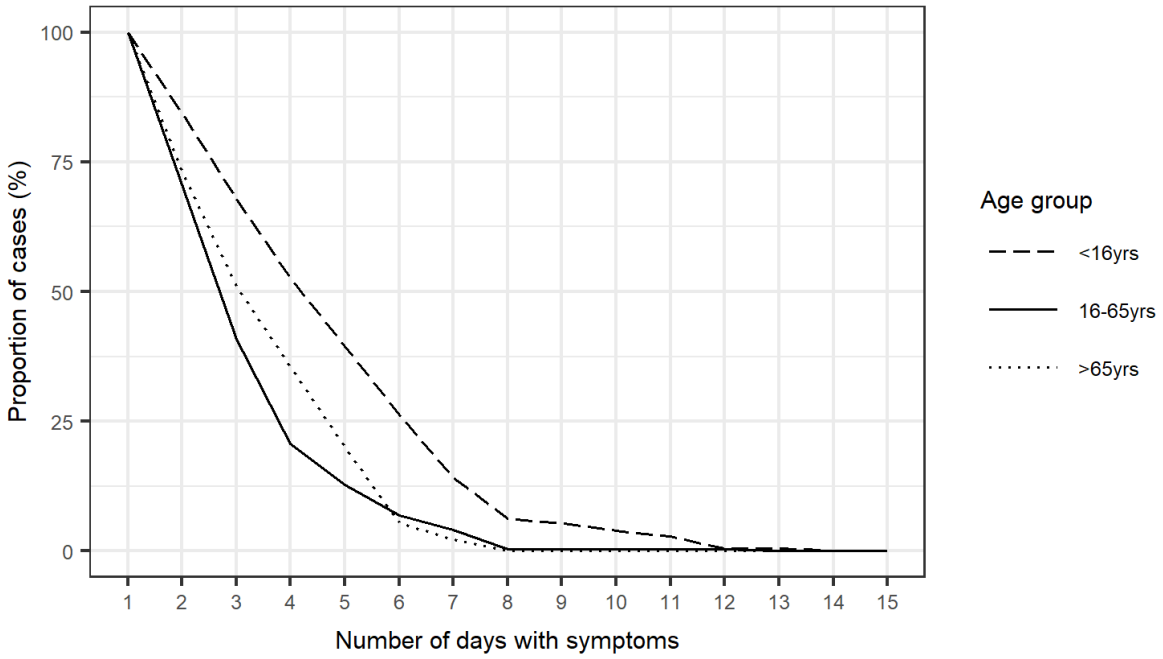
Adults also reported more nausea, abdominal pain, and headache. Elderly adults (>65 years) reported more episodes of prolonged diarrhoea with bacterial IID, with 67.7% of bacterial diarrhoea persisting for more than 3 days, but fever was a less common symptom in this age group compared to both children and younger adults.

The duration of diarrhoea in different age groups is shown in Figure 7.2. In contrast to adult age groups, children experienced more prolonged diarrhoea with viral pathogens, with a median duration of 4 days (IQR 2-6 days) compared to 2 days (IQR 1-3 days) and 3 days (IQR 1-4 days) for adults and elderly respectively, although this difference was not statistically significant. All age groups experienced a similar duration of bacterial diarrhoea, with a median duration of 4 days (IQR 2-5 days) for children, 4 days (IQR 2-6 days) for adults and 4.5 days (IQR 3-6 days) for elderly. Consequently, whilst both adult age groups showed a shorter median duration of diarrhoea with viral IID compared to bacterial, for children the duration of diarrhoea was comparable for both bacterial and viral organisms, and after 7 days 16% with bacterial IID still experienced diarrhoea, as did 14% with viral IID.

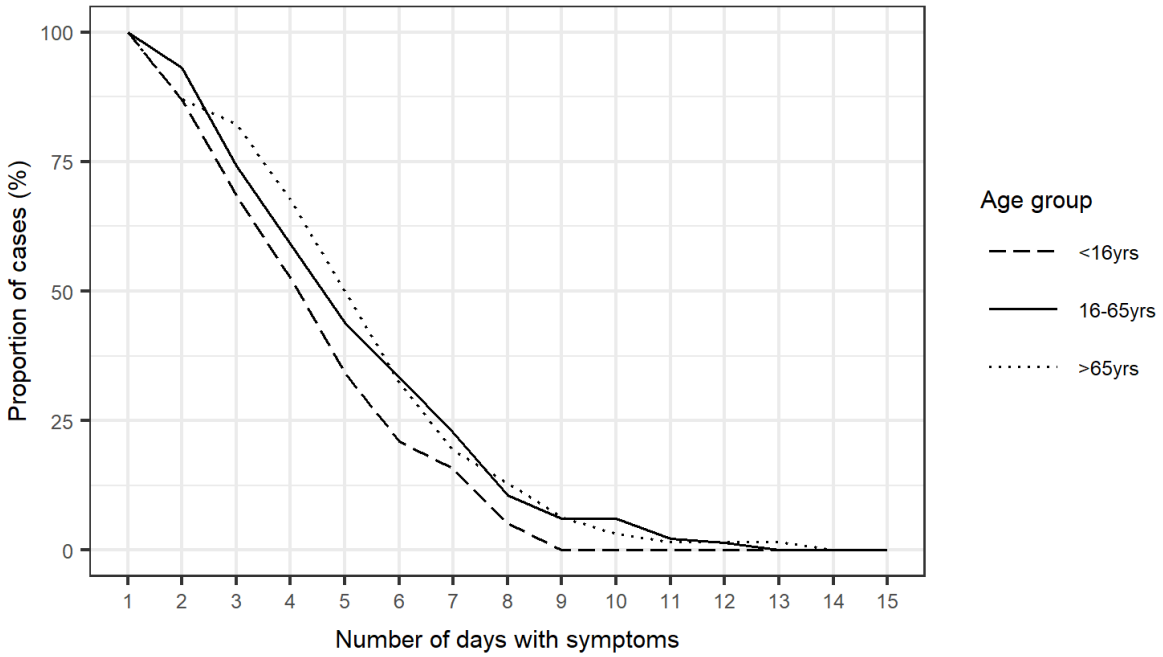
The duration of vomiting was shorter than for diarrhoea in all age groups, with median durations of 1-2 days for both bacterial and viral IID. Children had a longer duration of vomiting than the elderly age group (2 days compared to 1 day) for both bacterial and viral pathogens, although these were not statistically significant differences. The duration of nausea was similar across age groups and between bacterial and viral IID, with median durations of 2.5-3 days for bacterial organisms and 2-2.5 days for viral organisms.

Figure 7.2: Duration of diarrhoea by age group for a) viral IID and b) bacterial IID

a) Viral IID



b) Bacterial IID



Regression modelling

In children, the odds of disease caused by a bacterial organism were higher with diarrhoea in the absence of vomiting (OR 6.69, 95% CI 2.17-22.21), fever (OR 4.84, 95% CI 1.60-17.39) and onset in the summer months (OR 4.37, 95% CI 1.13-20.07). These symptoms were also associated with lower odds of a viral pathogen (Table 7.4). Abdominal pain was associated with increased odds of a bacterial pathogen, although the significance was borderline. Nausea was associated with lower odds of a bacterial cause and vomiting for more than 3 days increased the odds of viral IID, but neither finding was statistically significant.

Adults (16-65 years) had higher odds of a bacterial pathogen if they had diarrhoea but no vomiting (OR 3.04, 95% CI 1.63-5.74), bloody diarrhoea (OR 6.15, 95% CI 1.76-26.16) and diarrhoea lasting more than 3 days (OR 4.90, 95% CI 2.74-8.95). The odds of a bacterial IID were also increased in the presence of fever, and reduced if there was nausea, but these associations were not statistically significant. The odds of a viral pathogen in this age group were increased if onset was in winter (OR 4.82, 95% CI 1.93-12.42) or spring (OR 3.84, 95% CI 1.58-9.63) and there was a loss of appetite (OR 3.07, 95% CI 1.33-7.22).

For elderly adults (>65 years), the odds of bacterial IID were higher if there was diarrhoea but no vomiting (OR 3.20, 95% CI 1.44-7.30), and diarrhoea lasting more than 3 days (OR 3.31, 95% CI 1.49-7.59). These variables were also associated with lower odds of a viral pathogen. Bloody diarrhoea was associated with increased odds of bacterial IID, and loss of appetite with increased odds of viral IID, but neither were significant at the 5% level.

Mild to moderate correlations (phi coefficient <0.5) were identified between some symptoms in the different age groups, such as loss of appetite and nausea. For children, stronger correlations were identified between diarrhoea but no vomiting and nausea (phi = -0.55) and the duration of vomiting and duration of nausea (phi = 0.66). However, there was no evidence of multicollinearity within the regression models (VIF<5) and so all variables were included in the modelling.

Table 7.4: Multivariable logistic regression model outputs (odds ratio with 95% confidence intervals) for bacterial and viral pathogens, by age group.

Explanatory variable	Odds Ratio (95% confidence intervals)	
	Bacterial	Viral
Children (<16yrs)		
Onset in summer	4.37 (1.13-20.07)	0.29 (0.10-0.75)
Diarrhoea but no vomiting	6.69 (2.17-22.21)	0.24 (0.12-0.49)
Fever	4.84 (1.60-17.39)	0.43 (0.21-0.83)
Nausea	0.37 (0.11-1.26)	-
Abdominal pain	3.39 (1.01-13.92)	-
Vomiting lasting >3 days	-	2.59 (0.84-11.40)
Adults (16-65yrs)		
Onset in winter	0.31 (0.12-0.79)	4.82 (1.93-12.42)
Onset in spring	0.40 (0.16-0.97)	3.84 (1.58-9.63)
Diarrhoea but no vomiting	3.04 (1.63-5.74)	0.25 (0.14-0.44)
Bloody diarrhoea	6.15 (1.76-26.16)	0.21 (0.05-0.77)
Diarrhoea lasting >3 days	4.90 (2.74-8.95)	0.16 (0.09-0.28)
Loss of appetite	0.34 (0.14-0.79)	3.07 (1.33-7.22)
Fever	1.71 (0.93-3.21)	0.55 (0.29-1.01)
Nausea	0.55 (0.27-1.12)	-
Elderly (>65yrs)		
Diarrhoea but no vomiting	3.20 (1.44-7.30)	0.31 (0.14-0.69)
Bloody diarrhoea	7.38 (0.94-161.53)	0.14 (0.01-1.07)
Diarrhoea lasting >3 days	3.31 (1.49-7.59)	0.30 (0.13-0.67)
Loss of appetite	0.43 (0.14-1.22)	2.35 (0.82-7.03)

7.5 Discussion

This study has identified differences between the symptoms of bacterial and viral IID across different age groups. The odds of bacterial IID in children were increased by onset in the summer months, diarrhoea in the absence of vomiting, and fever. These symptoms were also associated with lower odds of a viral pathogen. In adults aged 16-65 years, diarrhoea but no vomiting, bloody diarrhoea, and diarrhoea lasting more than 3 days were associated with increased odds of a bacterial organism, whilst onset in the winter or spring and a loss of appetite were positively associated with viral IID. For elderly

age groups (>65 years) diarrhoea in the absence of vomiting and diarrhoea lasting more than 3 days were associated with higher odds of bacterial IID and lower odds of a viral cause. Only diarrhoea in the absence of vomiting emerged as a key symptom across all three age groups, associated with increased odds of a bacterial pathogen.

Seasonal patterns of IID are known to vary depending on the individual pathogen, but viral IID is generally associated with peaks in the winter and spring, and bacterial IID with peaks in the summer months.[245,250] However, this study suggests the association between seasonality and IID varies across age groups. Children experienced an increase in viral IID in the autumn, with an initial peak in late autumn/early winter, and a second larger peak in late winter/early spring. This is consistent with findings that cases of norovirus in children start earlier in the season than cases in adults.[100] Consequently, it was only during summer months that bacterial IID became the more likely cause of infection in a child. Whilst adults showed a more characteristic seasonality, with a large peak of viral IID over winter and spring, this was not the case for elderly adults (>65 years) who showed a less distinct seasonal trend. Furthermore, in the regression model, seasonality was not associated with the odds of either bacterial or viral IID in the elderly age group. Consequently, in our analysis season of onset was not a useful indicator to dissociate between bacterial and viral IID amongst elderly age groups.

Symptom profiling previously identified prolonged diarrhoea (lasting >3 days) as associated with increased odds of a bacterial pathogen,[250] and a shorter duration of diarrhoea has typically been associated with viral IID.[244] This study supports those findings for both adult age groups, but not for children. Children were found to have a longer median duration of diarrhoea in viral IID, compared to adult age groups, consistent with findings from other studies that children with norovirus, rotavirus and enteric adenovirus experience protracted diarrhoea.[23,217] However, this study also identified that children had a similar duration of diarrhoea with both bacterial and viral IID. This is not in keeping with findings of other studies which describe longer mean durations of diarrhoea for bacterial aetiologies in children.[23] One possible explanation is the setting of the study. This study used data from a community cohort and as such will capture milder illness which would otherwise be missed from studies based on hospital admissions or outpatient consultations. As this study identified little difference in the duration of bacterial and viral diarrhoea in children, the duration of diarrhoea may be an unhelpful indicator of underlying cause in children.

The prevalence of reported symptoms may also have an impact on the feasibility of using different symptoms in the development of symptom profiles. Bloody diarrhoea has typically been associated

with bacterial IID,[23,243,244,250] however in this study it was uncommon across all age groups, affecting only 2 children, 22 adults and 7 elderly cases. The low prevalence in this cohort made it a less useful indicator of bacterial IID in our analysis, especially amongst children and elderly. Nausea, headache, and abdominal pain were also less common in children and amongst those aged under 2 years there was a larger proportion of missing data for these variables. This is most likely because these symptoms are self-reported, rather than observed, and so cannot be identified in infants and very young children. Consequently, symptom profiles which incorporate these variables may be of limited value for children and adults who are non-verbal and unable to report the presence or absence of these symptoms.

The findings of this study also have implications for the use of epidemiological criteria in outbreak scenarios. Such criteria have used the diarrhoea-to-vomiting ratio, fever-to-vomiting ratio, and the average duration of illness to indicate the likely cause of an IID outbreak.[153–155] However, the finding that symptom profiles vary across age groups could make these epidemiological criteria less applicable to certain settings. For example, the short duration of diarrhoea to indicate norovirus infection may be less appliable to outbreaks in childcare and school settings. Likewise, the use of fever to indicate a bacterial pathogen may not be applicable to elderly care home outbreaks, as fever was less reported in this age group and did not discriminate between bacterial and viral aetiologies. Further work is needed to explore whether these criteria alter between different outbreak settings.

Limitations

This study provides a valuable comparison of different symptoms of IID across age groups. However, by undertaking a segmented analysis, the sample sizes were reduced, resulting in larger confidence intervals and less certainty around the estimates. To prevent the group sizes becoming too small to allow meaningful analysis, the age bands were split broadly into children, adults, and elderly. However, it is recognised that within these groups there may be further symptom variation according to age. This is particularly relevant to the extremes of age, with the very young and very old likely to experience IID differently to others within the broader categories of children and elderly. Unfortunately, the data utilised in this study did not allow this further breakdown by age, but future work should consider exploring these age categories further.

An additional limitation is the use of the broad categories of bacterial and viral IID, rather than developing organism-specific symptom profiles. The analysis was undertaken in this way due to the small number of individual organisms within each age band, which made further breakdown by

pathogen unfeasible. However, it should be considered whether the symptom profiles identified in this study could be attributed to different organisms predominating across the three age groups. This is particularly relevant to rotavirus, which disproportionately affects children and typically causes milder illness in adults. Nevertheless, norovirus remained the most common viral pathogen in children and adults alike and *Campylobacter* was the most prevalent bacterial organism in all three age groups. Further study should consider whether pathogen-specific symptom profiles could be developed across the different age groups.

7.6 Conclusion

Symptom profiles for bacterial and viral IID vary across age groups, with evidence of differing seasonality and prevalence of symptoms. Only diarrhoea in the absence of vomiting emerged as a key symptom across all three age groups, associated with increased odds of a bacterial pathogen. These findings have implications for clinicians, public health specialists and epidemiologists who use symptoms to guide presumptive diagnoses in the absence of microbiological confirmation. They may also impact on the use of epidemiological criteria in outbreak situations, as outbreaks in schools and childcare settings may exhibit different symptoms to those in elderly care home settings.

Chapter 8. Discussion

8.1 Summary

This chapter is a discussion of the main findings of this thesis. The results are discussed in terms of their contribution to our knowledge and understanding of the burden of infectious intestinal disease (IID) in schools and the potential role and utility of school attendance registers in the surveillance of IID in children. The strengths and limitations of this work will be considered as well as the implications of this thesis on future public health research, policy, and practice.

8.2 Key findings

Chapter 2 identified factors which increased the risk of outbreaks in schools and identified IID as a key cause of reported outbreaks, accounting for nearly half of all school outbreaks. IID, as well as influenza, were also associated with higher attack rates compared to rashes and other causes. Primary schools and all-through schools were at increased risk of experiencing outbreaks, as were larger schools, and are therefore key settings for health protection interventions. However, as schools do not have a statutory duty to report outbreaks it is likely many outbreaks go unreported and are not, therefore, captured within current surveillance data. This raised the question as to whether a school-based surveillance system could be used to improve the detection of cases and outbreaks of illness amongst children, removing the reporting bias inherent in existing data.

