The epidemiology of Shiga toxin-producing Escherichia coli (STEC) in England, and the impact of advanced diagnostics on our understanding and control

By

Lisa Anne Byrne

Thesis submitted for consideration for the degree of

Doctor of Philosophy in Health Sciences (by published work)

Warwick Medical School, University of Warwick

September, 2017

Contents

Acknowledgements

I sincerely thank my supervisor at the University of Warwick, Dr Noel McCarthy, for his support, encouragement and reassurance during the process of registration and in the production and submission of this thesis. I thank my colleagues, collaborators and friends, Natalie Adams, Dr Claire Jenkins, Dr Tim Dallman and Professor Bob Adak for their professional and personal support , both in respect to this body of work and beyond during the time I have worked for Public Health England. I thank Jamie Cunningham, for his continued support and encouragement in all my endeavours.

Submission declaration

I, Lisa Byrne, declare that the submitted material has not been previously submitted or is currently under submission for a degree, diploma, or similar qualification at any university or similar institution.

Summary

Shiga toxin-producing *E. coli* (STEC) are associated with human illness and are defined by the presence of phage encoded Shiga toxin genes. While a relatively rare cause of gastrointestinal illness, they have potential to cause severe illness including bloody diarrhoea and Haemolytic Ureamic Syndrome (HUS), a serious and life threatening condition affecting the blood and kidneys and the most common cause of acute renal failure in children in the UK. The main reservoir for STEC is cattle and transmission can occur through direct or indirect contact with animals, consumption of contaminated food or water and through person to person spread.

STEC first emerged as a pathogen of concern in the early 1980's and have since persisted as a pathogen of public health concern. Despite this, there was a paucity of published epidemiological data on STEC cases in the UK and elsewhere. In 2009, Public Health England introduced an enhanced surveillance system for STEC collecting standardised and detailed microbiological, clinical and epidemiological data on all cases in England. Recent advances in microbiological methods for the detection and typing of STEC have occurred.

In this thesis by published works, I use six publications utilising these methodological advancements. Together, these papers provide a detailed update on the epidemiology of STEC infection in England, including an assessment of the impact of molecular methods (PCR, MLVA and WGS), on our understanding and control of STEC. Specifically, I comprehensively describe the past and current epidemiology of STEC in England including clinical impact, and assess the impact of the recent introduction of advanced detection and typing techniques in England. The body of work as a whole provides a future perspective on where research is needed as well as informing operational implications for detecting and managing sporadic cases and clusters of infection.

Abbreviations

CI	Confidence Interval
CST	Common Source Threshold
DLV	Double Locus Variant
E. coli	*Escherichia coli*
	Electronic Foodborne and non-foodborne gastrointestinal Outbreak
eFOSS	Surveillance System
EHEC	Enterohaemorrhagic *Escherichia coli*
ESQ	Enhanced Surveillance Questionnaire
GBRU	Gastrointestinal Bacteria Reference Unit
GI	Gastrointestinal Illness
HC	Haemorrhagic Colitis
HPT	Health Protection Team
HUS	Haemolytic Uraemic Syndrome
IID	Infectious Intestinal Disease
IRR	Incidence Rate Ratio
MLST	Multi Locus Sequence Type
MLVA	Multi Locus Variable Number Tandem Repeat Analysis
NESSS	National Enhanced Surveillance System for STEC
NSF	Non Sorbitol Fermenting
PCR	Polymerase Chain Reaction
PHE	Public Health England
PT	Phage Type
SF	Sorbitol Fermenting
SLV	Single Locus Variant
SNP	Single Nucleotide Polymorphism
STEC	Shiga toxin-producing *Escherichia coli*
Stx	Shiga toxin
UK	United Kingdom
USA	Unites States of America
VTEC	Vero cytotoxin-producing *Escherichia coli*
WGS	Whole Genome Sequencing

Index of published work for consideration

Paper	Publication	Author contribution	Citations
1	**Byrne L,** Jenkins C, Launders, R. Elson Adak GK. The Epidemiology, microbiology and clinical impact of Shiga toxin-producing *Escherichia coli* in England, 2009-2012. *Epidemiol* Infect. 2015 Apr 29:1-13	LB designed the study, managed and cleaned the data, performed descriptive and statistical analyses, wrote the manuscript in liaison with co-authors and responded to reviewers as corresponding author.	20
2	Natalie L. Adams, **Lisa Byrne**, Geraldine A. Smith, Richard Elson, John Harris, Roland Salmon, Robert Smith, Sarah O'Brien, Goutam K. Adak, and Claire Jenkins. Shiga toxin-producing Escherichia coli O157 in England and Wales, 1983-2012. *Emerg Infect* Dis. 2016 Apr; 22(4):590-7. doi: 10.3201/eid2204.151485.	LB contributed to the study design and interpretation of data, extracted and cleaned outbreak data, wrote the outbreak section, drafted and critically revised the manuscript alongside co-authors.	7
3	**Byrne L,** Elson R**,** Dallman TJ, Perry N, Ashton P, Wain J, Adak GK, Grant KA, Jenkins C. Evaluating the use of multilocus variable number tandem repeat analysis of Shiga toxin-producing *Escherichia coli* O157 as a routine public health tool in England**.** *PLoS ONE* 9(1): Jan 2014	LB designed the study, managed and cleaned the data, performed all descriptive and statistical analyses, wrote the manuscript in liaison with co-authors and responded to reviewers as corresponding author.	13
4	Dallman TJ,' **Byrne L**, Ashton P,' Cowley LA, Perry NT, , Elson R,	LB (joint first author) led on the epidemiological aspects of	48

	Adak GK, Underwood A, Green J, Jenkins C, Grant KA Wain J. Whole Genome Sequencing for National Surveillance of Shiga Toxin Producing *Escherichia coli* O157. *Clin Infect Dis*. 2015 Apr 17. pii: civ318.	study design, provided epidemiological context to clusters described; Conducted the evaluation of MLVA versus WGS including survival analysis and contributed to the drafting and critical revision of the manuscript.	
5	**Byrne L**, Vanstone G. Perry N, Launders N, Adak GK, Godbole G, Grant K. Smith R., Jenkins C. The epidemiology and microbiology of Shiga-toxin producing *Escherichia coli* other than serogroup O157 in England 2009-2013. J *Med Microbiol*. 2014 Sep;63(Pt 9):1181-8	LB designed the study, managed and cleaned the data, performed all descriptive and statistical analyses, and wrote the manuscript in liaison with co-authors.	20
6	Dallman TJ, **Byrne L**, Launders N, Glen K, Grant KA, Jenkins C. The utility and public health implications of PCR and whole genome sequencing for the detection and investigation of an outbreak of Shiga toxin-producing Escherichia coli serogroup O26:H11. *Epidemiol Infect*. 2014 Oct 15:1-9.	LB (joint first author) provided epidemiological input into the study design, provided the epidemiological context to the study and contributed to the drafting and revising of the manuscript.	13

Introduction

The organism

Vero cytotoxin-producing *Escherichia coli* (VTEC) were first recognised in 1977, following reports that some strains of *E. Coli* from cases of human and animal disease produced cytotoxin acting on Vero cells; hence the name Verocytotoxin.[1] Due to the similarity of the toxins to those in S*higella dysenteriae,* VTEC are also known as Shiga toxin-producing *Escherichia coli* (STEC).For the work presented here, STEC nomenclature is used.

While STEC belong to the wider group of Enterohaemorrhagic *E. coli* (EHEC), they are characterised by the possession of phage encoded shiga toxins (Stx (*stx*)), their main pathogenicity factors. Intimin (*eae*), involved in the intimate attachment of the bacteria to gut mucosa, is also an important factor in causing disease.

The serogroup most associated with human disease in the UK is O157. Elsewhere, serogroups of STEC other than O157 (termed non-O157 STEC) have risen in prominence. Of particular note, was a large food-borne outbreak of STEC in Germany in 2011. The outbreak was caused by a novel strain of *E. coli* O104:H4 possessing Stx2a and intimin and caused 3,816 cases, 845 who developed Haemolytic Uraemic Syndrome (HUS) and 54 who died. [2] Elsewhere, Non-O157 strains are increasingly reported; since 2013 in Ireland non-O157 STEC are reported to cause more human cases per annum than O157.[3] In the UK, due to a diagnostic bias for detection of O157, the prevalence and significance of Non-O157 STEC is unknown.

Clinical significance

Although relatively rare compared with other gastrointestinal infections, STEC infections are of significant public health concern due to the potential disease severity. Symptoms can range from mild diarrhoea to include abdominal cramps, vomiting and severe bloody diarrhoea (Haemorrhagic Colitis (HC)). Infection can lead to the development of HUS a severe multisystem syndrome clinically characterised by a triad of acute kidney injury, microangiopathic haemolytic anaemia and thrombocytopenia.

Following ingestion of the bacteria, STEC enter the intestines and excrete the Stx toxin. These toxins enter the bloodstream via the gut mucosa, whereby they reach

target organs. Renal epithelium is enriched with receptors which Stx binds to resulting in the kidneys bearing the brunt of STEC toxicity. Extra-renal complications with HUS can occur, most frequently neurological complications, such as seizures, reported in 13 - 38% of cases. [4-12] Rarely, HUS is fatal.