Chapter 3 was a systematic review exploring the utility of school-based surveillance systems, focused on the role of school attendance registers in the surveillance of infectious disease outbreaks and occurrences amongst children. However, only studies concerning the surveillance of influenza could be found. The systematic review identified several features of such systems which could impact on the utility of these data in surveillance. The first was the specificity of the data collected, with syndrome-specific absences correlating better with other surveillance indicators than illness absence or all-cause absence. Similarly, they also provided more lead time than the other school absence indicators. The second was the use of absence thresholds to trigger alerts, either at an individual school level or at a health authority level. Such thresholds required a trade-off between accuracy and timeliness, with lower thresholds triggering earlier but correlating less well with existing surveillance indicators. The use of absence thresholds would have to be carefully investigated to identify the level at which alerts gave sufficient early warning of outbreaks or rising cases amongst children whilst minimising false positives. Thirdly, the frequency of data submissions from schools could impact on the value of school

absence data as an early warning system. The potential lead time of school absences was 1-2 weeks and so weekly data submissions may not be sufficient to provide advanced warning. Instead, daily data submissions should be considered to enhance the utility of these data for surveillance purposes. Finally, the consistency of data collection across and between areas was important for allowing data to be aggregated. If such a system was going to be used routinely for surveillance in the UK, it would need to be standardised across all schools for optimal benefit.

The finding that school absences had been used in the surveillance of influenza, combined with the high prevalence of IID outbreaks in schools, prompted the question as to the potential benefit of using school attendance registers in the surveillance of IID. Chapter 4 sought to collect and analyse school illness absences in Merseyside, UK to assess their correlation and lead time with other IID surveillance indicators. Unfortunately, insufficient data were received from Local Government to allow an in-depth analysis of this dataset, but a broad assessment of seasonal trends identified a positive association between school illness absences and viral IID laboratory reports, GP consultations and telehealth calls for children. A survey sent out to teachers identified several schools who collected information on the symptoms of illness when an illness absence was reported, although only a small number recorded these data electronically. However, such schools may provide a model for data collection which could be rolled out to schools more widely.

A key limitation of the above analysis was not only the small amount of school data received, but also the low number of cases reported in school-aged children within the surveillance data, when broken down by Local Government area. Consequently, further analysis was undertaken using national surveillance data to explore whether school IID outbreaks provided a lead time ahead of other IID surveillance indicators. This study, in Chapter 5, identified that school outbreaks provided a 3-week lead time ahead of care home and hospital outbreaks, as well as lead times ahead of norovirus laboratory reports and NHS 111 calls for vomiting and diarrhoea. Within the latter two datasets, children provided a lead time ahead of adults. Cases in children were also found to start increasing earlier in the season than adults, whilst school outbreaks increased earlier in five out of nine surveillance years. These findings would suggest that monitoring cases and outbreaks in children could provide an early warning of seasonal norovirus infection and, whilst it could not be demonstrated in this thesis, school attendance registers could be useful way of improving case detection in this age group.

However, many surveillance datasets are based on the broad syndrome of diarrhoea and vomiting. This makes it difficult to distinguish norovirus from other causes of IID and prompted the question as

to whether symptoms alone, in the absence of microbiological sampling, could be used to distinguish between different IID organisms. Chapter 6 explored this concept and identified that bacterial and viral IID were associated with different symptom profiles, which could be used to discriminate between them. However, symptom profiles were not identified which could be used to adequately distinguish individual pathogens and consequently assumptions regarding the causative organism should not be made based on symptoms alone. Such profiles could be used in syndromic surveillance systems, but also potentially in schools to improve the specificity of illness absence data.

To explore the application of these profiles to school settings, an additional analysis was undertaken on the same dataset to explore whether symptom profiles varied across different age groups (Chapter 7). This study identified differences between the symptoms of bacterial and viral IID in children, adults, and elderly, with variations in the prevalence and duration of certain symptoms and differing seasonal trends. Only diarrhoea in the absence of vomiting emerged as a key symptom across all three age groups, associated with increased odds of a bacterial pathogen. These findings not only impact on the use of symptom profiling in different age groups but could also have implications for the utility of epidemiological criteria in outbreak situations, as outbreaks in schools and childcare settings may exhibit different patterns of symptoms to those in elderly care home settings.

The findings of this thesis not only highlight the importance of IID as a cause of outbreaks in schools but also the potential for school outbreak data to be used as an early warning of the norovirus season before outbreaks start increasing in care homes and hospitals. There may be a role for school attendance registers in improving case and outbreak detection amongst children, however school attendance data in the UK in its current format is unable to distinguish IID from other causes of illness. Enhancing data collection in schools to provide syndrome-specific absences could significantly improve the utility of this dataset as a surveillance indicator. The use of symptom profiling has the potential to allow bacterial pathogens to be discriminated from viral IID, which could further enhance the specificity of surveillance data.

8.3 Research themes

Burden of IID in schools and school-aged children

Previous studies have identified schools as key settings for IID outbreaks, accounting for 10% of all IID outbreaks in the United States, 74% of shigellosis outbreaks and 5-15% of norovirus outbreaks.[62,105] The analysis of national outbreak data in Chapter 5 similarly found that schools accounted for 10% of reported outbreaks, whilst 33% occurred in hospitals and 57% in care homes. Children bore a

significant burden of IID infection, accounting for a third of GP consultations for IID and nearly half of all calls to NHS 111 for vomiting.

In the introduction, I identified rashes, IID and respiratory infections as common causes of outbreaks in schools.[101–104] However, the relative burden of IID in schools compared to other pathogens could not be identified from existing literature. In Chapter 2, I explored the causes of school outbreaks which had been reported to Public Health England over a two-year period and identified IID as the most common cause, accounting for 47% of school outbreaks. This was followed by rash (44%) and influenza (6%), with other causes accounting for less than 1% of reported school outbreaks each. However, the voluntary reporting of outbreaks by schools creates a potential source of bias when using these data to explore the burden of different syndromes in schools. The Public Health England guidance to exclude pupils and staff with diarrhoea and vomiting could lead to more absences and greater disruption in schools, making IID outbreaks more noticeable. Consequently, these outbreaks may be more likely to be reported to Public Health England. In contrast, the guidance for influenza does not include an exclusion period and potentially many will still attend school with milder symptoms. This would reduce the impact of an influenza outbreak on the day-to-day functioning of a school and could result in such outbreaks being under-recognised and under-reported compared to IID outbreaks. This reporting bias means influenza outbreaks may be under-represented within the existing surveillance data, compared to IID, and the incidence of influenza outbreaks in schools may be much greater than these data currently suggest. This would be consistent with the survey findings in Chapter 4, with influenza-like-illness (ILI) identified as causing the most illness in schools, whilst diarrhoea and vomiting was highlighted as causing more absences.

Chapter 2 also explored the attack rate in school outbreaks. The background literature suggested a mean attack rate of 28-30% in school outbreaks, with a median of 40-42 people affected per outbreak.[105–107] Within my analysis, outbreaks ranged in size from 2 to 300 cases, with a median of 10 cases per outbreak. IID outbreaks had a higher median case number of 20 cases per outbreak (IQR 11-33). Attack rates ranged from 0.1% to 74%, with a median attack rate of 3.6%. Attack rates varied depending on the cause of the outbreak, with the highest median attack rate occurring for IID outbreaks (8.3%) and the lowest for rash (1.1%). The estimates for both outbreak size and attack rates in this thesis are substantially lower than those identified within the background literature. This could be due to publication bias, with larger scale outbreaks more likely to be reported and published than smaller outbreaks. However, even Public Health England data are subject to investigation and reporting bias, and it is likely that outbreaks which go unreported are those with smaller case numbers

which do not cause significant disruption to the school. Consequently, the true attack rate in schools may be even lower than my findings suggest.

The background literature suggested a difference between school settings in terms of the burden of outbreaks and subsequent attack rates, based on factors such as pupil age and school size. [56,102,108] The analysis in Chapter 2 identified primary and all-through schools, as well as larger schools, as being at increased risk of outbreaks. Attack rates were found to be higher in primary schools, compared to secondary schools, but this finding was not statistically significant for IID or all-cause outbreaks. Interestingly, attack rates were also higher in smaller schools, although the median number of cases per outbreak did not vary significantly between schools of different sizes. This suggests the higher attack rate in small schools was driven by a smaller denominator rather than a higher number of cases. This raises the question as to what factors influence disease transmission in school settings and whether class size, rather than school size, is a more important determinant of case numbers within outbreaks. Studies exploring social contact networks suggest the majority of children's close contacts in school are within their immediate class [77,81] and these contacts are likely to have a higher potential for disease transmission than other contacts within the school environment.

The finding that primary schools were disproportionately affected by outbreaks is consistent with existing literature.[56,102,166,177] Chapter 4 also found that primary-aged children were responsible for a larger proportion of GP consultations and telehealth calls for IID, compared to secondary-aged children. However, this finding was not reflected in illness absence data (Chapter 4), with secondary schools having a higher median weekly rate of illness absence compared to primary schools (2.79% and 2.54% respectively) and a higher peak absence (4.46% and 3.98% respectively). Given the cause of illness absence cannot be determined from the data, a possible explanation for this finding is that illness absences in the two age groups are being driven by different causes. It may be that primary-aged pupils are more greatly affected by infectious illness, such as IID, which can result in outbreaks. In contrast, the higher illness absence rate in older pupils may be driven by non-infectious causes. The school survey in Chapter 4 would support this hypothesis, as secondary schools highlighted mental health, migraines, and chronic health conditions as common causes of illness absence amongst their pupils. In contrast, only diarrhoea and vomiting, and influenza-like-illness were identified as important causes of illness absence by the primary schools who responded to the survey.

Utility of school attendance registers in IID surveillance

In the introduction, I highlighted that current surveillance systems in the UK are primarily based on healthcare data and, as many cases of IID in school-aged children will be managed conservatively at home, are likely to significantly underestimate the impact of IID in children. Likewise, outbreaks in schools may go unrecognised and under-reported to public health teams.[135–137] School attendance registers offer a novel dataset which could be used to identify community cases of IID which might not otherwise be detected. I identified studies which had explored the use of these data to enhance the community surveillance of influenza, but none which considered their role in IID surveillance.

The systematic review in Chapter 3 provided a more detailed and thorough review of the literature surrounding school attendance registers and highlighted some important features which could affect the utility of these data in IID surveillance. The first was the specificity of the data, with syndrome-specific absences found to correlate better with other surveillance indicators and provide a more consistent lead-time. However, in the UK school absence data does not include the cause of illness and consequently it is not possible to distinguish IID from other syndromes. The challenges of using these non-specific data were highlighted in Chapter 4, where a winter peak in absences was observed to coincide with a peak in national norovirus cases. However, as influenza would also be circulating at that time, it cannot be known whether norovirus or influenza were responsible for the illness absence in children. The modelling of seasonal trends in different organisms, to apportion likely cause to peaks in illness absence, is problematic because of the limited laboratory testing done in school-aged children. As shown in Chapter 5 and Chapter 7 of this thesis, seasonal trends of IID may differ between age groups and consequently using cases in adults to model illness absence in schools could lead to an incorrect assumption that peaks of illness in adults are reflected in children. Furthermore, milder strains of pathogens such as influenza may go undetected by current surveillance systems and yet can still cause a significant number of outbreaks in schools.[102] Modelling surveillance data, even confirmed cases in children, could still risk ascribing peaks in illness absence to an incorrect cause, simply because current data are inadequate to detect milder illness within the community. Syndrome-specific absences would remove the need to model school data against trends in organisms to identify likely causes of absence. Such data are likely to be a more useful addition to syndromic surveillance in the UK. A key recommendation for future research is the need to explore the collection and application of syndrome-specific attendance data.

The school survey in Chapter 4 identified several schools which already collected data on symptoms of illness absence and working with these schools could provide a dataset for future research in this area. However, schools differed in terms of the exact nature of the information collected and how that

information was subsequently recorded. A standardised approach to collecting these data would need to be developed before this dataset could be considered for surveillance purposes. This could include having pre-determined categories of illness to distinguish between key syndromes. The recording of these data would also need to be considered, since not all schools were recording symptom data electronically. However, it also needs to be remembered that schools are not healthcare institutions and may have limited capacity to engage with health-related issues unless they directly impact on academic outcomes and targets. The success of a school-based surveillance system is likely to be improved if the collection and dissemination of attendance data can be streamlined using existing electronic attendance systems, minimising the additional workload for schools.

Another factor which could affect the utility of school attendance registers in IID surveillance is the frequency of data submissions. In the systematic review (Chapter 3), studies collected either daily or weekly attendance data, although weekly reports often contained details of daily absences. If data are being analysed retrospectively for research purposes the frequency of data submissions has less relevance, but it could have a significant bearing on its application to real-time surveillance. In the UK, schools record absences electronically both morning and afternoon, however discussions with Local Government revealed that the data upload does not necessarily happen daily, and many schools submit data to the Local Government once weekly. Furthermore, these data can still be altered and updated by schools with the final validated attendance data being submitted at the end of each term. Across the three Local Government areas involved in this research, each had a different arrangement in place with local schools regarding the timeliness of data submissions, which varied between daily, three times a week and weekly submissions. Using these data for real-time surveillance across a large footprint would require a standardised approach to the frequency of data submissions. Whether daily or weekly data submissions would be required would depend on the lead time offered by this dataset.

The systematic review (Chapter 3) suggested syndrome-specific absences for influenza peaked concurrently or 1-2 weeks ahead of other surveillance indicators. In comparison, studies using illness absence data showed a mixed picture of lead times, lag times and concurrent peaks. Analysing illness absence data in Chapter 4 revealed some of the limitations of using these data to explore lead or lag times. As the trends in illness absence can be caused by any organism, the timings of peaks compared to surveillance data could represent a peak in a different illness rather than a true lead or lag time. Chapter 5 explored lead and lag times using syndrome-specific outbreaks in schools. Whilst the exact organism was not confirmed, it could at least be distinguished from other causes of outbreaks such as influenza. This study found that school IID outbreaks occurred 2-3 weeks ahead of outbreaks in care homes and hospitals, NHS 111 calls for diarrhoea and vomiting, and laboratory norovirus reports.

Whilst it cannot be assumed that school absences for IID would provide the same lead time, it should be considered whether this would provide sufficient warning to allow additional protective measures to be put in place within other settings.

A final consideration for the utility of school-based data is its ability to link with other relevant datasets. Linking data across datasets requires a common identifier and for healthcare data fields such as the patient's name, date of birth, NHS number and address can be used to identify an individual patient. However, linking with school data requires different identifiers, which can be easily overlooked by public health teams and surveillance systems which are primarily focused on healthcare data. In Chapter 2, I linked school outbreaks to data from both the Department for Education and Ofsted to analyse whether school-level factors impacted on the risk of outbreaks occurring. School outbreak data commonly contains details of the name of the school and postcode. However, data linkage cannot be undertaken on these fields, as multiple schools in different areas may share the same name. Furthermore, the use of an abbreviation or a misspelling of the name by the person recording the data would make it impossible to identify the school by name alone. Postcode also cannot be used, as multiple schools, especially infant and junior schools may share the same site and the postcodes recorded on school websites may differ from the official postcode registered with the Department for Education. For the data linkage in Chapter 2, I used each school's Unique Reference Number (URN), which all schools are given when they are first established or when they convert to a different type of school. Linking data using a URN will not only correctly identify the school but will also ensure that it links to information relevant to the current status of the school. Unfortunately, public health teams do not record the school's URN when an outbreak is declared and consequently, I had to manually identify the URN for each school within the dataset. This process was time-consuming and limited the amount of data I was able to analyse. Including a URN for each school which reports an outbreak would significantly improve the utility of this dataset and allow outbreak data to be easily linked to other datasets thus providing additional data fields, such as the age range of the school and the postcode, which may be of value when analysing these data for surveillance purposes. This should also be considered for other school-based data, such as school attendance registers. Whilst I did not attempt to link school attendance data to other datasets within Chapter 4, ensuring a URN is available for each included school could enhance the scope of analyses possible with these data and consequently improve their application and utility in health surveillance.

Children as an early warning of seasonal IID infections

In the introduction, I explored the role of children as community transmitters of infection. For diseases which spread through close contact, social networks are an important way of understanding how diseases move through the population. Children have many close contacts at school and a high number of contact-hours at home.[81] Consequently, infections and outbreaks in schools can readily be transmitted into households and adult age groups. Children are already thought to play an important role in the spread of influenza and the childhood influenza vaccination programme has been shown to provide indirect protection to older, unvaccinated age groups.[92,93] Their role in the transmission of IID is less clear although modelling has suggested that paediatric norovirus vaccination could prevent 18-21 times more cases than elderly vaccination by providing both direct protection to children and indirect protection to adults.[133]

The findings of the systematic review in Chapter 3 would support the role of children in the transmission of influenza, with influenza-specific school absences peaking concurrently or 1-2 weeks ahead of other surveillance indicators. Influenza absences were also found to provide a lead time on the start, peak and endpoint of the H1N1 pandemic,[211] which would suggest that children are among the first affected by circulating influenza. Chapter 5 sought to explore this concept for IID by assessing whether outbreaks in schools occurred before outbreaks in care homes and hospitals. The study identified that school outbreaks occurred 3 weeks before outbreaks in other settings, which would suggest that IID starts circulating amongst younger age groups before spreading into care homes and hospitals. A descriptive analysis of norovirus laboratory reports also showed an earlier increase in cases amongst children, with a later rise in adult age groups. This is comparable to the findings of another study, exploring norovirus surveillance data in Germany.[100] Whilst the school outbreaks included in Chapter 5 were not norovirus-confirmed, norovirus spreads easily within semi-enclosed environments and has been identified as the predominant cause of school outbreaks.[56,62] Therefore, school outbreaks may be better attributed to seasonal norovirus, rather than any other IID organism.

Chapter 5 also investigated whether cases in children preceded those in adults within existing surveillance datasets. Seasonal trends in cases of IID amongst children provided a lead time ahead of adults for laboratory norovirus reports, and NHS 111 calls for both vomiting and diarrhoea, but occurred concurrently with adults for GP consultations. Breakpoint analysis revealed an earlier seasonal increase in cases amongst children compared to adults in all four surveillance indicators across the study period, except for GP consultations in one out of the six seasons studied. Whilst this provides further evidence of the role of children in the transmission of IID, the use of syndromic surveillance data for this analysis meant that norovirus could not be distinguished from other IID

pathogens within GP and NHS 111 data and therefore it is not clear whether this finding should be attributed to norovirus alone or all IID. It is unlikely that children have the same role in the transmission of different IID pathogens. Animals are the major reservoir of bacterial IID and consequently transmission is primarily from contaminated food or water, or directly from an infected animal.[53,68,69] In contrast, humans are the natural reservoir of viral IID and the most common pathway of transmission is via person-to-person contact.[34,62] Consequently, the role of children in the transmission of viral IID is likely to be of greater importance. Analysing trends in syndromic data could mask differences between organisms and make it difficult to draw firm conclusions regarding any single IID pathogen.