The overall proportion of STEC cases which are hospitalised and go on to develop HUS is largely unknown. A previous study undertaken on data collected in 2000 in England, estimated 38% of STEC O157 cases were hospitalised,[13] but no other population based estimates of hospitalisation due to STEC have since been published. In some outbreaks, progression to HUS has been reported in 20% to 65% of cases [14-17].Using data from surveillance of STEC cases, estimates of HUS range between 6.1% to 7.9% cases, (13, 16, 17) and the last reported data in England (1992-4) estimated that 12% of STEC cases developed HUS. (14)

Children are at greater risk and numerous studies indicate that children aged below five years most frequently develop HUS. [5-10, 12, 18-24]

Risk factors for human infection

During the early years of the emergence of STEC, outbreak investigations implicated meat and dairy products as sources of infection and STEC was first isolated from Cattle in England 1989. [25] Subsequently, numerous studies on livestock were undertaken and STEC was frequently isolated from cattle, sheep and goats, with cattle considered the main reservoir of STEC.[26, 27]

Transmission to humans occurs through consumption of contaminated food or water, direct or indirect contact with infected animals or their environment and through person to person spread. Each transmission route can cause sporadic infection as well as outbreaks. STEC has a very low infectious dose and ingesting less than 100 organisms can cause illness.[28-31] The incubation period (between exposure and the occurrence of symptoms) is usually between one and eight days (commonly 3-4 days).

While outbreak investigations provide valuable information on viable vehicles and sources of infection, elucidating risk factors for sporadic infections is challenging. Several analytical studies were undertaken in the early 1990s to ascertain risk factors for sporadic infection and each indicated contact with animal excreta in the

environment as a major risk factor.[32-34] The relative role of transmission routes, their current relevance and trends over time are unknown.

Surveillance and epidemiology of STEC in England

The first laboratory surveillance report for England was published in 1993, for the period 1989-1991 and described a continued increase in detection of STEC cases. The authors described the Stx2 toxin subtype as being more often associated with both HC and HUS, which was reported for 15% of STEC cases. The distinct seasonality of STEC, with most cases occurring in the warmer months of May to September was presented, a phenomenon observed in other countries in the Northern Hemisphere. The incidence was highest in children aged 1-4 years and decreased with each age group, increasing again in the over 85's. Infection rates were higher in adult females, with females over 35 having twice the rate of STEC than males. Incidence varied by region, and was highest in the North of England. A further report was published for the period 1992-1994, and reiterated these findings.

The last and only other surveillance publication for England and Wales covered the period 1995-1998.[35] Further increases in detection of STEC were noted, but otherwise the demographic and geographic distribution of cases remained the same.

There had been no update on the national epidemiology of STEC since 1998. Testing of STEC was only performed on specimens from patients experiencing HC or HUS, which was extended to include all diarrhoeal specimens in 1997. This major change in screening criteria will have impacted on the measurable burden of STEC overall and by sub-groups and no assessment of the change was undertaken until the work presented here. Furthermore, these reports relied on data collected through laboratory referral forms, rather than through a surveillance system. This passive system is subject to incomplete and inaccurate data collection; age and sex were only available for 89-93% of strains submitted. Geographical regions were assigned based on sending laboratories and not patients' residence. Foreign travel and reporting of HUS were not routinely collected and so figures presented were likely underestimated. No clinical data were collected to assess the burden of illness and no other data had been collected on the burden of STEC in terms of hospitalisation and death since the IID1 study in 2000.[13]

Changes in the incidence of infection, host characteristics and microbiological subtypes over time had not been examined and a timely review of the burden of STEC was needed. Further, aside from outbreak reports, assessment of risk factors among sporadic cases had not been made since the 1990's and an up to date assessment on who is at risk of STEC infection and where the risks of infection lie in England was required.

Aims and Objectives

Despite the continued emergence of STEC as a pathogen of public health concern over three decades, there was a paucity of published epidemiological data on clinical STEC cases from the UK and elsewhere. Recently advances in surveillance and microbiological methods pertaining to STEC in England have occurred.

With this in mind, the aim of this body of work was to provide a detailed update on the epidemiology of STEC infection in England including assessing the impact of molecular methods on our understanding and control. To achieve this aim, the research had three objectives.

These were:

1. Characterising the population at risk of STEC and identifying where the risks of infection lie

2. Identifying host and pathogen factors which influence clinical outcome

3. Evaluating the impact of enhanced detection and typing methods on our understanding of the epidemiology and public health impact of STEC

Six publications addressing these objectives are included for consideration.