There may, however, be differences between the syndromic indicators in terms of the organisms and age groups they are more likely to capture. Previous research suggests that there are 23 norovirus cases in the community for every one which presents to the GP.[29] In comparison, 1 in 7 people with *Campylobacter*, and 1 in 3 people with *Salmonella,* consulted their GP.[29] Consequently, GP consultation data may be of greater value in the surveillance of bacterial IID and a less sensitive indicator of norovirus in the community. Additionally, there is evidence in the literature that vomiting may be a more prevalent feature of norovirus amongst children whilst diarrhoea is more common amongst adults.[154,235] Consequently, NHS 111 calls for vomiting may be a better indicator of norovirus amongst children than calls for diarrhoea. This could explain why, in Chapter 5, trends in school outbreaks were found to correlate better with NHS 111 vomiting calls than diarrhoea calls (r_s 0.80 and r_s 0.59 respectively).

The findings of Chapter 5, when considered in the context of existing literature, would suggest that cases in children provide an early warning of seasonal norovirus infection. Whether this finding is applicable to all viral IID pathogens or only to norovirus is difficult to assess. The number of viral isolates for each organism within routine laboratory data is limited (Table 4.2), making it difficult to assess and compare trends across different age groups. Even the number of cases from the Second Study of Infectious Intestinal Disease in the Community (IID2 Study), which conducted more extensive testing and typing of symptomatic cases, were too small to allow an analysis of individual pathogens (Table 7.2). Testing within community outbreaks is also scarce, so whilst norovirus might be the most likely cause, this is an assumption which is often unconfirmed. Enhancing testing in community outbreaks to confirm the causative organism could improve the utility of outbreak data for surveillance purposes. The potential benefits of using data on cases and outbreaks of IID in children to provide an early warning of seasonal norovirus warrants further investigation. Future study should explore whether the lead time offered by these data is sufficient to put protective interventions in place within

other settings and whether implementing these interventions earlier is effective at reducing the rate of norovirus infection in adult age groups.

Enhancing syndromic surveillance

Surveillance data were used throughout this thesis to explore infectious intestinal disease in schools and school-aged children. In Chapter 2, surveillance data on IID outbreaks in schools were used to identify school-level risk factors for outbreak occurrence. In Chapter 4, laboratory, GP and telehealth data were used to assess the potential utility of school absenteeism data as a surveillance indicator. Chapter 5 combined surveillance data on outbreaks, laboratory reports, GP consultations and telehealth calls to investigate whether children provided a lead time ahead of adults for seasonal IID infection. Surveillance data has the advantage of covering a large population group over a prolonged period, providing more cases and a larger sample size than would be possible from a prospective cohort study. However, with the exception of laboratory data, most surveillance indicators are based on syndromic definitions, which capture wide-ranging causes of diarrhoea and vomiting. Those which are based on clinical diagnoses, such as GP data, can distinguish a likely infective cause from a non-infective cause, but this will still capture a broad range of pathogens which vary in terms of their source, pathways of transmission and control.

A recurring theme within this thesis is the need for greater specificity within the data. The systematic review in Chapter 3 highlighted the value of using syndrome-specific absences compared to broader absence definitions. Chapter 4 also demonstrated the limitations of using all-cause illness absence data, as it was unable to distinguish between norovirus and influenza, both of which are important causes of absence, prevalent amongst children, and demonstrate similar seasonal trends. The analysis and interpretation in Chapter 5 was affected by being unable to distinguish bacterial and viral IID, which made it difficult to determine the role of children in the spread of different IID infections.

Both Chapter 6 and Chapter 7 sought to address this issue by exploring whether symptoms alone could be used to make inferences about the causative organism, in the absence of laboratory testing. Chapter 6 identified different symptom profiles for both bacterial and viral IID, which could be used to enhance syndromic surveillance. Whilst there was insufficient data to explore symptom profiles for individual pathogens, given the fundamental differences between bacterial and viral IID in terms of their reservoirs and transmission pathways, this would be a valuable distinction to be able to make within the data. This would have enhanced the analysis in Chapter 5 and enabled seasonal trends in

viral and bacterial IID to be discerned. It could also have added to the analysis of outbreak data in Chapter 2, as school-level risk factors may differ for viral and bacterial IID outbreaks.

A further application of symptom profiles would be to enhance school illness absence data and with that consideration, Chapter 7 explored whether profiles varied by age group. Important differences were found between symptoms in children, adults and elderly which would need to be taken into account if such profiles were used for syndromic surveillance. The only symptom which distinguished bacterial and viral IID in all three age groups was diarrhoea in the absence of vomiting. Whilst single symptoms are unlikely to be as accurate as profiles based on multiple symptoms, they have the advantage of being simple and easy to apply in non-clinical settings. The school survey in Chapter 4 explored whether symptom data was collected by schools for each illness absence. Of the schools who did ask for further information, the main cause or a brief description of illness were commonly sought. Asking about the presence or absence of both diarrhoea and vomiting is likely to be easily incorporated into the information already gathered and would not only distinguish IID from other causes of illness, but also indicate the broad category of bacterial or viral IID. The use of more detailed symptom profiles, incorporating symptoms such as nausea, abdominal pain and loss of appetite, combined with the season of illness, are likely to be too complex for schools to collect for each reported absence.

Enhancing syndromic surveillance to be able to distinguish bacterial and viral IID would improve the utility of surveillance datasets and our ability to monitor the epidemiology of IID infections. Limited case numbers of individual pathogens meant that symptom profiles could not be explored for separate organisms, but such profiles could also be of benefit to monitor key pathogens such as norovirus, rotavirus and *Campylobacter*. However, symptom profiles must balance sensitivity and specificity with feasibility, and complex profiles are unlikely to be practical within the time constraints of many services which provide surveillance data. This is further complicated by the finding that profiles alter across age groups, which could result in complex algorithms which combine age and season with a list of reported symptoms. Simple case definitions which apply across age groups, such as diarrhoea in the absence of vomiting, have the advantage of being easy to use and could be incorporated into data collection from non-clinical, community settings such as schools.

8.4 Strengths and limitations

Acquiring school attendance data

One of the main challenges of this research was obtaining school attendance data from Local Government. All three Local Government areas involved were interested and supportive of the

research, however there were multiple issues in terms of accessing and extracting the data. This was made more complicated by the variation in data collection and data handling between the Local Government areas. All schools in the area used SIMS (School Information Management System) to electronically record attendance data. The attendance data was then transferred via a B2B data transfer to the Capita One system. The Capita One system could be accessed by Local Government and was primarily used to search for attendance data for individual pupils, which was particularly relevant for Social Services who monitor the attendance of at-risk children in their area. The Capita One system also generated some simple reports based on aggregated attendance data. The only statutory requirement for school attendance data was a termly submission to the Department for Education and it was these figures which were used by Local Government to monitor performance.

In discussion with Local Government, I explored several different ways of accessing weekly attendance data for individual schools. The first was whether the Capita One system could be searched at school level, but unfortunately the system did not have this functionality and could only be searched by individual pupil. The second was whether the raw data could be downloaded from Capita One and then aggregated by school to ensure anonymity. But here I found a gap in the knowledge of the data flow between schools and Local Government, and it was generally unknown who processed the raw data or how that data could be accessed beyond the standard reports the system already generated. Only one Local Government area could name their IT supplier but could not provide a contact for me to follow up with. The third approach was to explore whether the reports generated by Capita One could be altered to include a summary of weekly attendance, broken down by school. However, the only way of pursuing this was to put in a request via the service desk, which I was unable to do outside of the Local Government system. This put the onus on the Local Government to be willing to make the request and follow up any issues with Capita One. Only one Local Government area was willing to do this, with another willing to adopt the same approach only if it was successful in their neighbouring area.

The one Local Government area that took this forward managed to extract raw attendance data for two academic years, broken down by school. However, this was a complex and time-consuming process due to the limitations of the computer systems, which struggled to run complex queries without crashing. I had originally asked for a breakdown by year group and sex, however the system was unable to handle the additional breakdown and ultimately these data could not be extracted. Due to competing priorities, there was a limit to the time that could be spared by those within Local Government to attempt to resolve these issues. With the benefit of hindsight, I wonder whether the additional time spent attempting to extract data at year group level meant fewer years of data were

ultimately provided, since those involved reached a point where they were unable to give any more time to this research. Perhaps had I not asked for the additional detail within the data they would have been able to provide more years of historical attendance data which ultimately would have enabled a more detailed analysis to be undertaken in Chapter 4. This highlights one of the challenges of working with third parties to access data, as organisations and individuals are required to undertake additional work which may not align with their own priorities. This becomes especially problematic if data extraction is complex or time consuming. This also creates a tension for the researcher, as the research question may not be answerable without a certain level of detail within data. However, requesting too much data or a breakdown which is too complex may act as a barrier for organisations wanting to work with and engage in research.

Reporting of illness absence in schools

An important critique of using school illness absence data in surveillance is the reliability and accuracy of the information received from parents. Illness absence is considered an authorised absence within schools and consequently pupils who are absent for unauthorised reasons, such as holidays during term time or a pupil refusing to attend, may be falsely reported as being unwell. The selective information received from parents was highlighted by one school as the main reason they thought collecting data on symptoms of illness would not be beneficial for schools (Chapter 4). In the discussion of that chapter, I considered the implications of false illness reports on the utility of these data in surveillance. I concluded that whilst these inaccuracies were likely to lead to an over-estimate of the burden of illness amongst children, if they occurred randomly throughout the year, such reports should not greatly affect the overall trends which would be used for detecting peaks and outbreaks of seasonal illness. If false-absence reports are patterned, for example in the weeks surrounding school holidays, analysing these trends over multiple years would allow them to be considered when modelling school absences.

Similar concerns regarding the reliability of parental reports were raised by members of the Public and Patient Involvement (PPI) panel who reviewed the school survey prior to it being disseminated. Members noted that teachers must rely on the honesty of parents, and this could affect the integrity of the data for research purposes. However, one panel member who had been a headteacher, Ofsted inspector and a Governor of a school provided some interesting insights from the perspective of the school. They commented that whilst the information received from the parent may be false, the school has little incentive to challenge this, as unauthorised absences count against a school during an Ofsted inspection and consequently could affect their chances of getting a good rating. Therefore, schools

may be reluctant to pursue the cause of illness too rigorously. This could make some schools hesitant to engage in research in this area and may have contributed to the low response rate to the survey in Chapter 4. However, it was also noted that this did not apply to outbreaks, in which schools are more likely to be proactive about addressing the cause of illness and ensuring parents have the necessary information.

If school attendance data are to be considered for health surveillance, it would be beneficial to engage more closely with schools and parents to explore influencing factors around absence reporting. This may be an important step in encouraging schools to engage with research in this area, as any underlying concerns can be addressed and where appropriate, reassurance given that these data will not be used to reflect the performance of an individual school. This forms the basis for one of my recommendations for further research in this area.

Utilising secondary datasets

This thesis used multiple secondary datasets, including surveillance data on laboratory reports, GP consultations and telehealth calls (Chapters 4 and 5), data on reported outbreaks (Chapters 2 and 5), and local GP consultation data collected to support an evaluation of the rotavirus vaccine (Chapter 4). The school attendance data which informed Chapter 4 was also a secondary dataset, as were the data from the IID2 Study which underlay the analyses in Chapters 6 and 7. The use of secondary data in this thesis allowed a significantly larger amount of data to be used than would have been possible from prospective data collection within the time available. This proved invaluable when analysing and comparing trends over time, especially when data were broken down by age group, area, or organism. Even using large national surveillance datasets, the case numbers for some subgroups were small and data had to be aggregated over multiple years to allow the analyses to be undertaken. It is therefore unlikely that prospective data collection would have provided enough data to inform these studies.

However, the use of secondary datasets also creates additional challenges as the format or detail within each dataset may not be optimal for answering a given research question. For example, the specificity of the data cannot be altered, and this was a limitation in several of the chapters in this thesis. The analysis of school illness absence data would have been enhanced if syndrome-specific data could have been collected from schools. Even if less data had been collected as a result, the additional information regarding symptoms would have removed the need for statistical modelling to identify the likely cause of illness absence. A descriptive analysis alone could then have revealed trends in IID as well as other causes of illness absence. Similarly, the analysis of outbreak and syndromic surveillance

data in Chapter 5, used to identify whether children provided a lead time ahead of adults, was limited by the broad IID definition, which did not allow norovirus cases and outbreaks to be distinguished from other causes of IID. A prospective cohort study could have addressed this issue by conducting sampling for both individual cases and outbreaks to confirm the causative organism.

An additional limitation of using surveillance datasets is the reporting bias inherent within these data. Current surveillance data are dependent on contact with healthcare services and certain groups are more likely to both need and seek healthcare advice when symptomatic, such as the very young, people with comorbidities and the elderly.[35,37,222,223]. Laboratory testing policies may also influence the detection of certain pathogens in different age groups.[145] Consequently, certain groups will be under-represented in surveillance data and school-aged children are likely to be one of these groups. Reporting bias is also an important limitation of the school outbreak data utilised in Chapters 2 and 5. Schools do not have a mandatory duty to report outbreaks and it should not be assumed that the underreporting of outbreaks occurs randomly. Prospective data collection to inform these analyses could have removed some of the reporting biases by actively following up individuals and schools to identify cases and outbreaks of illness.

Chapters 6 and 7 used data from the Second Study of Infectious Intestinal Disease in the Community (IID2 Study), which was a large, prospective cohort study. Whilst the data were not collected purposively to meet the research objectives of this thesis, the design of the IID2 study removed many of the reporting biases inherent within surveillance data. It also addressed the issues around sampling, as cases were actively followed up and sampled as part of the original research study. As a national cohort study, it was able to attain a sample size which would not have been achievable within the resource limitations of this thesis. Despite this, the number of individual pathogens were too few to allow organism-specific analyses to be undertaken in Chapter 6. In Chapter 7, when data were broken down by age and organism, the group sizes were small resulting in large confidence intervals and less certainty around the estimates. This illustrates the challenging trade-off between the large sample size offered by routine datasets and the enhanced data specificity obtained from study-specific data collection. In the absence of being able to achieve both these objectives within a single dataset, incorporating both large routine datasets and smaller study datasets into an analysis, and triangulating the results, could address some of these issues. This was what I had intended to do in Chapter 4 by working with a smaller number of local schools to obtain syndrome-specific absence data. Unfortunately, I was not able to obtain any data directly from schools, but the collection of these data is one of my recommendations for future research.

Generalisability of this work

The degree to which the findings of this thesis can be generalised to other settings varies between the different studies. Chapter 4 was focused on the Merseyside area and even across the three Local Government areas included in the study, there were different barriers to accessing data. The level of school data collected and how it was used by Local Government also varied. Consequently, issues in acquiring school attendance data are unlikely to be the same across all Local Government areas in England. However, the level of data collected and recorded by schools is expected to be similar and the limitations of using illness absence data are applicable across the country.

Chapter 2 used national-level data to explore risk factors for outbreak occurrence and attack rate in schools. These findings therefore have national relevance but are not necessarily applicable to school settings in other countries. Whilst factors such as the age range and size of schools were included in the analysis, data on other variables which could influence outbreak risk, such as class size and school policies, could not be captured. Whilst these factors are unlikely to vary significantly within the United Kingdom (UK), they may differ substantially in other countries, reducing the generalisability of these findings outside of the UK.

Of wider relevance are the findings of the systematic review (Chapter 3). The systematic review included studies from across the world but was limited to OECD countries which had established health surveillance systems already in place. Consequently, the findings of the review are not applicable to countries which may be using school attendance data in rural settings, in the absence of alternative forms of health surveillance. Whilst Chapter 5 used national-level data, the finding that norovirus cases in children started increasing earlier in the season than cases in adults adds to the wider knowledgebase surrounding the role of children in seasonal norovirus infection and has application beyond the UK. The symptom profiles developed in Chapters 6 and 7 are also relevant to the global community. However, organisms were grouped to generate these profiles and consequently they may be affected by significant changes in the prevalence of individual organisms within each group. This could limit their application to countries which are affected by different IID pathogens to those commonly found in the UK.

8.5 Recommendations for future research

Based on the findings of this thesis, the following are recommended to guide future work in this field:

Recommendation 1: Future research should explore the collection of syndrome-specific illness absence data from schools in the UK.

The specificity of school absence data was highlighted as an important feature of school-based surveillance systems in the systematic review, with syndrome-specific absences providing good correlation and more lead time ahead of health surveillance data. The inability to distinguish prevalent winter pathogens such as norovirus and influenza is a major limitation of existing illness absence data in the UK. However, the survey in Chapter 4 revealed that some schools already request information on the symptoms of illness and if such data collection could be standardised, data from different schools could be aggregated for research and surveillance purposes. Future research should consider utilising and enhancing the electronic attendance reporting systems currently in place to allow syndrome-specific data to be captured across a larger number of schools and geographical areas whilst minimising the additional workload required by schools.

Recommendation 2: Insight work is needed with parents and schools to better understand the influencing factors behind illness absence reporting in schools.

Both the school survey and members of the Public and Patient Involvement Panel highlighted concerns regarding the accuracy of illness reporting from parents. Children who are absent from school for unauthorised reasons may be falsely reported to be unwell to avoid further repercussions. Furthermore, schools may have their own reasons for not wanting to challenge the cause of absence as too many unauthorised absences reflect poorly on a school's performance. A better understanding of parental and school-level factors which influence illness absence reporting would not only improve the interpretation of this dataset but also help identify and address barriers to engagement with research in this area.

Recommendation 3: The Unique Reference Number (URN) should be documented for all schools which report outbreaks to Public Health England.

Appropriate identifiers need to be used for school-based data to support linkage across datasets. For schools, the name and postcode are inadequate to correctly identify the school. Furthermore, schools may be closed, and new schools opened on the same site. The URN of a school is unique and not only allows a school to be identified across different datasets, but also ensures the correct status of a school is captured. The URN should be incorporated into the collection of all school-based surveillance data, including illness absence data. Ensuring a URN is documented for each school could enhance the possibilities of data linkage and increase the utility of both school outbreak data and school attendance data for surveillance and research purposes.

Recommendation 4: The potential benefit of monitoring cases and outbreaks of IID amongst children to provide an early warning of seasonal norovirus infection warrants further exploration.

The findings of this thesis suggest that cases and outbreaks of IID amongst children could provide a 1-3 week lead time ahead of cases and outbreaks in adult age groups. This early warning of seasonal IID could be used to put preventative measures in place before norovirus enters care home and hospital settings. However, both school outbreaks and syndromic surveillance data are not norovirus specific and will also capture other causes of IID. The use of school outbreak data as an early warning indicator may be improved by enhancing sampling in community outbreaks to confirm the causative organism. Further investigation is needed to explore whether the lead time offered by these data is sufficient to put protective interventions in place in other settings and whether implementing these interventions earlier is effective at reducing the rate of norovirus infection in adult age groups.

Recommendation 5: The practical application of symptom profiles to syndromic surveillance systems needs to be examined and evaluated.

Epidemiological criteria have already been developed which utilise symptoms to make inferences as to the underlying organism in outbreak scenarios. This research has identified symptoms which could be used to distinguish bacterial from viral IID for individual cases. However, the application and integration of these profiles into syndromic surveillance systems requires further consideration. Detailed symptom profiles which incorporate multiple symptoms combined with season of illness may not be practical within the time constraints of many services which provide surveillance data. This is further complicated by the change in symptom profiles across age groups. Whether single symptoms, such as diarrhoea in the absence of vomiting, can provide sufficient discrimination between bacterial and viral IID should be considered and further study should explore the feasibility of applying these profiles in practice.

8.6 Conclusion

This thesis aimed to improve knowledge and understanding of the burden of infectious intestinal disease in schools and school-aged children, and the potential role and utility of school attendance registers in the surveillance of IID. Through this research I have sought to describe the burden of IID on illness absence and outbreaks in school settings, to consider the role and utility of school attendance registers in the surveillance of IID, to explore whether children provide an early warning of IID infections in the community, and to investigate whether symptoms alone can distinguish between different IID organisms to enhance syndromic surveillance systems.

The findings of this research not only highlight the importance of IID as a cause of outbreaks in schools but also the potential for school outbreaks to be used as an early warning of seasonal norovirus infection before outbreaks start increasing in care homes and hospitals. Whilst there is a potential role for school attendance registers in the surveillance of IID, there are currently significant challenges in accessing these data and distinguishing IID from other causes of illness absence. Improving access to these data and exploring the collection of syndrome-specific school absence data should be the focus of future work in this area. The use of symptom profiles to distinguish bacterial from viral IID could enhance not only data collection from schools but also existing syndromic surveillance systems. The feasibility of applying these profiles in practice needs to be explored, but such profiles have the potential to improve the utility of these data for both surveillance and research purposes, enhancing our understanding of the epidemiology of IID infections in the UK.

References

1. **Department for Education**. Schools, pupils, and their characteristics. National Statistics. 2020. (www.gov.uk/government/statistics/schools-pupils-and-their-characteristics-january-2020). Accessed 29 September 2020.
2. **European Education and Culture Executive Agency.** The Structure of the European Education Systems 2018/19. 2019. (op.europa.eu/en/publication-detail/-/publication/9f68b65e-cc36-11e8-9424-01aa75ed71a1/language-en/format-PDF). Accessed 14 January 2021.
3. **Gov.UK**. School admissions: School starting age. 2020. (www.gov.uk/schools-admissions/school-starting-age). Accessed 14 January 2021.
4. **Gov.UK**. School leaving age. (www.gov.uk/know-when-you-can-leave-school). Accessed 14 January 2021.
5. **Department for Education**. Types of school. (www.gov.uk/types-of-school). Accessed 14 January 2021.
6. **Department for Education**. National curriculum. 2014. (www.gov.uk/government/collections/national-curriculum). Accessed 14 January 2021.
7. **Independent Schools Inspectorate.** (www.isi.net/). Accessed 14 January 2021.
8. **Department for Education**. Health and safety: responsibilities and duties for schools. 2018. (www.gov.uk/government/publications/health-and-safety-advice-for-schools/responsibilities-and-duties-for-schools#other-areas-and-activities-to-consider). Accessed 20 January 2021.
9. **Public Health England**. National Child Measurement Programme operational guidance: 2020. 2020. (www.gov.uk/phe%5Cnhttps://www.gov.uk/government/uploads/system/uploads/attachment_data/file/377902/NCMP_operational_guidance.pdf). Accessed 20 January 2021.
10. **Public Health England**. Immunising primary school children against flu. 2015. (assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/434137/PHE_9342_Immunising_primary_schoolchildren_against_flu_A4_06_web.pdf). Accessed 20 January 2021.
11. **Tiley K, *et al.*** School-based vaccination programmes: An evaluation of school immunisation delivery models in England in 2015/16. *Vaccine* 2020; **38**: 3149–3156.
12. **Fonseca S, Forsyth H, Neary W**. School hearing screening programme in the UK: Practice and performance. *Archives of Disease in Childhood* 2005; **90**: 154–156.
13. **Public Health England**. Child vision screening. 2019. (https://www.gov.uk/government/publications/child-vision-screening). Accessed 20 January

2021.

14. **Jackson MI**. Understanding links between adolescent health and educational attainment. *Demography* 2009; **46**: 671–694.
15. **Haas SA, Fosse NE**. Health and the educational attainment of adolescents: Evidence from the NLSY97. *Journal of Health and Social Behavior* 2008; **49**: 178–192.
16. **Haas SA**. Health selection and the process of social stratification: The effect of childhood health on socioeconomic attainment. *Journal of Health and Social Behavior* 2006; **47**: 339–354.
17. **von dem Knesebeck O, Verde PE, Dragano N**. Education and health in 22 European countries. *Social Science and Medicine* 2006; **63**: 1344–1351.
18. **Mackenbach JP, *et al.*** Variations in the relation between education and cause-specific mortality in 19 European populations: A test of the 'fundamental causes' theory of social inequalities in health. *Social Science and Medicine* 2015; **127**: 51–62.
19. **Eide ER, Showalter MH**. Estimating the relation between health and education: What do we know and what do we need to know? *Economics of Education Review* 2011; **30**: 778–791.
20. **Schuring M, *et al.*** The effects of ill health on entering and maintaining paid employment: Evidence in European countries. *Journal of Epidemiology and Community Health* 2007; **61**: 597–604.
21. **Kroll LE, Lampert T**. Unemployment, social support and health problems: results of the GEDA study in Germany, 2009. *Deutsches Arzteblatt* 2011; **108**: 47–52.
22. **Tam CC, *et al.*** The second study of infectious intestinal disease in the community (IID2 study). UK Food Standards Agency, 2012; project no B18021.
23. **Uhnoo I, Olding-Stenkvist E, Kreuger A**. Clinical features of acute gastroenteritis associated with rotavirus, enteric adenoviruses, and bacteria. *Archives of Disease in Childhood* 1986; **61**: 732–738.
24. **Graham DY, *et al.*** Epidemiology of Helicobacter pylori in an asymptomatic population in the United States. Effect of age, race, and socioeconomic status. *Gastroenterology* 1991; **100**: 1495–1501.
25. **Tam CC, *et al.*** Changes in causes of acute gastroenteritis in the United Kingdom over 15 years: microbiologic findings from 2 prospective, population-based studies of infectious intestinal disease. *Clinical Infectious Diseases* 2012; **54**: 1275–1286.
26. **Hall AJ, *et al.*** Incidence of acute gastroenteritis and role of norovirus, Georgia, USA, 2004-2005. *Emerging Infectious Diseases* 2011; **17**: 1381–1388.
27. **De Wit MA, *et al.*** Sensor, a population-based cohort study on gastroenteritis in the Netherlands: incidence and etiology. *American Journal of Epidemiology* 2001; **154**: 666–674.

28. **Marshall JA, *et al.*** Incidence and characteristics of endemic Norwalk-like virus-associated gastroenteritis. *Journal of Medical Virology* 2003; **69**: 568–578.
29. **Tam CC, *et al.*** Longitudinal study of infectious intestinal disease in the UK (IID2 study): incidence in the community and presenting to general practice. *Gut* 2012; **61**: 69–77.
30. **Ahmed SM, Lopman BA, Levy K**. A systematic review and meta-analysis of the global seasonality of norovirus. *PLoS ONE* 2013; **8**: e75922.
31. **Bernstein DI**. Rotavirus overview. *The Pediatric Infectious Disease Journal* 2009; **28**: S50–S53.
32. **Parashar UD, *et al.*** Global illness and deaths caused by rotavirus disease in children. *Emerging Infectious Diseases* 2003; **9**: 565–572.
33. **Harris JP, *et al.*** Evaluating rotavirus vaccination in England and Wales. Part I. Estimating the burden of disease. *Vaccine* 2007; **25**: 3962–3970.
34. **Crawford SE, *et al.*** Rotavirus infection. *Nature Reviews Disease Primers* 2017; **3**: 1–16.
35. **Acheson D, Allos BM**. Campylobacter jejuni infections: update on emerging issues and trends. *Clinical Infectious Diseases* 2001; **32**: 1201–1206.
36. **Nichols GL, *et al.*** Campylobacter epidemiology: a descriptive study reviewing 1 million cases in England and Wales between 1989 and 2011. *BMJ Open* 2012; **2**: e001179. doi:10.1136/bmjopen-2012-001179
37. **Glass RI, Parashar UD, Estes MK**. Norovirus gastroenteritis. *New England Journal of Medicine* 2009; **361**: 1776–1785.
38. **Rockx B, *et al.*** Natural history of human Calicivirus infection: A prospective cohort study. *Clinical Infectious Diseases* 2002; **35**: 246–253.
39. **Lopman BA, *et al.*** Clinical manifestation of norovirus gastroenteritis in health care settings. *Clinical Infectious Diseases* 2004; **39**: 318–324.
40. **Mattner F, *et al.*** Risk groups for clinical complications of norovirus infections: An outbreak investigation. *Clinical Microbiology and Infection* 2006; **12**: 69–74.
41. **Hall AJ**. Noroviruses: the perfect human pathogens? *Journal of Infectious Diseases* 2012; **205**: 1622–1624.
42. **Arias C, *et al.*** Epidemiological and clinical features of norovirus gastroenteritis in outbreaks: A population-based study. *Clinical Microbiology and Infection* 2010; **16**: 39–44.
43. **Patel M, *et al.*** Noroviruses: A comprehensive review. *Journal of Clinical Virology* 2009; **44**: 1–8.
44. **Bartsch SM, *et al.*** Global economic burden of norovirus gastroenteritis. *PLoS ONE* 2016; **11**: e0151219.
45. **Tam CC, O'Brien SJ**. Economic cost of campylobacter, norovirus and rotavirus disease in the United Kingdom. *PLoS ONE* 2016; **11**: e0138526.

46. **Tate JE, *et al.*** 2008 estimate of worldwide rotavirus-associated mortality in children younger than 5 years before the introduction of universal rotavirus vaccination programmes: a systematic review and meta-analysis. *The Lancet Infectious Diseases* 2012; **12**: 136–141.
47. **Burnett E, *et al.*** Global impact of rotavirus vaccination on childhood hospitalizations and mortality from diarrhea. *The Journal of Infectious Diseases* 2017; **215**: 1666–1672.
48. **Lopman BA, *et al.*** Infant rotavirus vaccination may provide indirect protection to older children and adults in the United States. *The Journal of Infectious Diseases* 2011; **204**: 980–986.
49. **Markkula J, *et al.*** Rotavirus epidemiology 5–6 years after universal rotavirus vaccination: persistent rotavirus activity in older children and elderly. *Infectious Diseases* 2017; **49**: 388–395.
50. **Kaakoush NO, *et al.*** Global epidemiology of campylobacter infection. *Clinical Microbiology Reviews* 2015; **28**: 687-720. doi: 10.1128/CMR.00006-15.
51. **Tam CC, *et al.*** Incidence of Guillain-Barré syndrome among patients with *campylobacter* infection: a General Practice research database study. *The Journal of Infectious Diseases* 2006; **194**: 95–97.
52. **Liu L, *et al.*** Global, regional, and national causes of child mortality: An updated systematic analysis for 2010 with time trends since 2000. *The Lancet* 2012; **379**: 2151–2161.
53. **Elliott EJ**. Acute gastroenteritis in children. *BMJ* 2007; **334**: 35-40.
54. **Murphy MS**. Guidelines for managing acute gastroenteritis based on a systematic review of published research. *Archives of Disease in Childhood* 1998; **79**: 279–284.
55. **Barker SF, *et al.*** Cost of gastroenteritis in Australia: A healthcare perspective. *PLoS ONE* 2018; **13**: e0195759.
56. **Lee MB, Greig JD**. A review of gastrointestinal outbreaks in schools: effective infection control interventions. *Journal of School Health* 2010; **80**: 588–598.
57. **Bartsch SM, O'Shea KJ, Lee BY**. The clinical and economic burden of norovirus gastroenteritis in the United States. *The Journal of Infectious Diseases* 2020; **222**: 1910–1919.
58. **Allen CW, Diamond-Myrsten S, Rollins LK**. School absenteeism in children and adolescents. *American Family Physician* 2018; **98**: 738–744.
59. **Aucejo EM, Romano TF**. Assessing the effect of school days and absences on test score performance. *Economics of Education Review* 2016; **55**: 70–87.
60. **Marks PJ, *et al.*** A school outbreak of Norwalk-like virus: evidence for airborne transmission. *Epidemiology and Infection* 2003; **131**: 727–736.
61. **Lopman B, *et al.*** Environmental transmission of norovirus gastroenteritis. *Current Opinion in Virology* 2012; **2**: 96-102.

62. **Hall AJ, *et al.*** Acute gastroenteritis surveillance through the national outbreak reporting system, United States. *Emerging Infectious Diseases* 2013; **19**: 1305.
63. **Sattar SA, *et al.*** Institutional outbreaks of rotavirus diarrhoea: Potential role of fomites and environmental surfaces as vehicles for virus transmission. *Journal of Hygiene* 1986; **96**: 277–289.
64. **Hall AJ, *et al.*** Epidemiology of foodborne norovirus outbreaks, United States, 2001-2008. *Emerging Infectious Diseases* 2012; **18**: 1566–1573.
65. **Barker J, Stevens D, Bloomfield SF**. Spread and prevention of some common viral infections in community facilities and domestic homes. *Journal of Applied Microbiology* 2001; **91**: 7-21.
66. **Hardstaff JL, *et al.*** Foodborne and food-handler norovirus outbreaks: a systematic review. *Foodborne Pathogens and Disease*. 2018; **15**: 589–597.
67. **Maalouf H, *et al.*** Strain-dependent norovirus bioaccumulation in oysters. *Applied and Environmental Microbiology* 2011; **77**: 3189–3196.
68. **Fhogartaigh CN, Edgeworth JD**. Bacterial gastroenteritis. *Medicine* 2009; **37**: 586–593.
69. **Taylor EV, *et al.*** Common source outbreaks of Campylobacter infection in the USA, 1997-2008. *Epidemiology and Infection* 2013; **141**: 987–996.
70. **Unicomb LE, *et al.*** Outbreaks of campylobacteriosis in Australia, 2001 to 2006. *Foodborne Pathogens and Disease* 2009; **6**: 1241–1250.
71. **Kothary MH, Babu US**. Infective dose of foodborne pathogens in volunteers: A review. *Journal of Food Safety* 2001; **21**: 49–68.
72. **Beutels P, *et al.*** Social mixing patterns for transmission models of close contact infections: exploring self-evaluation and diary-based data collection through a web-based interface. *Epidemiology and Infection* 2006; **134**: 1158-1166.
73. **Edmunds WJ, *et al.*** Mixing patterns and the spread of close-contact infectious diseases. *Emerging Themes in Epidemiology* 2006; **3**: 10.
74. **Béraud G, *et al.*** The French Connection: the first large population-based contact survey in France relevant for the spread of infectious diseases. *PLoS ONE* 2015; **10**: e0133203.
75. **Mossong J, *et al.*** Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Medicine* 2008; **5**: e74.
76. **Read JM, Eames KT, Edmunds WJ**. Dynamic social networks and the implications for the spread of infectious disease. *Journal of The Royal Society Interface* 2008; **5**: 1001–1007.
77. **Stehlé J, *et al.*** High-resolution measurements of face-to-face contact patterns in a primary school. *PLoS ONE* 2011; **6**: e23176.
78. **Fournet J, Barrat A**. Contact patterns among high school students. *PLoS ONE* 2014; **9**: e107878.