Methods

The National Enhanced Surveillance System for STEC (NESSS) was implemented on 1st January 2009, and collects standardised and detailed microbiological, clinical and epidemiological data on all STEC cases in England. Using these data I analysed a large cohort of cases with respect to the epidemiology and clinical impact of STEC infection in Papers 1, 3 and 5 (Objectives 1-3). Over the following years, novel molecular methods for the detection and typing of STEC were trialled and the inclusion of microbiological data in NESSS facilitated my assessment and evaluation of these methods and their impact upon our knowledge of clinical infection in England (Table 1).

Electronic Foodborne and non-foodborne gastrointestinal Outbreak Surveillance System (eFOSS)

In 1992, Public Health England began standardized surveillance of general outbreaks of gastrointestinal disease. The methods are reported in detail elsewhere but briefly, local PHE units report standardised epidemiologic data on all outbreaks of gastrointestinal diseases *[36]*.

National enhanced surveillance system for STEC (NESSS)

From 1st January 2009, local laboratories reported presumptive isolates of STEC to Health Protection Teams (HPTs), who undertook public health management of the cases as per guidelines.[37] Stool specimens from patients were sent to the local hospital laboratory where they were tested for the presence of STEC O157. Locally confirmed isolates were sent for confirmation and typing at Gastrointestinal Bacteria Reference Laboratory (GBRU), London, which provides the national reference service for STEC in England and Wales. If STEC O157 was not isolated from a patient with clinical symptoms indicative of STEC (e.g. a child with HUS), the specimen was sent to GBRU and tested for the presence of non-O157 STEC. In cases where a faecal specimen was unavailable a serum sample was sent for serological detection of *E. coli* O157.

Case management by HPTs included taking a detailed case history for the seven days prior to illness using the VTEC Enhanced Surveillance Questionnaire (ESQ) (Appendix IV). These questionnaires were forwarded to the National Gastrointestinal Infections (GI) department and entered into a web-based database for inclusion in NESSS. On a

daily basis GBRU test results were imported into NESSS and reconciled with case data collected through the ESQ. Multiple isolates were received for a number of infected cases (due to follow-up samples being taken for clearance). These results were reviewed and one isolate chosen to represent the case. This would be the first isolate confirming infection or the report with the most information. Cases were assigned epidemiological and microbiological case definitions as set out in Paper 1.

A detailed description of the isolation and typing of STEC at GBRU including changes over time is provided in Paper 2. Molecular methods (Polymerase Chain Reaction (PCR), Multi Locus Variable Number Tandem Repeat Analysis (MLVA) and Whole Genome Sequencing (WGS)) were assessed through the work presented in this thesis and detailed methods are contained within the individual papers.

Table 1. Summary of key surveillance and testing activities utilised for the composition of this work

Surveillance/Testing activity	Time period in place[1]	Time period analysed[2]	Research objective(s)	Papers
Electronic Foodborne and non-foodborne gastrointestinal Outbreak Surveillance System (eFOSS)	1992 to date	01/01/1992-31/12/2012	1	Paper 2
National Enhanced Surveillance system for STEC in England	01/01/2009- to date	01/01/2009-31/12/2012	1-3	Papers 1,3 and 5
Multi-locus variable number tandem repeat analysis (MLVA) for typing of all STEC O157 isolates routinely	01/05/2012-01/08/2015	01/05/2012-31/10/2012	3	Papers 3 and 4
Phased implementation of Polymerase Chain Reaction (PCR) for Gastrointestinal Infections (GI) at frontline hospital laboratories	01/12/2013- *to date*	01/12/2013-	2,3	Papers 5 and 6
Pilot study for whole genome sequencing(WGS) of STEC isolates by GBRU[1]	01/04/2012-31/03/2014	01/05/2012-31/12/2013	3	Papers 4,5 and 6

1. This pilot study included sequencing a subset of genomes from 1983 through to 2014, although the work was undertaken from 1st April 2012 through to 31th March 2014.
2. For the analyses I undertook in the papers only.

Objective 1: Characterising the population at risk of STEC and identifying where the risks of infection lie

Introduction

STEC infections are not evenly distributed amongst the population and some groups are more at risk of infection and severe disease. However prior to the published work undertaken here , the last surveillance publication for England and Wales covered the period 1995-1998 and presented limited data collected through laboratory surveillance.[35] Changes in the incidence of infection, host characteristics and microbiological subtypes over time had not been examined.

Defining risk factors for infection is also important. A small number of case control studies in the 1990s identified contact with animal excreta as an important risk factor for infection in Great Britain.[32-34] Analysing outbreak data collected between 1992 and 2002, identified two distinct aetiologies for foodborne outbreaks of STEC O157 infection, relating to milk and red meat products.[38] However no review of outbreaks, including non-foodborne outbreaks, or consideration of changes in exposure risks over time had been undertaken in England.