79. **Conlan AJK, *et al.*** Measuring social networks in British primary schools through scientific engagement. *Proceedings of the Royal Society B: Biological Sciences* 2011; **278**: 1467–1475.
80. **Salathe M, *et al.*** A high-resolution human contact network for infectious disease transmission. *Proceedings of the National Academy of Sciences of the United States of America* 2010; **107**: 22020–22025.
81. **Glass LM, Glass RJ**. Social contact networks for the spread of pandemic influenza in children and teenagers. *BMC Public Health* 2008; **8**: 61.
82. **Mikolajczyk RT, *et al.*** Social contacts of school children and the transmission of respiratory-spread pathogens. *Epidemiology and Infection* 2008; **136**: 813–822.
83. **Petrie JG, *et al.*** Influenza transmission in a cohort of households with children: 2010-2011. *PLoS ONE* 2013; **8**: e75339.
84. **Longini IM, *et al.*** Estimating household and community transmission parameters for influenza. *American Journal of Epidemiology* 1982; **115**: 736–751.
85. **Viboud C, *et al.*** Risk factors of influenza transmission in households. *British Journal of General Practice* 2004; **54**: 684–689.
86. **Monto AS**. Studies of the community and family: acute respiratory illness and infection. *Epidemiologic reviews* 1994; **16**: 351–73.
87. **Cauchemez S, *et al.*** Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza. *Proceedings of the National Academy of Sciences of the United States of America* 2011; **108**: 2825–30.
88. **Calatayud L, *et al.*** Pandemic (H1N1) 2009 virus outbreak in a school in London, April-May 2009: An observational study. *Epidemiology and Infection* 2010; **138**: 183–191.
89. **France AM, *et al.*** Household transmission of 2009 influenza A (H1N1) virus after a school-based outbreak in New York City, April–May 2009. *The Journal of Infectious Diseases* 2010; **201**: 984–992.
90. **Vynnycky E, *et al.*** Estimating the impact of childhood influenza vaccination programmes in England and Wales. *Vaccine* 2008; **26**: 5321–5330.
91. **Heikkinen T, *et al.*** Should healthy children be vaccinated against influenza? *European Journal of Pediatrics* 2006; **165**: 223-228.
92. **Monto AS**. Interrupting the transmission of respiratory tract infections: theory and practice. *Clinical Infectious Diseases* 1999; **28**: 200–204.
93. **Pebody RG, *et al.*** Uptake and impact of vaccinating primary school-age children against influenza: experiences of a live attenuated influenza vaccine programme, England, 2015/16. *Eurosurveillance* 2018; **23**: 1700496.
94. **Gallagher KM, *et al.*** Measles, Mumps, and Rubella. *Vaccines: A Biography* 2010. Published

online: 10 November 2009. doi.org/10.1007/978-1-4419-1108-7_13

95. **Rajmil L**. Role of children in the transmission of the COVID-19 pandemic: A rapid scoping review. *BMJ Paediatrics Open* 2020; **4**: e000722.

96. **Ludvigsson JF**. Children are unlikely to be the main drivers of the COVID-19 pandemic – A systematic review. *Acta Paediatrica* 2020; **109**: 1525–1530.

97. **de Graaf M, Villabruna N, Koopmans MP**. Capturing norovirus transmission. *Current Opinion in Virology* 2017; **22**: 64–70.

98. **Chan MCW, *et al.*** Virus genotype distribution and virus burden in children and adults hospitalized for norovirus gastroenteritis, 2012–2014, Hong Kong. *Scientific Reports* 2015. Published online: 17 June 2015. doi:10.1038/SREP11507.

99. **Kambhampati A, Koopmans M, Lopman BA**. Burden of norovirus in healthcare facilities and strategies for outbreak control. *Journal of Hospital Infection* 2015; **89**: 296–301.

100. **Bernard H, *et al.*** Epidemiology of norovirus gastroenteritis in Germany 2001-2009: eight seasons of routine surveillance. *Epidemiology and Infection* 2014; **142**: 63–74.

101. **Joseph C, *et al.*** A review of outbreaks of infectious disease in schools in England and Wales 1979–88. *Epidemiology and Infection* 1990; **105**: 419–434.

102. **Zhao H, Joseph CA, Phin N**. Outbreaks of influenza and influenza-like illness in schools in England and Wales, 2005/06. *Eurosurveillance* 2007; **12**: 3–4.

103. **Czumbel I, *et al.*** Management and control of communicable diseases in schools and other child care settings: systematic review on the incubation period and period of infectiousness. *BMC Infectious Diseases* 2018. Published online: 02 May 2018. doi:10.1186/s12879-018-3095-8.

104. **Thorrington D, *et al.*** Social and economic impacts of school influenza outbreaks in England: survey of caregivers. *Journal of School Health* 2017; **87**: 209–216.

105. **Matthews JE, *et al.*** The epidemiology of published norovirus outbreaks: A review of risk factors associated with attack rate and genogroup. *Epidemiology and Infection* 2012; **140**: 1161–1172.

106. **Daniels NA, *et al.*** Foodborne disease outbreaks in United States schools. *Pediatric Infectious Disease Journal* 2002; **21**: 623–628.

107. **Evans HS, Maguire H**. Outbreaks of infectious intestinal disease in schools and nurseries in England and Wales 1992 to 1994. *Communicable Disease Report CDR Review* 1996; **6**: R103-8.

108. **Finnie TJR, *et al.*** An analysis of influenza outbreaks in institutions and enclosed societies. *Epidemiology and Infection* 2014; **142**: 107–113.

109. **Cleary V, Slaughter R, Heathcock R**. An infection control programme in primary schools and the wider public health impact. *British Journal of Infection Control* 2003; **4**: 11–17.

110. **Schmidt WP, *et al.*** Formative research on the feasibility of hygiene interventions for influenza control in UK primary schools. *BMC Public Health* 2009; **9**: 1–8.

111. **Kaltenthaler EC, *et al.*** Faecal contamination on children's hands and environmental surfaces in primary schools in Leeds. *Epidemiology and Infection* 1995; **115**: 527–534.

112. **Willmott M, *et al.*** Effectiveness of hand hygiene interventions in reducing illness absence among children in educational settings: a systematic review and meta-analysis. *Archives of Disease in Childhood* 2016; **101**: 42–50.

113. **Meadows E, Saux N Le**. A systematic review of the effectiveness of antimicrobial rinse-free hand sanitizers for prevention of illness-related absenteeism in elementary school children. *BMC Public Health* 2004; **4**: 50.

114. **Prazuck T, *et al.*** Reducing gastroenteritis occurrences and their consequences in elementary schools with alcohol-based hand sanitizers. *Pediatric Infectious Disease Journal* 2010; **29**: 994–998.

115. **Nandrup-Bus I**. Mandatory handwashing in elementary schools reduces absenteeism due to infectious illness among pupils: A pilot intervention study. *American Journal of Infection Control* 2009; **37**: 820–826.

116. **Sandora TJ, Shih MC, Goldmann DA**. Reducing absenteeism from gastrointestinal and respiratory illness in elementary school students: A randomized, controlled trial of an infection-control intervention. *Pediatrics* 2008; **121**: e1555–e1562.

117. **Uhari M, Möttönen M**. An open randomized controlled trial of infection prevention in child day-care centers. *Pediatric Infectious Disease Journal* 1999; **18**: 672–677.

118. **Krilov LR, *et al.*** Impact of an infection control program in a specialized preschool. *American Journal of Infection Control* 1996; **24**: 167–173.

119. **Roberts L, *et al.*** Effect of infection control measures on the frequency of diarrheal episodes in child care: A randomized, controlled trial. *Pediatrics* 2000; **105**: 743–746.

120. **Public Health England**. Health protection in schools and other childcare facilities. 2017. (www.gov.uk/government/publications/health-protection-in-schools-and-other-childcare-facilities). Accessed 25 September 2020.

121. **Public Health England**. Preventing person-to-person spread following gastrointestinal infections: guidelines for public health physicians and environmental health officers. 2004. (assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/336841/guidance_on_preventing_person-to-person_infections_4_04.pdf). Accessed 6 January 2021.

122. **Richardson M, *et al.*** Evidence base of incubation periods, periods of infectiousness and exclusion policies for the control of communicable diseases in schools and preschools.

Pediatric Infectious Disease Journal 2001; **20**: 380–391.

123. **Gemmetto V, Barrat A, Cattuto C**. Mitigation of infectious disease at school: Targeted class closure vs school closure. *BMC Infectious Diseases* 2014; **14**: 695.
124. **Sugisaki K, *et al.*** Effective school actions for mitigating seasonal influenza outbreaks in Niigata, Japan. *PLoS ONE* 2013. Published online: September 2013. doi:10.1371/journal.pone.0074716.
125. **Jackson C, *et al.*** School closures and influenza: systematic review of epidemiological studies. *BMJ Open* 2013; **3**: 2149.
126. **Rodriguez C V, *et al.*** Association between school closure and subsequent absenteeism during a seasonal influenza epidemic. *Epidemiology* 2009; **20**: 787–792.
127. **Egger JR, *et al.*** The effect of school dismissal on rates of influenza-like illness in New York City schools during the spring 2009 novel H1N1 outbreak. *Journal of School Health* 2012; **82**: 123–130.
128. **Kawai S, *et al.*** Influenza vaccination of schoolchildren and influenza outbreaks in a school. *Clinical Infectious Diseases* 2011; **53**: 130–136.
129. **Pan Y, *et al.*** Influenza vaccination in preventing outbreaks in schools: A long-term ecological overview. *Vaccine* 2017; **35**: 7133–7138.
130. **King Jr. JC, *et al.*** Direct and indirect impact of influenza vaccination of young children on school absenteeism. *Vaccine* 2012; **30**: 289–293.
131. **Davis MM, *et al.*** Countywide school-based influenza immunization: Direct and indirect impact on student absenteeism. *Pediatrics* 2008; **122**: E260–E265.
132. **Hallowell BD, Parashar UD, Hall AJ**. Epidemiologic challenges in norovirus vaccine development. *Human Vaccines & Immunotherapeutics* 2019; **15**: 1279–1283.
133. **Steele MK, *et al.*** Targeting pediatric versus elderly populations for norovirus vaccines: a model-based analysis of mass vaccination options. *Epidemics* 2016; **17**: 42–49.
134. **Department for Education**. School attendance: Guidance for maintained schools, academies, independent schools and local authorities. 2020. (assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/907535/School_attendance_guidance_for_2020_to_2021_academic_year.pdf). Accessed 6 January 2021.
135. **Mellou K, *et al.*** Epidemiological investigation of two parallel gastroenteritis outbreaks in school settings. *BMC Public Health* 2013. Published online: March 2013. doi:10.1186/1471-2458-13-241.
136. **Ha C, Rios LM, Pannaraj PS**. Knowledge, attitudes, and practices of school personnel regarding influenza, vaccinations, and school outbreaks. *Journal of School Health* 2013; **83**:

554–561.

137. **Mann P, *et al.*** Alert system to detect possible school-based outbreaks of influenza-like illness. *Emerging Infectious Diseases* 2011; **17**: 262–264.
138. **Tanabe Y, *et al.*** Infection control in nursery schools and schools using a school absenteeism surveillance system. *The Tohoku Journal of Experimental Medicine* 2019; **247**: 173–178.
139. **Public Health England**. Notifications of infectious diseases (NOIDs). (https://www.gov.uk/government/collections/notifications-of-infectious-diseases-noids). Accessed 4 March 2021.
140. **Elliot AJ, *et al.*** Syndromic surveillance – a public health legacy of the London 2012 Olympic and Paralympic Games. *Public Health* 2013; **127**: 777–781.
141. **Meakins SM, *et al.*** General outbreaks of infectious intestinal disease (IID) in hospitals, England and Wales, 1992-2000. *Journal of Hospital Infection* 2003; **53**: 1–5.
142. **Harris JP, *et al.*** The development of web-based surveillance provides new insights into the burden of norovirus outbreaks in hospitals in England. *Epidemiology and Infection* 2014; **142**: 1590–1598.
143. **Kara-Zaïtri C, Gelletlie R, Schweiger M**. The development and deployment of a national web-based system for communicable disease control in England. *International Journal of Infectious Diseases* 2012; **16**: e133.
144. **UK Food Standards Agency**. A report of the study of infectious intestinal disease in England. 2000. (www.food.gov.uk/sites/default/files/media/document/iid1_study_final_report.pdf). Accessed 4 January 2021.
145. **Atchison CJ, *et al.*** Clinical laboratory practices for the detection of rotavirus in England and Wales: can surveillance based on routine laboratory testing data be used to evaluate the impact of vaccination? *Eurosurveillance* 2009; **14**: 19217.
146. **Humphries RM, Linscott AJ**. Laboratory diagnosis of bacterial gastroenteritis. *Clinical Microbiology Reviews* 2015; **28**: 3–31.
147. **Harcourt SE, *et al.*** Use of a large general practice syndromic surveillance system to monitor the progress of the influenza A(H1N1) pandemic 2009 in the UK. *Epidemiology and Infection* 2012; **140**: 100–105.
148. **Elliot AJ, *et al.*** Establishing an emergency department syndromic surveillance system to support the London 2012 Olympic and Paralympic Games. *Emergency Medicine Journal* 2012; **29**: 954–60.
149. **Todkill D, *et al.*** Utility of ambulance data for real-time syndromic surveillance: a pilot in the West Midlands region, United Kingdom. *Prehospital and Disaster Medicine* 2017; **32**: 667–672.
150. **Baker M, *et al.*** Early warning and NHS Direct: a role in community surveillance? *Journal of*

Public Health 2003; **25**: 362–368.

151. **World Health Organization**. Diarrhoeal disease. (https://www.who.int/news-room/fact-sheets/detail/diarrhoeal-disease). Accessed 3 March 2021.
152. **Majowicz SE, *et al.*** A common, symptom-based case definition for gastroenteritis. *Epidemiology and Infection* 2008; **136**: 886–894.
153. **Kaplan JE, *et al.*** Epidemiology of Norwalk gastroenteritis and the role of Norwalk virus in outbreaks of acute nonbacterial gastroenteritis. *Annals of Internal Medicine* 1982; **96**: 756–761.
154. **Hedberg CW, Osterholm MT**. Outbreaks of food-borne and waterborne viral gastroenteritis. *Clinical Microbiology Reviews* 1993; **6**: 199–210.
155. **Dalton CB, *et al.*** Outbreaks of enterotoxigenic Escherichia coli infection in American adults: A clinical and epidemiologic profile. *Epidemiology and Infection* 1999; **123**: 9–16.
156. **Turcios RM, *et al.*** Reevaluation of epidemiological criteria for identifying outbreaks of acute gastroenteritis due to norovirus: United States, 1998-2000. *Clinical Infectious Diseases* 2006; **42**: 964–969.
157. **Ginsberg J, *et al.*** Detecting influenza epidemics using search engine query data. *Nature* 2009; **457**: 1012–1014.
158. **Milinovich GJ, *et al.*** Internet-based surveillance systems for monitoring emerging infectious diseases. *The Lancet Infectious Diseases* 2014; **14**: 160–168.
159. **Polgreen PM, *et al.*** Using internet searches for influenza surveillance. *Clinical Infectious Diseases* 2008; **47**: 1443–1448.
160. **Elliot AJ, *et al.*** Internet-based remote health self-checker symptom data as an adjuvant to a national syndromic surveillance system. *Epidemiology and Infection* 2015; **143**: 3416–3422.
161. **Hulth A, Rydevik G, Linde A**. Web queries as a source for syndromic surveillance. *PLoS ONE* 2009; **4**: e4378.
162. **Heffernan R, *et al.*** New York City syndromic surveillance systems. *Morbidity and Mortality Weekly Report* 2004; **53 Suppl**: 23–27.
163. **Edge VL, *et al.*** Syndromic surveillance of gastrointestinal illness using pharmacy over-the-counter sales. *Canadian Journal of Public Health* 2004; **95**: 446–450.
164. **Bollaerts K, *et al.*** Timeliness of syndromic influenza surveillance through work and school absenteeism. *Archives of Public Health* 2010; **68**: 115–120.
165. **Williams NJ, *et al.*** Comparison of 3 school-based influenza surveillance indicators: lessons learned from 2009 pandemic influenza A (H1N1)-Denver metropolitan region, Colorado. *Journal of Public Health Management and Practice* 2013; **19**: 119–125.
166. **Mogto CAK, *et al.*** School absenteeism as an adjunct surveillance indicator: experience during

the second wave of the 2009 H1N1 pandemic in Quebec, Canada. *PLoS ONE* 2012. Published online: March 2012. doi:10.1371/journal.pone.0034084.

167. **Lenaway DD, Ambler A**. Evaluation of a school-based influenza surveillance system. *Public Health Reports* 1995; **110**: 333–7.
168. **Schmidt WP, Pebody R, Mangtani P**. School absence data for influenza surveillance: a pilot study in the United Kingdom. *Eurosurveillance* 2010; **15**: 19467.
169. **Adams NL, *et al.*** Socioeconomic status and infectious intestinal disease in the community: A longitudinal study (IID2 study). *European Journal of Public Health* 2018; **28**: 134–138.
170. **Rose TC, *et al.*** Socioeconomic status is associated with symptom severity and sickness absence in people with infectious intestinal disease in the UK. *BMC Infectious Diseases* 2017. Published online: 23 June 2017. doi:10.1186/s12879-017-2551-1.
171. **Hungerford D, *et al.*** Influenza-associated hospitalisation, vaccine uptake and socioeconomic deprivation in an English city region: An ecological study. *BMJ Open* 2018. Published online: 1 December 2018. doi:10.1136/bmjopen-2018-023275.
172. **British Educational Suppliers Association (BESA): Key UK education statistics.** (www.besa.org.uk/key-uk-education-statistics/). Accessed 7 August 2020.
173. **Department for Education.** (www.gov.uk/government/organisations/department-for-education). Accessed 20 August 2019.
174. **Ofsted.** (www.gov.uk/government/organisations/ofsted/about). Accessed 7 August 2020.
175. **National Statistics: English indices of deprivation 2019.** (www.gov.uk/government/statistics/english-indices-of-deprivation-2019). Accessed 7 August 2020.
176. **The R Project for Statistical Computing.** (www.r-project.org/). Accessed 7 August 2020.
177. **Wang Y, *et al.*** Age, primary symptoms, and genotype characteristics of norovirus outbreaks in Shanghai schools in 2017. *Scientific Reports* 2018. Published online: 1 December 2018. doi:10.1038/s41598-018-33724-0.
178. **Glass K, McCaw JM, McVernon J**. Incorporating population dynamics into household models of infectious disease transmission. *Epidemics* 2011; **3**: 152–158.
179. **House T, Keeling MJ**. Household structure and infectious disease transmission. *Epidemiology and Infection* 2009; **137**: 654–661.
180. **Vivancos R, *et al.*** Factors affecting prevention and control of viral gastroenteritis outbreaks in care homes. *Journal of Hospital Infection* 2012; **82**: 129–132.
181. **Gavigan P, McCullers JA**. Influenza: Annual seasonal severity. *Current Opinion in Pediatrics* 2019; **31**: 112–118.
182. **Pebody RG, *et al.*** Uptake and impact of vaccinating school age children against influenza

during a season with circulation of drifted influenza A and B strains, England, 2014/15. *Eurosurveillance* 2015; **20**: 30029.