An in-depth review of those at risk of STEC and where risk factors of infection lie in England was therefore required. The introduction of the National Enhanced Surveillance System for STEC (NESSS) on 1st January 2009, afforded the opportunity, alongside outbreak surveillance, to address these gaps through the work undertaken in Papers 1, 2 and 5.

Methods

To provide a detailed national picture on clinical STEC infections and assess risks for infection, data from NESSS were described for the first time for a four year period in Paper 1. Incidence rates were calculated by age, gender, ethnicity, geographic region, and rurality of residential post-code. Comparisons of incidence were made between demographic groups through calculation of Incidence Rate Ratios (IRR). The age and gender distribution of STEC cases were compared to those of *Salmonella*

enterica and *Campylobacter* sp. Environmental exposures were examined for primary, domestic sporadic STEC cases. In Paper 2, a review of 30 years of outbreaks of STEC O157 in England and Wales was undertaken, including a description of the causes and settings of outbreaks and relative changes in those over time in order to define where risk factors for STEC lie. In Paper 5, a matched case-case study to compare exposures amongst non-O157 STEC cases to O157 cases was performed.

Results

In Paper 1, a total of 3,717 confirmed and probable STEC cases were reported with a crude incidence of 1.80/100,000 person years, varying by age, sex, ethnicity and region of residence. Incidence was highest among children aged 1 to 4 years (7.63/100,000 person-years), and declined with each subsequent age group reaching a trough in adults aged 20-59. Over 40% of cases were in children, compared to 7.0% of campylobacteriosis and 17.8% of salmonellosis cases. Females had a higher incidence of STEC than males (RR=1.24, p<0.001), but the difference was most prominent in adults aged 20-59 (RR=1.7, p<0.001). Incidence was higher in white ethnic groups as compared to non-white ethnic groups (RR=1.43, p<0.001). The highest incidence was in the Yorkshire and Humber region (2.48/100,000 person-years) and lowest in London (1.04/100,000 person-years). Incidence of STEC was over four times higher among people residing in rural areas than urban areas (RR=4.39, p<0.001).

A fifth of cases comprised known outbreaks and a further fifth of cases were travel related. Non-O157 cases appeared to be more often travel associated than O157 cases (52·3% vs. 20·6%, P = 0·021). However, in Paper 5 non-O157 cases (n=69) were at increased odds (OR=3.29, p<0.001) of contact with farm animals.

Amongst sporadic cases, exposure to farm animals and/or their faeces were associated with infection and were reported twice as often among cases living in rural areas than urban areas (p<0.001). Reported direct and indirect contact with farm animals/faeces increased significantly with increased rurality, in both farm settings (p<0·001) and non-farm settings (p<0·001). Where farm settings were reported, the majority (84.5%) of urban cases reported visiting an open farm or holidaying on a farm, whereas the majority (58.9%) of rural cases reported living, working or having access to a farm through relatives or neighbours.

In Paper 2, vehicles and settings were examined for 335 outbreaks reported between 1982 and 2012. These comprised 3,107 cases, representing 17.4% of cases overall. The median size of an outbreak was five cases (range 2-252). Overall, food vehicles contributed the highest number of outbreaks and associated cases, although they decreased proportionately in the later period of analysis (table 2) .

Thirty-six outbreaks were reportedly attributed to consuming contaminated meat, although most (66.7%) occurred prior to 2003 (n=24). The first implicated food vehicle in an outbreak in this study was raw potatoes in 1985, and six outbreaks associated with raw vegetables or salads were subsequently reported throughout the study period. The largest national outbreak of STEC in Great Britain in 2011, affecting 252 cases was linked to handling raw leeks and potatoes. Between 1983 and 2002, 14/17 dairy related outbreaks were associated with raw milk or post pasteurisation contamination of milk products, but no dairy-related outbreaks were observed thereafter.

Person to person spread as a transmission route in outbreaks increased over time; This appeared to be driven by outbreaks in child-care facilities occurring with increasing frequency, demonstrating that person to person transmission remains an important driver of outbreaks. The number of open farm outbreaks also increased. In the first 20 years of the study, 14 open farm outbreaks were reported. Between 2003 and 2012, 31 petting farm outbreaks were reported, including an outbreak in September 2009, affecting 93 cases.

Table 2. Reported outbreaks of STEC O157 by transmission routes in England and Wales 1983-2002 and 2003-2012.

Time period	1982-2002				2003-2012			
Transmission route/ *subgroup*	No. outbreaks	*%*	No. outbreak cases	*%*	No. outbreaks	*%*	No. outbreak cases	*%*
Person to person	42	*24.4*	332	*22.9*	51	*31.3*	444	*26.8*
Hospital	5	*2.9*	33	*2.3*	*1*	*0.6*	*4*	*0.2*
Care home	10	*5.8*	50	*3.5*	*6*	*3.7*	*26*	*1.6*
Nursery	20	*11.6*	217	*15.0*	*39*	*23.9*	*395*	*23.8*
Other person to person	7	*4.1*	32	*2.2*	*5*	*3.1*	*19*	*1.1*
Animal Contact	21	*12.2*	115	*7.9*	44	*27.0*	313	*18.9*
Foodborne	62	*36.0*	717	*49.5*	38	*23.3*	704	*42.4*
Raw vegetables	5	*2.9*	70	*4.8*	*2*	*1.2*	*263*	*15.9*
Meat[1]	24	*14.0*	280	*19.3*	*12*	*7.4*	*283*	*17.1*
Dairy	17	*9.9*	229	*15.8*	*0*	*0.0*	*0*	*0.0*
Other food	0	*0.0*	0	*0.0*	*5*	*3.1*	*31*	*1.9*
Food vehicle unknown	16	*9.3*	138	*9.5*	*19*	*11.7*	*127*	*7.7*
Waterborne[2]	8	*4.7*	58	*4.0*	6	*3.7*	53	*3.2*
Other	0	*0.0*	0	*0.0*	2	*1.2*	17	*1.0*
Unknown vehicle	39	*22.7*	226	*15.6*	22	*13.5*	128	*7.7*
All outbreaks	172	*100.0*	1448	*100.0*	163	*100.0*	1659	*100.0*

1. Meat outbreaks include handling and consuming undercooked or raw meat and consuming cooked products which have been contaminated.
2. Waterborne outbreaks include drinking or swimming/paddling contaminated water.

Discussion

Paper 1 provided a detailed update on the epidemiology of STEC in England, reporting on a large dataset of cases over a four year period. Collection of standardised information through an enhanced system increased the reliability of the data and provided a more accurate reflection of cases attributable to outbreaks, secondary transmission and travel compared to laboratory surveillance alone.

As described, the crude incidence of STEC was relatively low, varying by age, gender, ethnicity and geographical area of residence. However, there is a diagnostic gap for STEC for two reasons. First, many people with enteric pathogens do not present to health-care and of those who do, only a proportion will have a specimen taken. Of those who do, only a subset will yield a pathogen at local laboratories which will be referred to GBRU for typing. Two investigations aiming to determine the true incidence of infectious intestinal disease (IID) in England were undertaken in 1993-96 (IID1) and 2008-9 (IID2).[39, 40] IID1 estimated that only one in six people with IID presented to a GP. For STEC, IID2 estimated that for each laboratory confirmed case, seven go undetected in the community. Severe illness, recent foreign travel and lower socioeconomic status were all found to be associated with GP presentation.[41] Clinical specimens are likely therefore to reflect more severe disease or foreign travel in those with less severe disease. Second, frontline diagnostics exploit the Non Sorbitol-Fermenting (NSF) properties of STEC O157 and mean that Non-O157 STEC, which often ferment sorbitol are undetected by these methods and their prevalence in England is unknown. Incidence rates reported in Paper 1 can therefore largely be inferred as those for STEC serogroup O157 only.

Children have long been established as at greatest risk of STEC infection and these data were consistent with previous reports from the UK and elsewhere. [35, 42-44] While overall incidence is likely under-estimated, as described above, national surveillance data may also suffer ascertainment bias which could impact upon incidence among different groups. Female adults and children could be more likely to present to health-care and provide a specimen emphasising the impact of disease in those groups. The age distribution for STEC was much more heavily skewed towards children than that for either *Salmonella enterica* or *Campylobacter* sp. However,

STEC disease is more severe in children and any bias may be amplified by the severity of their symptoms.

The low infectious dose and propensity for person-to-person transmission in childcare facilities also increases risk in these age groups.[45-50] Environmental transmission routes are also important for STEC, more so than for *S. enterica* or *Campylobacter* sp; as children practice poorer hygiene than adults, they are more likely to acquire infection, which may also contribute to their higher incidence. The increased incidence from infants under one year to those aged one to four years may reflect this; infants under one are also less mobile and less exposed to contaminated environments.

Higher incidence rates in females had been reported previously in England.[35, 44] However the reasons for this are unknown and the disparity was not apparent for *S. enterica* or *Campylobacter* sp. Host factors may place women at greater risk of severe symptoms and increase their likelihood of presenting to healthcare. [51] Alternatively, adult females may have different behaviours with regards to consumption, food handling and childcare increasing their risk of infection. Incidence rates in England were assessed for the first time by ethnicity (Paper 1), and were higher among those of white ethnicity than those of non-white ethnicity. Interpretation of these data are limited as ethnicity was only reported for 38·5% of cases and ethnicity categories were broad. Collecting more detailed ethnicity data would enable these differences to be explored in more detail, ideally alongside other markers known to impact health outcomes such as socioeconomic status.