183. **Vukovic V, *et al.*** Identifying people at risk for influenza with low vaccine uptake based on deprivation status: A systematic review. *European Journal of Public Health* 2020; **30**: 132–141.
184. **World Health Organization**. Public health surveillance. (www.who.int/topics/public_health_surveillance/en/). Accessed 1 October 2019.
185. **Jajosky RA, Groseclose SL**. Evaluation of reporting timeliness of public health surveillance systems for infectious diseases. *BMC Public Health* 2004; **4**: 29.
186. **Morse SS**. Public health surveillance and infectious disease detection. *Biosecurity and Bioterrorism: Biodefense Strategy, Practice, and Science* 2012; **10**: 6–16.
187. **Henning KJ**. What is syndromic surveillance? *MMWR supplements* 2004; **53**: 5–11.
188. **Elliot A**. Syndromic surveillance: the next phase of public health monitoring during the H1N1 influenza pandemic? *Eurosurveillance* 2009; **14**: 19391.
189. **Heffernan R, *et al.*** Syndromic surveillance in public health practice, New York City. *Emerging Infectious Diseases* 2004; **10**: 858–864.
190. **Hope K, *et al.*** Syndromic surveillance: is it a useful tool for local outbreak detection? *Journal of Epidemiology and Community Health* 2006; **60**: 374–5.
191. **Katz R, *et al.*** Redefining syndromic surveillance. *Journal of Epidemiology and Global Health* 2011; **1**: 21–31.
192. **Loveridge P, *et al.*** Vomiting calls to NHS Direct provide an early warning of norovirus outbreaks in hospitals. *Journal of Hospital Infection* 2010; **74**: 385–393.
193. **Bork KH, *et al.*** Surveillance of ambulance dispatch data as a tool for early warning. *Eurosurveillance* 2006; **11**: 13–14.
194. **Moore K**. Real-time syndrome surveillance in Ontario, Canada: the potential use of emergency departments and Telehealth. *European Journal of Emergency Medicine* 2004; **11**: 3–11.
195. **Blozik E, Grandchamp C, Von Overbeck J**. Influenza surveillance using data from a telemedicine centre. *International Journal of Public Health* 2012; **57**: 447–452.
196. **Azor-Martinez E, *et al.*** The impact of common infections on school absenteeism during an academic year. *American Journal of Infection Control* 2014; **42**: 632–637.
197. **White CG, *et al.*** Reduction of illness absenteeism in elementary schools using an alcohol-free instant hand sanitizer. *The Journal of School Nursing* 2001; **17**: 248–265.
198. **Principi N, *et al.*** Burden of influenza in healthy children and their households. *Archives of Disease in Childhood* 2004; **89**: 1002–1007.
199. **Neuzil KM, Hohlbein C, Zhu YW**. Illness among schoolchildren during influenza season - Effect

on school absenteeism, parental absenteeism from work, and secondary illness in families. *Archives of Pediatrics & Adolescent Medicine* 2002; **156**: 986–991.

200. **Cate TR**. Clinical manifestations and consequences of influenza. *The American Journal of Medicine* 1987; **82**: 15–19.
201. **Glezen WP, Couch RB**. Inter pandemic influenza in the Houston area 1974-1976. *New England Journal of Medicine* 1978; **298**: 587–592.
202. **Cox NJ, Subbarao K**. Global epidemiology of influenza: past and present. *Annual Review of Medicine* 2000; **51**: 407–421.
203. **Donaldson A, *et al.*** School-based surveillance of acute infectious disease in children: a systematic review. *PROSPERO* 2019; CRD4201911.
204. **OECD Countries 2019.** (http://worldpopulationreview.com/countries/oecd-countries/). Accessed 20 September 2019.
205. **Besculides M, *et al.*** Evaluation of school absenteeism data for early outbreak detection, New York City. *BMC Public Health* 2005; **5**: 105.
206. **Crawford GB, *et al.*** Influenza and school-based influenza-like illness surveillance: a pilot initiative in Maryland. *Public Health Reports* 2011; **126**: 591–596.
207. **Kara EO, *et al.*** Absenteeism in schools during the 2009 influenza A(H1N1) pandemic: a useful tool for early detection of influenza activity in the community? *Epidemiology and Infection* 2012; **140**: 1328–1336.
208. **Kightlinger L, Horan V**. School illness absenteeism during 2009 influenza A (H1N1) pandemic--South Dakota, 2009-2010. *South Dakota Medicine* 2013; **66**: 177,179-181.
209. **Ma T, *et al.*** Syndromic surveillance of influenza activity in Sweden: an evaluation of three tools. *Epidemiology and Infection* 2015; **143**: 2390–2398.
210. **Mook P, *et al.*** Pilot scheme for monitoring sickness absence in schools during the 2006/07 winter in England: can these data be used as a proxy for influenza activity? *Eurosurveillance* 2007; **12**: E11-12.
211. **Suzue T, *et al.*** The new school absentees reporting system for pandemic influenza A/H1N1 2009 infection in Japan. *PLoS ONE* 2012. Published online: February 2012. doi:10.1371/journal.pone.0030639.
212. **Aldridge RW, *et al.*** Are school absences correlated with influenza surveillance data in England? Results from Decipher My Data-a research project conducted through scientific engagement with schools. *PLoS ONE* 2016. Published online: March 2016. doi:10.1371/journal.pone.0146964.
213. **Chu A, *et al.*** Assessing the relative timeliness of Ontario's syndromic surveillance systems for early detection of the 2009 influenza H1N1 pandemic waves. *Canadian Journal of Public*

Health 2013; **104**: e340–e347.

214. **Enanoria WTA, *et al.*** The epidemiology and surveillance response to pandemic influenza A (H1N1) among local health departments in the San Francisco Bay area. *BMC Public Health* 2013. Published online: March 2013. doi:10.1186/1471-2458-13-276.
215. **Kirk MD, *et al.*** World Health Organization estimates of the global and regional disease burden of 22 foodborne bacterial, protozoal, and viral diseases, 2010: a data synthesis. *PLoS Medicine* 2015; **12**: e1001921.
216. **Roberts JA, *et al.*** The study of infectious intestinal disease in England: socio-economic impact. *Epidemiology and infection* 2003; **130**: 1–11.
217. **Harris JP, *et al.*** Norovirus strain types found within the second infectious intestinal diseases (IID2) study an analysis of norovirus circulating in the community. *BMC Infectious Diseases* 2019; **19**: 1–8.
218. **Shah MP, Hall AJ**. Norovirus Illnesses in Children and Adolescents. *Infectious Disease Clinics of North America* 2018; **32**: 103–118.
219. **Kirkwood CD, Streitberg R**. Calicivirus shedding in children after recovery from diarrhoeal disease. *Journal of Clinical Virology* 2008; **43**: 346–348.
220. **Simmons K, *et al.*** Duration of immunity to norovirus gastroenteritis. *Emerging Infectious Diseases* 2013; **19**: 1260.
221. **Wall PG, *et al.*** Food poisoning: notifications, laboratory reports, and outbreaks--where do the statistics come from and what do they mean? *Communicable Disease Report CDR review* 1996; **6**: R93-100.
222. **Velázquez FR**. Protective effects of natural rotavirus infection. *The Pediatric Infectious Disease Journal* 2009; **28**: S54–S56.
223. **Kirk MD, Veitch MG, Hall GV**. Gastroenteritis and food-borne disease in elderly people living in long-term care. *Clinical Infectious Diseases* 2010; **50**: 397–404.
224. **Iturriza-Gómara M, Cunliffe N**. Rotavirus vaccine: a welcome addition to the immunisation schedule in the UK. *BMJ (Clinical research ed.)* 2013; **346**: f2347.
225. **Office for National Statistics**. 2011 Census: Population estimates by five-year age bands, and household estimates, for Local Authorities in the United Kingdom. 2013. (www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/2011censuspopulationestimatesbyfiveyearagebandsandhouseholdestimatesforlocalauthoritiesintheunitedkingdom). Accessed 20 August 2019.
226. **Public Health England**. Laboratory reporting to Public Health England: A guide for diagnostic laboratories. 2020. (assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file

/926838/PHE_Laboratory_reporting_guidelines_October-2020-v3.pdf). Accessed 1 September 2021.

227. **National Health Service**. NHS 111. (www.nhs.uk/using-the-nhs/nhs-services/urgent-and-emergency-care/nhs-111/). Accessed 20 August 2019.
228. **Hungerford D, *et al.*** Ecological assessment of the direct and indirect effects of routine rotavirus vaccination in Merseyside, UK using data from multiple health systems: a study protocol. *BMJ Open* 2014; **4**: e006161.
229. **Hungerford D, *et al.*** Rotavirus vaccine impact and socioeconomic deprivation: an interrupted time-series analysis of gastrointestinal disease outcomes across primary and secondary care in the UK. *BMC Medicine* 2018; **16**: 10.
230. **Office for National Statistics**. Estimates of the population for the UK, England, Wales, Scotland and Northern Ireland - Mid-2012 edition. (https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland). Accessed 16 January 2023.
231. **Jackson C, Vynnycky E, Mangtani P**. The relationship between school holidays and transmission of influenza in England and Wales. *American Journal of Epidemiology* 2016; **184**: 644–651.
232. **Lindsay L, *et al.*** A decade of norovirus disease risk among older adults in upper-middle and high income countries: A systematic review. *BMC Infectious Diseases* 2015; **15**: 1–16.
233. **Murata T, *et al.*** Prolonged norovirus shedding in infants ≤6 months of age with gastroenteritis. *Pediatric Infectious Disease Journal* 2007; **26**: 46–49.
234. **Phillips G, *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; **171**: 1014–1022.
235. **Götz H, *et al.*** Clinical spectrum and transmission characteristics of infection with Norwalk-like virus: findings from a large community outbreak in Sweden. *Clinical Infectious Diseases* 2001; **33**: 622–628.
236. **De Wit MA, Koopmans MP, Duynhoven YT van**. Risk factors for norovirus, sapporo-like virus, and group A rotavirus gastroenteritis. *Emerging Infectious Diseases* 2003; **9**: 1563.
237. **Phillips G, *et al.*** Risk factors for symptomatic and asymptomatic norovirus infection in the community. *Epidemiology and Infection* 2011; **139**: 1676–1686.
238. **Zeileis A, *et al.*** Testing and dating of structural changes in practice. *Computational Statistics & Data Analysis* 2003; **44**: 109–123.
239. **Hughes SL, *et al.*** Monitoring telehealth vomiting calls as a potential public health early

warning system for seasonal norovirus activity in Ontario, Canada. *Epidemiology and Infection* 2019. Published online: 2019. doi:10.1017/S0950268818003357.

240. **Eames KTD, Tilston NL, Edmunds WJ**. The impact of school holidays on the social mixing patterns of school children. *Epidemics* 2011; **3**: 103–108.
241. **Eames KTD**. The influence of school holiday timing on epidemic impact. *Epidemiology and Infection* 2014; **142**: 1963–1971.
242. **Ali ST, Kadi AS, Ferguson NM**. Transmission dynamics of the 2009 influenza A (H1N1) pandemic in India: the impact of holiday-related school closure. *Epidemics* 2013; **5**: 157–163.
243. **Liu LJ, *et al.*** Diagnostic value of bacterial stool cultures and viral antigen tests based on clinical manifestations of acute gastroenteritis in pediatric patients. *European Journal of Clinical Microbiology and Infectious Diseases* 2005; **24**: 559–561.
244. **De Wit MA, *et al.*** Etiology of gastroenteritis in sentinel general practices in The Netherlands. *Clinical Infectious Diseases* 2001; **33**: 280–288.
245. **Wiegering V, *et al.*** Gastroenteritis in childhood: A retrospective study of 650 hospitalized pediatric patients. *International Journal of Infectious Diseases* 2011; **15**: e401–e407.
246. **Chen SY, *et al.*** Molecular epidemiology and clinical manifestations of viral gastroenteritis in hospitalized pediatric patients in Northern Taiwan. *Journal of Clinical Microbiology* 2007; **45**: 2054–2057.
247. **Guasticchi G, *et al.*** Syndromic surveillance: sensitivity and positive predictive value of the case definitions. *Epidemiology and Infection* 2009; **137**: 662–671.
248. **Scallan E, *et al.*** Factors associated with seeking medical care and submitting a stool sample in estimating the burden of foodborne illness. *Foodborne Pathogens and Disease* 2006; **3**: 432–438.
249. **Moyo SJ, *et al.*** Comprehensive analysis of prevalence, epidemiologic characteristics, and clinical characteristics of monoinfection and coinfection in diarrheal diseases in children in Tanzania. *American Journal of Epidemiology* 2017; **186**: 1074–1083.
250. **Donaldson AL, *et al.*** Symptom profiling for infectious intestinal disease (IID): A secondary data analysis of the IID2 study. *Epidemiology and Infection* 2019. Published online: 2019. doi:10.1017/S0950268819001201.
251. **Aoki Y, *et al.*** Duration of norovirus excretion and the longitudinal course of viral load in norovirus-infected elderly patients. *Journal of Hospital Infection* 2010; **75**: 42–46.
252. **Anderson EJ, *et al.*** Rotavirus in adults requiring hospitalization. *Journal of Infection* 2012; **64**: 89–95.
253. **Nakajima H, *et al.*** Winter seasonality and rotavirus diarrhoea in adults. *Lancet* 2001; **357**: 1950.

Appendix A: Systematic review PRISMA 2009 checklist

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	Page 51
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	Page 53
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	Page 54
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	Page 55
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	Page 55
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	Page 55
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	Page 55
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Page 56
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	Page 56
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	Page 56

Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	Page 56
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	N/A
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	Page 57
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	Page 57

Appendix B: Systematic review search terms by database

OVID MEDLINE(R) <1946 to Present>

1	population surveillance/ or public health surveillance/ or sentinel surveillance/ or surveillance .mp.
2	syndromic surveillance.mp.
3	attend*.mp.
4	Absenteeism/ or absen*.mp.
5	registers.mp.
6	1 or 2 or 3 or 4 or 5
7	school.mp. or Schools/
8	school aged children.mp.
9	school children.mp.
10	7 or 8 or 9
11	Infectious disease.mp. or Communicable Diseases/
12	Outbreaks.mp. or Disease Outbreaks/
13	epidemics.mp. or EPIDEMICS/
14	pandemics.mp. or PANDEMICS/
15	bugs.mp.
16	11 or 12 or 13 or 14 or 15
17	6 and 10 and 16

Web of Science

1	**TOPIC:** (surveillance OR "population surveillance" OR "public health surveillance" OR "sentinel surveillance" OR "syndromic surveillance" OR attend* OR absen* OR register)
2	**TOPIC:** (school* OR "school*children" OR "school*aged children")
3	**TOPIC:** ("infectious disease" OR "communicable disease" OR outbreak* OR epidemic* OR pandemic* OR bug*)
4	#1 AND #2 AND #3

Pubmed

1	(("surveillance"[Title/Abstract]) OR "attendance"[Title/Abstract]) OR "absenteeism"[Title/Abstract]) OR "register"[Title/Abstract]

2 (("school"[Title/Abstract]) OR "school children"[Title/Abstract]) OR "school aged children"[Title/Abstract]

3 (((((("infectious disease"[Title/Abstract]) OR "communicable disease"[Title/Abstract]) OR "outbreak"[Title/Abstract]) OR "pandemic"[Title/Abstract]) OR "epidemic"[Title/Abstract]) OR "bugs"[Title/Abstract]

4 #1 AND #2 AND #3

Scopus

1 ABS (surveillance OR attendance OR absenteeism OR register)

2 ABS (school* OR "school*children" OR "school*aged children")

3 ABS ("infectious disease" OR "communicable disease" OR outbreak* OR epidemic* OR pandemic* OR bug*)

4 #1 AND #2 AND #3

Science Direct*

1 (surveillance OR attendance OR absenteeism OR register) AND (school) AND ("infectious disease" OR "communicable disease" OR outbreak OR pandemic)

Biosis Previews

1 **TOPIC:** (surveillance OR "population surveillance" OR "public health surveillance" OR "sentinel surveillance" OR "syndromic surveillance" OR attend* OR absen* OR register)

2 **TOPIC:** (school* OR "school*children" OR "school*aged children")

3 **TOPIC:** ("infectious disease" OR "communicable disease" OR outbreak* OR epidemic* OR pandemic* OR bug*)

4 #1 AND #2 AND #3

Open Grey

1 Health surveillance AND school

ProQuest

1 Ab(surveillance) AND ab(school) AND ab(disease outbreaks)

*Limited to 8 Boolean operators. Terms searched for within title, abstract and keywords.

Appendix C: Systematic review data extraction form

Background information

Article title	
Authors	
Journal	
Year of publication	

Methods

Country (+/- area)	
Prospective or retrospective	
Age group(s)	
School type	
Sample size	
Period of data collection	
Organism / syndrome	
Purpose of system (case or outbreak ascertainment, pandemic or seasonal trends)	
Case/absence definition (i.e. missed class, half day, full day)	
Outbreak definition / absence threshold (if applicable)	
Primary outcome measure	
School-level data collected - Numerator - Denominator	
Specificity of data recorded (all-cause, illness, syndrome specific, micro confirmed)	
Timeliness of data reporting from schools	
Spatial and temporal aggregation of data	
Other surveillance systems / data used	
Methods of data analysis	

Findings

Absence rates (+ 95% CI)	
Correlation with other surveillance measures (+ p-values)	
Lead / lag time compared to other surveillance	
Sensitivity of case/outbreak detection	

Appendix D: Study protocol manuscript

School attendance registers for the syndromic surveillance of infectious intestinal disease (IID) in UK children: a study protocol

Abstract

Background:

Infectious intestinal disease (IID) is common and children are more likely than adults both to suffer from IID and to transmit infection onto others. Before the introduction of the vaccine, rotavirus was the leading cause of severe childhood diarrhoea, with norovirus and *Campylobacter* predominate pathogens. Public health surveillance of IID is primarily based on healthcare data and as such illness which is managed within the community will often go undetected. School attendance registers offer a novel dataset that has the potential to identify community cases and outbreaks of IID which would otherwise be missed by current health surveillance systems. Whilst studies have explored the role of school attendance registers in the monitoring of influenza amongst children, no studies have been identified which consider this approach in the surveillance of IID.

Objective:

The aim of this study is to explore the role and utility of school attendance registers in the detection and surveillance of IID in children. The secondary aims are to estimate the burden of IID on school absenteeism and to assess the impact of the rotavirus vaccine on illness absence amongst school-aged children.

Methods:

This study is a retrospective analysis of school attendance registers to investigate whether school absences due to illness can be used to capture seasonal trends and outbreaks of infectious intestinal disease amongst school-aged children. School absences in Merseyside, UK will be compared and combined with routine health surveillance data from primary care, laboratories and telehealth services. These data will be used to model spatial and temporal variations in the incidence of IID and to apportion likely cause to changes in school absenteeism trends. This will be used to assess the potential utility of school attendance data in the surveillance of IID and to estimate the burden of IID absenteeism in schools. It will also inform an analysis of the impact of the rotavirus vaccine on disease within this age-group.

Results:

This study has received ethical approval from the University of Liverpool Research Ethics Committee (Reference number 1819). Use of General Practice data has been approved for the evaluation of rotavirus vaccination in Merseyside by NHS Research Ethics Committee, South Central-Berkshire REC Reference: 14/SC/1140.

Conclusions:

This study is unique in considering whether school attendance registers could be used to enhance the surveillance of IID. Such data have multiple potential applications and could improve the identification of outbreaks within schools, allowing early intervention to reduce transmission both within and outside of school settings. These data have the potential to act as an early warning system, identifying infections circulating within the community before they enter healthcare settings. School attendance data could also inform the evaluation of vaccination programmes such as rotavirus and, in time, norovirus.

Key words: Syndromic surveillance, schools, children, absenteeism, infectious intestinal disease, diarrhoea and vomiting, school attendance registers

Introduction

Infectious intestinal diseases (IID) are common in both high and low income countries, causing an estimated 2 billion cases globally each year.[1] Norovirus is the leading cause of IID, with *Campylobacter* the most common bacterial cause.[1–3] In children, rotavirus has been a major cause of severe IID until the licensing of the vaccine in 2006.[4] The high incidence of IID infection results in significant disease burden and economic costs due to work and school absenteeism, lost earnings, reduced workforce productivity and increased healthcare usage.[5–7] In the UK alone, IID has been estimated to result in one million additional General Practice consultations each year,[6] and norovirus, rotavirus and *Campylobacter* combined cost the UK economy an estimated £150 million per annum.[5] Over 80% of total costs are borne by patients, driven by lost income and out-of-pocket expenses.[5]

Children are disproportionately affected by IID, with those under 5 years accounting for 38% of foodborne cases globally.[1] Children are thought to be important transmitters of IID infection and experience prolonged symptoms and viral shedding, reduced immunity and higher levels of infectiousness.[8–12] The majority of a child's close contacts are based at school and home [13,14]

and infections, especially viruses, can spread easily through these semi-enclosed environments.[15] This not only increases the risk of outbreaks within school settings, but also provides a pathway through which infections can spread from schools into the wider community.[13,16,17] There is evidence that children may be the first affected by seasonal and pandemic disease,[18–21] and hence enhancing infectious disease surveillance in schools could not only improve the health of children, but could also provide advanced warning before infections start to circulate in the wider community.

Public health surveillance of IID is primarily based on healthcare data such as laboratory reports, statutory notifications, hospital admissions, primary care consultations and calls to remote telehealth services.[22,23] The majority of IID cases, however, will be managed in the community, without involvement from healthcare services. As a result, current surveillance is likely to be significantly underestimating the impact of IID. Furthermore, there is an inherent bias in the surveillance of IID, as certain groups are more susceptible to complications and therefore more likely to present to healthcare, such as the very young, the comorbid and the elderly.[2,3,24,25] Laboratory testing policies can also be targeted towards detecting pathogens in these high risk groups,[26] further increasing the surveillance bias. Enhancing the surveillance of IID and improving detection of community cases of disease would provide important information on the epidemiology of these infections. Such data would be of value to support the evaluation of public health interventions, such as rotavirus vaccination and, in time, norovirus vaccination. As vaccinations can alter the epidemiology of infection,[27] it is crucial we are able to accurately monitor the long-term impact and effectiveness of these interventions, not just on healthcare services but also prevalence in the community.

School attendance registers offer a novel dataset which could be used to identify community cases of IID which might not otherwise be detected. School absenteeism data have shown potential in the surveillance of both seasonal and pandemic influenza,[28–35] but no studies have been identified which consider their role in monitoring IID. Whilst mild cases of diarrhoea and vomiting will not necessitate contact with healthcare services, they are likely to still result in an absence from school for the duration of the illness and, in line with Public Health guidance, an additional 48hrs after symptoms have resolved.[36] This provides a routine dataset which has the potential to capture illness from the day of onset.

This study aims to explore the role and utility of school attendance registers in the detection and surveillance of IID in children. The secondary aims are to estimate the burden of IID on school absenteeism and to assess the impact of the rotavirus vaccine on illness absence amongst school-aged children.

Methods

Study setting

This study will take place in Local Government areas within Merseyside in the North West of England. Merseyside is a predominately urban, metropolitan county with a population of 1.38 million, over 240 000 of whom are school-aged children.[37] It comprises five Local Government areas, which range in size from 145 000 residents to over 450 000 residents.[37] For this study, the population of interest is children aged 4 to 16 years who are registered at a school within Merseyside.

Study design

The study will be a retrospective analysis of school absenteeism data to investigate whether school attendance registers can be used to capture seasonal trends and outbreaks of infectious intestinal disease amongst school-aged children. Whilst these data are routinely collected by Local Government for school attendance management,[38] this is a novel application of this dataset. In the United Kingdom (UK), all absences due to illness are given a single code, which distinguishes them from absences due to other causes, including those to attend medical appointments. However, the nature of the illness is not reported. Routine health surveillance data from primary care, laboratories and the NHS 111 telehealth service will be used to model spatial and temporal variations in the incidence of IID and to apportion likely cause to changes in school absenteeism trends. This will allow an assessment to be made of the potential value and lead time of school absenteeism data in the surveillance of IID and the overall burden of IID on illness absenteeism. The impact of the rotavirus vaccine, which was introduced in the UK in 2013, will also be explored. As none of the school-aged children included in this study will have received the rotavirus vaccine, this study will capture the impact of vaccinating infants on herd immunity and reducing illness absenteeism amongst older, unvaccinated children.[27]

Data sources

School absenteeism data is available at individual school level. Attendance data for schools providing primary (4-11 year olds) and secondary (11-16 year olds) education, regardless of type of school, will be sought from Local Government in Merseyside, with data broken down by school and year group. Total absences and absences due to illness will be requested. Details of the number of children in each school and year group will also be obtained to allow corresponding rates to be calculated.

Laboratory data reported to Public Health England (PHE) North West will be used to obtain organism-specific rates of IID within the different geographical areas. These data are routinely collected and reported to Public Health England from diagnostic and reference laboratories.[39] Public Health England also holds data from NHS 111, which is a telehealth service that operates across England.[40]

Calls to NHS 111 (and its precursor, NHS Direct) for diarrhoea and/or vomiting will be used to indicate probable cases of IID. NHS 111 and NHS Direct data are held securely by the Public Health England Real-time Syndromic Surveillance team (ReSST) and can be accessed with permission via PHE.

Primary care consultations for diarrhoea and/or vomiting will be used as another measure of probable cases of IID. These data have recently been collected from Clinical Commissioning Groups and General Practices across Merseyside to inform an evaluation of the rotavirus vaccine.[41] Read Codes were used to distinguish acute cases of IID from cases linked to chronic conditions or non-infective causes.[41] These data can be accessed from the University of Liverpool in an anonymised format as a secondary dataset to further inform the evaluation of the rotavirus vaccine.

To facilitate the spatio-temporal modelling, numbers from each dataset will be aggregated to weekly rates to enable a common timescale. The spatial measurement will depend upon the data source; for school absenteeism data the postcode of the school will be used alongside the catchment area (where appropriate). Primary care consultation data has been mapped to Lower Super Output Areas (LSOA), which represents a geographical area with between 1 000 and 3 000 residents.[42] Laboratory data contains full postcodes, but to protect the anonymity of patients these will be reduced to LSOAs before the data is transferred to the research team for analysis. Telehealth data contains only the postcode district of patients,[43] which is a larger geographical aggregation than LSOA, ensuring anonymity. Denominator populations will be derived from the Office for National Statistics (ONS) mid-year population estimates.[42] Comparison of derived population estimates will be made with the Health and Safety Laboratories National Population Database [44] and the most suitable denominator population will be used.

To allow the analysis to be conducted at year-group level, the surveillance data will include details of the age of the patient (year of birth), and their sex. All other Personally Identifiable Information (PII) will be removed from the data before it is transferred to the research team. The outcomes of the analysis will be based on aggregated data.

Study period

Data will be examined retrospectively from July 2007 to June 2016, capturing nine IID seasons. Each season is considered to start in calendar week 27 and end in calendar week 26 of the following year.

Population sample

This study will focus on three of the five Local Government areas within Merseyside in order to reflect the coverage of primary care data collected to inform an evaluation of the rotavirus vaccine.[45] The population sample was estimated using data from the Department for Education, which holds a record of all Local Government registered schools.[46] Data were based on the 2017/2018 academic year, limited to schools providing primary and secondary education (ages 4-16 years).

The total number of schools across the three Local Government areas is 372, consisting of 299 primary schools and 103 secondary schools. Thirty of the schools deliver both primary and secondary education. The total pupil population across all included schools is 140 164. Assuming that each year one in four pupils are affected by IID,[47] in each academic year we estimate there would be approximately 35 000 cases of IID in schools within the study area. As data will be requested over a 9-year period, the total number of cases across the study period is estimated to be 315 000.

Case definitions

The case definitions used within each dataset are outlined in Textbox 1.

Textbox 1. Case definitions

School attendance registers

- Absence with registration code 'I' (Illness, not medical or dental appointments)

NHS 111 calls

- Calls for vomiting
- Calls for diarrhoea

General Practice consultations (read codes in parenthesis)

- Diarrhoea and vomiting (19G)
- Diarrhoea symptom NOS (19F6)
- Viral gastroenteritis (A07y0)
- Diarrhoea (19F2)
- Gastroenteritis—presumed infectious origin (A0812)
- Diarrhoea of presumed infectious origin (A083)
- Infantile viral gastroenteritis (A07y1)
- Infectious gastroenteritis (A0803)
- Enteritis due to rotavirus (A0762)
- Infectious diarrhoea (A082)

Laboratory detections

- Detection of bacterial, viral or protozoal IID organisms in a faecal specimen

Recruitment and consent

Recruitment will be conducted at a Local Government level. Local Government will be approached via their Public Health departments and invited to participate in this study. Consent for use of aggregated school attendance data will be sought from the Local Government, who carry the legal responsibility for the data and its usage. As the data are aggregated and anonymised, consent will not be sought from individual schools or parents.

Data analysis

A descriptive analysis will be undertaken of each dataset to examine and describe the temporal trends and seasonality of illness absenteeism rates, and of confirmed and probable cases of IID. The analysis will be stratified by age to capture varying rates of disease in different year groups. Rotavirus-specific incidence data will be obtained from laboratory reports. The mathematical and statistical analysis will include an organism-specific dynamic transmission model as well as mixed effect regression analysis to apportion cause to the variations in absenteeism and to estimate organism-specific incidence rates. The complexity of dynamical models will be decided during the project based on the outputs of the descriptive analysis. Rotavirus modelling will include an interrupted time series analysis to explore changes in school absenteeism rates pre- and post- the introduction of the vaccine. This will support an assessment of the impact of vaccination on disease transmission in the community. Other organisms which commonly cause IID in children will also be included in the analysis (e.g. norovirus and *Campylobacter*) to test the ability of illness absenteeism data to accurately detect seasonal trends and outbreaks of disease. This will inform an assessment of the suitability of school attendance registers as a potential form of disease surveillance in the community and its role in the long-term monitoring of vaccine-preventable diseases.

Results

This study received ethical approval from the University of Liverpool Research Ethics Committee (Reference number 1819). Use of General Practice data has been approved for the evaluation of rotavirus vaccination in Merseyside by NHS Research Ethics Committee, South Central-Berkshire REC Reference: 14/SC/1140. Study findings will be submitted to open access peer-reviewed journals and presented at scientific conferences and meetings, including meetings with stakeholders.

Discussion

Current surveillance of infectious intestinal disease is predominantly based on healthcare data and therefore illness which is managed within the community will often go undetected. This study is unique

in considering whether school absenteeism data could be used to enhance the surveillance of IID. The findings could have several significant applications. These data could support the improved identification of outbreaks in schools, allowing early intervention to reduce transmission both within and outside of the school setting. As children may be the first affected by seasonal illness, these data have the potential to act as an early warning system, identifying infections circulating within the community before they enter healthcare settings. Absenteeism data could also be utilised to inform the evaluation of vaccination programmes such as rotavirus and potentially, in time, norovirus. Similarly, these data could be used to monitor the impact of health improvement programmes such as hand washing interventions.

However, there are some limitations which should be considered. The most pertinent is the low specificity of the case definition for illness absenteeism. As a single code is used for all causes of illness absenteeism, these data cannot distinguish between absences caused by IID and absences from other illnesses such as respiratory tract infections. Therefore, the burden of IID on absenteeism cannot be directly measured and modelling of routine surveillance data is required to apportion likely cause to changes in absenteeism rates. A further consideration in this study is the spatial measure available within each dataset; the NHS 111 telehealth service does not collect information below the level of postcode district and hence the statistical modelling, when including this dataset, will be restricted to this geographical level. This limits the ability of this analysis to test whether school absenteeism data can detect localised outbreaks of IID within communities. However, this reflects a limitation within our current surveillance systems and one which school attendance data has the potential to rectify. Future work should consider the feasibility of collecting symptom-specific absence information from schools to enhance the specificity of the data and support the syndromic surveillance of a broader range of childhood infectious disease.

Acknowledgements

Author contributions

AD, JPH, RV, DH and SOB all contributed to the study design. IH contributed to the statistical methods. AD wrote the first draft of the manuscript. All authors read and approved the final manuscript.

Funding statement

This 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 Quadram Institute [Grant number NIHR HPRU 2012-10038]. DH is funded by a National Institute for

Health Research (NIHR) Post-doctoral Fellowship (PDF-2018-11-ST2-006). IH is supported by the UKRI through the JUNIPER modelling consortium (grant no. MR/V038613/1). The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, the Department of Health and Social Care or Public Health England.

Conflicts of Interest

DH report grants on the topic of rotavirus vaccines, outside of the submitted work, from GlaxoSmithKline Biologicals, Sanofi Pasteur and Merck and Co (Kenilworth, NJ, USA) after the closure of Sanofi Pasteur-MSD in December 2016. RV report grants on the topic of rotavirus vaccines, outside of the submitted work, from GlaxoSmithKline Biologicals. AD, JPH, IH and SOB have nothing to disclose.