The highest incidence rates were reported in the Northern and Western regions, while cases were more likely to live in rural than urban areas. Further work to assess the regional distribution of STEC infections accounting for rurality alongside other factors, such as livestock density is needed to elucidate this relationship. The increased risk associated with rurality and the wide geographical variations in incidence observed across the UK and Ireland appear to reflect environmental exposure to STEC in areas where ruminants are raised.[43, 52-56] Cases living in rural areas were over twice as likely to have had contact with livestock as those from urban areas. These findings demonstrate that environmental/animal contact is an important risk factor for infection amongst sporadic cases. Farm exposures differed between urban and rural cases, with the former being exposed more often to open

farms and rural cases to private farms. Open farms have the potential to cause outbreaks,[14, 57] and guidance exist around control of infection in those settings.[58, 59] The findings demonstrate however, that private farms are a significant driver in the burden of sporadic STEC infection and represent an uncontrolled risk. In Paper 1, non-O157 cases appeared to be more often travel associated than O157 cases. However in Paper 5, which included an additional subset of Non-O157 cases in an adjusted analysis demonstrated that non-O157 cases were at increased odds of contact with farm animals suggesting the domestic animal reservoir is a driver for infection. In both studies non-O157 case numbers were small and these finding should be interpreted with caution.

The sustained importance of person-to-person transmission, food-borne infection and zoonotic transmission in outbreaks over time was demonstrated in paper 2. However, there were apparent changes in settings and vehicles over time. Outbreaks occurring in prisons, hospitals and care-homes declined while outbreaks in schools and nurseries increased. This may suggest improved infection control measures in residential institutional settings, and historical examination of outbreaks of other pathogens would provide further insight into this. The apparent increase in outbreaks in child-care facilities may be reflective of increased child-care utilisation over the period, with more children accessing nurseries and therefore a larger population is at risk, although it was not possible to evaluate this. It is of concern that shifts in the settings of outbreaks have been towards those affecting children, the group most vulnerable to severe disease.

Outbreaks associated with unpasteurised dairy products, post pasteurisation contamination, or pasteurisation failures which were prominent up to the early 2000's were not reported after 2002. Following a large outbreak in Wales in 2005 due to cross contamination of meat supplied to schools which led to a public inquiry, meat related outbreaks were subsequently rarely detected. The decline in their occurrence suggests that efforts to reduce risk in catering, retail and meat hygiene sectors were effective.

Outbreak surveillance data are reliant on correct detection of outbreaks and outbreak cases. A proportion of outbreaks reported were attributed to an unknown source, but some may have represented a coincidental occurrence of unrelated cases which did not have a common source. Alternatively, epidemiological links between

related cases may have been diluted through including unrelated cases in investigations, hampering the identification of source. Individual outbreak investigations were not evaluated in this work but may have been subject to potential investigation biases towards known risk factors for STEC, which will have changed over time following high profile outbreaks. For example, the first reports of STEC in the USA linked the pathogen to hamburgers and it was nicknamed the 'burger bug'. In England, handling raw vegetables was the first implicated vehicle in an outbreak, and the vehicle linked to the biggest national outbreak, in 2011, [60, 61]yet vegetable sources were seldom reported in the intervening years.

Interpreting outbreak data over time is subject to some limitations, as prior to 1998, only outbreaks of HC or HUS would have been detected. In the first report on STEC outbreaks published on eFOSS data, only 8% of STEC O157 cases in 1992-1994 were attributed to outbreaks. [62] If specific subtypes are associated with different vehicles and can cause more severe disease, than a preponderance to detect outbreaks of more severe disease may have biased identification of certain sources of infection. This is plausible based on the description of niche strain type, such as PT8 and travel in Paper 1. Despite these caveats, Paper 2 indicates that the prominent vehicles in outbreaks reported previously (contaminated meat and milk) no longer apply to the same extent and efforts to reduce risk need to be focused elsewhere.

Impact and implications

The risk of STEC infection in the population is unevenly distributed; women, children and those living in rural areas are at a greater risk of infection through a likely combination of host factors and behaviours which expose them to infection. Paper 1 provided a foundation for further work, describing the distribution of disease and identifying areas for further exploration and has been cited by 18 articles. The finding that adult females are at greater risk is unexplained and warrants further investigation using robust, analytical studies accounting for other *a priori* risk factors. Expanding such studies to include a valid measure of ethnicity and socioeconomic variables to greater understand the uneven distribution of disease is needed. The studies demonstrate that living in rural environments and exposure to animals and their environment is associated with sporadic infection. However, both surveillance

of animal populations and exploration of the interaction of humans with these environments are needed if interventions are to be considered.