Abbreviations

IID: Infectious Intestinal Disease

LSOA: Lower Super Output Area

NHS: National Health Service

PHE: Public Health England

ReSST: Real-time Syndromic Surveillance team

UK: United Kingdom

References

1. Kirk MD, Pires SM, Black RE, et al. World Health Organization Estimates of the Global and Regional Disease Burden of 22 Foodborne Bacterial, Protozoal, and Viral Diseases, 2010: A Data Synthesis. *PLOS Med*. 2015;12(12):e1001921. doi:10.1371/journal.pmed.1001921

2. Acheson D, Allos BM. Campylobacter jejuni Infections: Update on Emerging Issues and Trends. *Clin Infect Dis*. 2001;32(8):1201-1206. doi:10.1086/319760

3. Glass RI, Parashar UD, Estes MK. Norovirus Gastroenteritis. *N Engl J Med*. 2009;361(18):1776-1785. doi:10.1056/NEJMra0804575

4. Tate JE, Burton AH, Boschi-Pinto C, Steele AD, Duque J, Parashar UD. 2008 estimate of worldwide rotavirus-associated mortality in children younger than 5 years before the introduction of universal rotavirus vaccination programmes: a systematic review and meta-analysis. *Lancet Infect Dis*. 2012;12(2):136-141. doi:10.1016/S1473-3099(11)70253-5

5. Tam CC, O'Brien SJ. Economic Cost of Campylobacter, Norovirus and Rotavirus Disease in the United Kingdom. *PLoS One*. 2016;11(2):e0138526. doi:10.1371/journal.pone.0138526

6. Tam CC, Rodrigues LC, Viviani L, et al. Longitudinal study of infectious intestinal disease in the UK (IID2 study): incidence in the community and presenting to general practice. *Gut*. 2012;61(1):69-77. doi:10.1136/gut.2011.238386

7. Roberts JA, Cumberland P, Sockett PN, et al. The study of infectious intestinal disease in England: socio-economic impact. *Epidemiol Infect*. 2003;130(1):1-11. doi:10.1017/s0950268802007690

8. Harris JP, Iturriza-Gomara M, Allen DJ, Kelly S, O'Brien SJ. Norovirus strain types found within the second infectious intestinal diseases (IID2) study an analysis of norovirus circulating in the community. *BMC Infect Dis*. 2019;19(1):1-8. doi:10.1186/s12879-019-3706-z

9. Shah MP, Hall AJ. Norovirus Illnesses in Children and Adolescents. *Infect Dis Clin North Am*. 2018;32(1):103-118. doi:10.1016/j.idc.2017.11.004

10. Kirkwood CD, Streitberg R. Calicivirus shedding in children after recovery from diarrhoeal disease. *J Clin Virol*. 2008;43(3):346-348. doi:10.1016/J.JCV.2008.08.001

11. Simmons K, Gambhir M, Leon J, Lopman B. Duration of Immunity to Norovirus Gastroenteritis. *Emerg Infect Dis*. 2013;19(8):1260. doi:10.3201/EID1908.130472

12. Uhnoo I, Olding-Stenkvist E, Kreuger A. Clinical features of acute gastroenteritis associated with rotavirus, enteric adenoviruses, and bacteria. *Arch Dis Child*. 1986;61(8):732-738. doi:10.1136/adc.61.8.732

13. Béraud G, Kazmercziak S, Beutels P, et al. The French Connection: The First Large Population-Based Contact Survey in France Relevant for the Spread of Infectious Diseases. *PLoS One*. 2015;10(7):e0133203. doi:10.1371/journal.pone.0133203

14. Glass LM, Glass RJ. Social contact networks for the spread of pandemic influenza in children and teenagers. *BMC Public Health*. 2008;8(1):61. doi:10.1186/1471-2458-8-61

15. Barker J, Stevens D, Bloomfield SF. Spread and prevention of some common viral infections in community facilities and domestic homes. *J Appl Microbiol*. 2001;91(1):7-21. doi:10.1046/j.1365-2672.2001.01364.x

16. Cauchemez S, Bhattarai A, Marchbanks TL, et al. Role of social networks in shaping disease transmission during a community outbreak of 2009 H1N1 pandemic influenza. *Proc Natl Acad Sci U S A*. 2011;108(7):2825-2830. doi:10.1073/pnas.1008895108

17. Mossong J, Hens N, Jit M, et al. Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. *PLoS Med*. 2008;5(3):e74. doi:10.1371/journal.pmed.0050074

18. Bernard H, Hoehne M, Niendorf S, Altmann D, Stark K. Epidemiology of norovirus gastroenteritis in Germany 2001-2009: eight seasons of routine surveillance. *Epidemiol Infect*. 2014;142(1):63-74. doi:10.1017/S0950268813000435

19. Cate TR. Clinical manifestations and consequences of influenza. *Am J Med*. 1987;82(6):15-19. doi:10.1016/0002-9343(87)90555-9

20. Glezen WP, Couch RB. Inter pandemic influenza in the Houston area 1974-1976. *N Engl J Med*. 1978;298(11):587-592. doi:10.1056/nejm197803162981103

21. Cox NJ, Subbarao K. Global Epidemiology of Influenza: Past and Present. *Annu Rev Med*. 2000;51(1):407-421. doi:10.1146/annurev.med.51.1.407

22. Elliot AJ, Morbey RA, Hughes HE, et al. Syndromic surveillance – a public health legacy of the London 2012 Olympic and Paralympic Games. *Public Health*. 2013;127(8):777-781. doi:10.1016/J.PUHE.2013.05.007

23. Wall PG, de Louvois J, Gilbert RJ, Rowe B. Food poisoning: notifications, laboratory reports, and outbreaks--where do the statistics come from and what do they mean? *Commun Dis Rep CDR Rev*. 1996;6(7):R93-100. PMID: 8680502

24. Velázquez FR. Protective Effects of Natural Rotavirus Infection. *Pediatr Infect Dis J*. 2009;28(Supplement):S54-S56. doi:10.1097/INF.0b013e3181967c03

25. Kirk MD, Veitch MG, Hall GV. Gastroenteritis and Food-Borne Disease in Elderly People Living in Long-Term Care. *Clin Infect Dis*. 2010;50(3):397-404. doi:10.1086/649878

26. Atchison CJ, Lopman BA, Harris CJ, Tam CC, Iturriza Gómara M, Gray JJ. Clinical laboratory practices for the detection of rotavirus in England and Wales: can surveillance based on routine laboratory testing data be used to evaluate the impact of vaccination? *Eurosurveillance*. 2009;14(20):19217. doi:10.2807/ese.14.20.19217-en

27. Iturriza-Gómara M, Cunliffe N. Rotavirus vaccine: a welcome addition to the immunisation schedule in the UK. *BMJ*. 2013;346:f2347. doi:10.1136/bmj.f2347

28. Besculides M, Heffernan R, Mostashari F, Weiss D. Evaluation of school absenteeism data for early outbreak detection, New York City. *BMC Public Health*. 2005;5. doi:10.1186/1471-2458-5-105

29. Chu A, Savage R, Whelan M, et al. Assessing the Relative Timeliness of Ontario's Syndromic Surveillance Systems for Early Detection of the 2009 Influenza H1N1 Pandemic Waves. *Can J Public Heal*. 2013;104(4):e340-e347. doi:10.17269/cjph.104.3721

30. Kara EO, Elliot AJ, Bagnall H, et al. Absenteeism in schools during the 2009 influenza A(H1N1) pandemic: a useful tool for early detection of influenza activity in the community? *Epidemiol Infect*. 2012;140(7):1328-1336. doi:10.1017/S0950268811002093

31. Kightlinger L, Horan V. School illness absenteeism during 2009 influenza A (H1N1) pandemic--South Dakota, 2009-2010. *S D Med*. 2013;66(5):177,179-181. PMID: 23798263

32. Kom Mogto CA, De Serres G, Douville Fradet M, et al. School absenteeism as an adjunct surveillance indicator: experience during the second wave of the 2009 H1N1 pandemic in Quebec, Canada. *PLoS One*. 2012;7(3):e34084. doi:10.1371/journal.pone.0034084

33. Lenaway DD, Ambler A. Evaluation of a school-based influenza surveillance system.. *Public Health Rep*. 1995;110(3):333-337. PMID: 7610226

34. Ma T, Englund H, Bjelkmar P, Wallensten A, Hulth A. Syndromic surveillance of influenza activity in Sweden: an evaluation of three tools. *Epidemiol Infect*. 2015;143(11):2390-2398. doi:10.1017/S0950268814003240

35. Schmidt WP, Pebody R, Mangtani P. School absence data for influenza surveillance: A pilot study in the United Kingdom. *Eurosurveillance*. 2010;15(3):1-6. doi:10.2807/ese.15.03.19467-en

36. Public Health England. Health protection in schools and other childcare facilities. 2017. https://www.gov.uk/government/publications/health-protection-in-schools-and-other-childcare-facilities [Accessed Sept 1, 2021].

37. Office for National Statistics. 2011 Census: Population Estimates by Five-Year Age Bands, and Household Estimates, for Local Authorities in the United Kingdom. 2013. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/2011censuspopulationestimatesbyfiveyearagebandsandhouseholdestimatesforlocalauthoritiesintheunitedkingdom [Accessed Sept 1, 2021].

38. Department for Education. School Attendance: Guidance for Maintained Schools, Academies, Independent Schools and Local Authorities. 2020. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/907535/School_attendance_guidance_for_2020_to_2021_academic_year.pdf [Accessed Sept 1, 2021].

39. Public Health England. Laboratory Reporting to Public Health England: A Guide for Diagnostic Laboratories. 2020. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_d

ata/file/926838/PHE_Laboratory_reporting_guidelines_October-2020-v3.pdf [Accessed Sept 1, 2021].

40. National Health Service. NHS 111. Available from: https://www.nhs.uk/using-the-nhs/nhs-services/urgent-and-emergency-care/nhs-111/ [Accessed Sept 1, 2021].

41. Hungerford D, Vivancos R, French N, Iturriza-Gomara M, Cunliffe N. Ecological assessment of the direct and indirect effects of routine rotavirus vaccination in Merseyside, UK using data from multiple health systems: a study protocol. *BMJ Open*. 2014;4(11):e006161. doi:10.1136/bmjopen-2014-006161

42. Office for National Statistics. Lower layer Super Output Area population estimates (supporting information). 2020. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/lowersuperoutputareamidyearpopulationestimates [Accessed Sept 1, 2021].

43. Office for National Statistics: Postal geography. Available from: https://www.ons.gov.uk/methodology/geography/ukgeographies/postalgeography [Accessed Sept 1, 2021].

44. Health and Safety Laboratories: The National Population Database. Available from: https://www.hsl.gov.uk/what-we-do/data-analytics/national-population-database [Accessed Sept 1, 2021].

45. Hungerford D, Vivancos R, Read JM, Iturriza-Gómara M, French N, Cunliffe NA. Rotavirus vaccine impact and socioeconomic deprivation: an interrupted time-series analysis of gastrointestinal disease outcomes across primary and secondary care in the UK. *BMC Med*. 2018;16(1):10. doi:10.1186/s12916-017-0989-z

46. GOV.UK: Department for Education. Available from: https://www.gov.uk/government/organisations/department-for-education [Accessed Sept 1, 2021].

47. Tam CC, Viviani L, Adak B, et al. The Second Study of Infectious Intestinal Disease in the Community (IID2 Study). UK Food Standards Agency, 2012; project no B18021. http://www.food.gov.uk/sites/default/files/media/document/711-1-1393_IID2_FINAL_REPORT.pdf [Accessed Sept 1, 2021].

Appendix E: Ethics approval letter

Central University Research Ethics Committee for Physical Interventions

28 November 2017

Dear Prof O'Brien

I am pleased to inform you that your application for research ethics approval has been approved. Application details and conditions of approval can be found below. Appendix A contains a list of documents approved by the Committee.

Application Details

Reference:	1819
Project Title:	The role of school absence registers in the evaluation of the rotavirus vaccine
Principal Investigator/Supervisor:	Prof Sarah O'Brien
Co-Investigator(s):	Dr John Harris, Mr Dan Hungerford, Prof Neil French, Prof Nigel Cunliffe, Prof Miren Iturriza-Gomara, Dr Anna Donaldson
Lead Student Investigator:	-
Department:	Public Health and Policy
Approval Date:	28/11/2017
Approval Expiry Date:	Five years from the approval date listed above

The application was **APPROVED** subject to the following conditions:

Conditions of approval

- All serious adverse events must be reported via the Research Integrity and Ethics Team (ethics@liverpool.ac.uk) within 24 hours of their occurrence.
- If you wish to extend the duration of the study beyond the research ethics approval expiry date listed above, a new application should be submitted.
- If you wish to make an amendment to the research, please create and submit an amendment form using the research ethics system.
- If the named Principal Investigator or Supervisor leaves the employment of the University during the course of this approval, the approval will lapse. Therefore it will be necessary to create and submit an amendment form using the research ethics system.

- It is the responsibility of the Principal Investigator/Supervisor to inform all the investigators of the terms of the approval.

Kind regards,

Central University Research Ethics Committee for Physical

Interventions ethics@liverpool.ac.uk

0151-795-8355

Appendix - Approved Documents

(Relevant only to amendments involving changes to the study documentation)

The final document set reviewed and approved by the committee is listed below:

Document Type	File Name	Date	Version
Evidence Of Peer Review	Peer Review Proforma - Reviewer 1		
Evidence Of Peer Review	Peer Review Proforma - Reviewer 2		
Study Proposal/Protocol	STUDY PROTOCOL FINAL - Nov 2017	21/11/2017	Final

Appendix F: Illness absenteeism survey for schools in Merseyside

Public Health England

UNIVERSITY OF LIVERPOOL

What impact does sickness absence have on your school?

We in Public Health are interested in supporting schools in Merseyside to reduce sickness absence and illness amongst school-aged children.

We know that children are more likely to get illnesses such as diarrhoea and vomiting, coughs and colds. Children are not only more likely to catch bugs, but they are also more likely to pass them on to someone else. Most of these infections are short lived and children often recover quickly, but because they are so common and spread so easily, they can cause significant disruption and absence from schools across an academic year.

We at Public Health England and the University of Liverpool are interested in exploring which infections most affect illness absence from schools, and how we can use information on illness absence to catch the spread of infections early. Picking up sickness early and preventing it spreading between children could improve the health and wellbeing of children and reduce rates of illness absence from school.

This survey is intended to explore what illness-absence information is currently collected by schools and whether schools record and use additional information on the symptoms of illness. You do not have to take part. If you do, your responses will be kept anonymous. All data will be securely stored and managed in accordance with the Data Protection Act 2018. Ethics approval for this survey has been granted by the University of Liverpool Research Ethics Committee.

About your school

This section is intended to help us understand the type and size of school you work in.

1. What is your role within the school?

2. What type of school do you work in?

Please select as applicable.

- [ ] Pre-school
- [ ] Primary (Infant only)
- [ ] Primary (Junior only)
- [ ] Primary (Infant and Junior)
- [ ] Primary and Secondary
- [ ] Secondary only
- [ ] Special
- [ ] Boarding
- [ ] Other, please specify

3. How many children are currently registered at the school?

Please select one.

- [ ] Less than 50
- [ ] 50-99
- [ ] 100-249
- [ ] 250-499
- [ ] 500-999
- [ ] 1000 or more

Illness absence in school-aged children

The next questions are intended to capture your experience of illness in your school.

4. Which of the following do you think cause the most **illness** in your school?

Please select one or two main causes of illness in your school.

- ☐ Diarrhoea and vomiting
- ☐ Rashes
- ☐ Coughs, colds or flu-like illness
- ☐ Injury
- ☐ Pre-existing health conditions
- ☐ Other, please specify

5. Which illnesses do you think cause the most **absence** in your school?

Please select one or two main causes of illness-absence in your school.

- ☐ Diarrhoea and vomiting
- ☐ Rashes
- ☐ Coughs, colds or flu-like illness
- ☐ Injury
- ☐ Pre-existing health conditions
- ☐ Other, please specify

6. Which age groups do you think are most affected by illness absence in your school?

Please select one or two age groups which you feel are most affected by illness absence.

- ☐ Early years foundation (pre-school and reception)
- ☐ Key stage 1 (Years 1-2)
- ☐ Key stage 2 (Years 3-6)
- ☐ Key stage 3 (Years 7-9)
- ☐ Key stage 4 (Years 10-11)
- ☐ Key stage 5 (Years 11-12)
- ☐ All ages equally affected by illness

7. Have you ever contacted Public Health England for advice or support because of illness or illness absence at your school?

- ○ Yes
- ○ No

8. Do you have a school exclusion policy for illnesses such as diarrhoea and vomiting?

- ○ Yes
- ○ No

Illness absence data

This section explores what information you collect about illness absence and how you use that information within your school.

9. How can parents or care givers inform you of sickness absence?

Select as many as apply.

- [ ] Email
- [ ] Phone call to reception staff
- [ ] Automated answer phone
- [ ] Online
- [ ] Text message
- [ ] Letter
- [ ] Other, please specify

10. Do you collect information from parents/care givers on cause or symptoms of illness?

- ○ Yes
- ○ No

11. If yes, what information do you collect?

Please select one or more options.

- [ ] Main cause of illness
- [ ] Brief description of illness
- [ ] Whatever a parent chooses to tell us
- [ ] Pre-determined 'category' of illness
- [ ] Doctor's note / sick note
- [ ] We do not collect information on their symptoms
- [ ] Other, please specify

12. How is this information recorded?

Please select one.

- [ ] Asked but not recorded
- [ ] Paper record
- [ ] Electronic record
- [ ] Not applicable
- [ ] Other, please specify

13. How do you use this information?

Please select as many as apply.

- [ ] Not routinely used
- [ ] Used to validate reason for absence
- [ ] Used to monitor sickness within school
- [ ] Used to pick up spread of illness amongst children
- [ ] Used to share relevant information with parents about illnesses
- [ ] Reported to Local Authority
- [ ] Not applicable
- [ ] Other, please specify

14. What impact, if any, does this have on how you manage illness in your school?

15. Do you think there would be a benefit to schools if information was routinely collected on causes/symptoms of illness absence?

- ○ Yes
- ○ No

Future work

16. Would you be interested in taking part in further work to explore how information about illness absence could be used to better understand and prevent illness in children?

- ○ Yes
- ○ No

17. If yes, please leave contact details below. These details will only be used to contact you in relation to this work and will not be shared outside of the project team.

School name:

School address:

Named contact for school:

Contact telephone number:

Contact email address:

Preferred time to be contacted:

Appendix G: Ethics amendment approval letter

Central University Research Ethics Committee for Physical Interventions

28 June 2019

Dear Prof O'Brien,

I am pleased to inform you that the amendment to your study has been approved. Amendment details and conditions of approval can be found below. If applicable, Appendix A contains a list of documents approved by the Committee.

Amendment details

Reference:	1819 (amendment)
Project Title:	The role of school absence registers in the evaluation of the rotavirus vaccine
Principal Investigator:	Prof Sarah O'Brien
Co-Investigator(s):	Prof Miren Iturriza-Gomara, Prof Neil French, Prof Nigel Cunliffe, Mr Dan Hungerford, Dr Anna Donaldson, Dr John Harris
Lead Student Investigator:	-
Department:	Public Health and Policy
Approval Date:	28/06/2019

The amendment was **APPROVED** subject to the following conditions:

Conditions of approval

- All serious adverse events must be reported to the Committee (ethics@liv.ac.uk) in accordance with the procedure for reporting adverse events.
- If it is proposed to make further amendments to the study, please create and submit an amendment form within the research ethics system.
- It is the responsibility of the Principal Investigator or Supervisor to inform all the investigators of the terms of the approval.

Kind regards,

Central University Research Ethics Committee for Physical Interventions
ethics@liverpool.ac.uk
0151-795-8355

Appendix - Approved documents

If applicable, the final document set reviewed and approved by the committee is listed below:

Document Type	File Name	Date	Version
Default	Illness absenteeism in schools - survey	31/05/2019	1