The analysis of outbreaks highlighted that microbiological risks in the food chain may change but persist to the extent that transmission of STEC over time is sustained. Since the study period an outbreak associated with raw drinking milk occurred in 2014,[63] and one associated with unpasteurised cheese in Scotland, causing a child death, occurred in 2016(unpublished data). Additionally, outbreaks associated with salad vegetables have increased in frequency and size, and fit the profile of common PTs which are nationally dispersed, which would have been undetected using the previous criterion.[64-66] These outbreaks highlight the continued, yet dynamic, microbiological risk from contaminated products and the importance of monitoring trends in consumption preferences which can impact upon foodborne infection risks. Efforts should focus on assessing microbiological risk in foods and proactively tackling biosecurity risks which can lead to contamination of ready to eat produce as well as increasing public awareness regarding risk.

Objective 2: Identifying host and pathogen factors which influence clinical outcome

Introduction

In outbreaks, progression to HUS has been reported to vary from 20% to 65% of cases [14-17]. Few countries monitor HUS at a population level, but Scotland, the US and Denmark report estimates of 6.1%, 6.3% and 7.9% respectively.(13,16,17) In England, the last reported data at a national level (1992-4) estimated that 12% of STEC cases developed HUS.[44] However, these data were from a period when specimens from cases of HC or HUS only were screened for STEC which will have likely accounted for the greater estimate than the other countries.[67]

The risk of progression to HUS following STEC infection is heterogeneous and children are at greatest risk with numerous studies reporting that children aged below five years most frequently develop HUS. [5-10, 12, 18-24] STEC strains encoding the Stx2 toxin, particularly the Stx2a subtype more often cause HUS than other strains [68-72].

The proportion of STEC cases which develop HUS overall and by sub-groups in England was unknown. The routine collection of clinical data reconciled with microbiological results for STEC cases through NESSS enabled me for the first time, to examine clinical severity, including development of HUS among STEC cases in Papers 1 and 5.

Methods

Clinical data collected from NESSS were described , including the frequency of cases being hospitalised, developing HC or HUS overall and by case demographic, STEC serogroup, STEC O157 phage type (PT), and Stx type. Comparisons were made between microbiological subtypes (PT and serogroup) and clinical presentation using Fisher's exact test to assess statistical significance ($P < 0{\cdot}05$). In Paper 5, the pathogenicity traits of a set of Non-O157 strains were described including subtyped s*tx* genes, and *eae* and *aggR* genes (additional pathogenicity factors). To assess for

associations with HUS, comparisons were made between groups and Fisher's exact test used to assess statistical significance. Mantel– Haenzel odds ratios (ORs) were calculated to compare the risk of developing HUS within different groups.

Results

In Paper 1, clinical data were available for 3,267 symptomatic confirmed STEC cases, as well as 92 probable cases (confirmed serologically but STEC was not isolated). Diarrhoeal symptoms were indicated for 92·8% (3,117) cases, including HC for 61·0% (2,050). Abdominal pain was frequently reported (79·2%, 2,662), while fever (32·5%, 1,092), nausea (46·6%, 1564) and vomiting (37·3%, 1,252) were reported less often. Hospitalisation was recorded for 34·3% (1,151) and development of HUS for 6·4% (215) of cases. Thirteen deaths were reported, although cause of death was unknown; Nine were female and all but one were adult cases with a mean age of 65 years (95% CI 52·7–77·3).

Disease severity differed by age, gender and strain characteristics. For most age groups, HC and hospitalisation were similarly reported for both sexes. Three quarters of HUS cases were children aged 0–14 years, with a mean age of 14·6 years (95% CI 11·7– 17·5). HUS was more frequently reported in females than males in both children and adults aged over 60. The highest proportion of STEC cases progressing to HUS was 15% in females aged 1–4 years.

The most commonly detected serogroup was O157, accounting for 3,558 (98·8%) of cases and just nineteen different non-O157 serogroups comprising 44 cases were detected. Thirty-six different phage types (PTs) were isolated from cases with STEC O157 infection. The most common overall were PT21/28 and PT8, which contributed between 54·1% and 69.2% of STEC O157 each year. Phage type 8 was more often isolated in cases who had travelled abroad whereas PT 21/28 cases appeared to be most often domestically acquired.

Among STEC O157 cases, vomiting was more frequent in cases of PT21/28 than PT8 (40·1% vs. 34·2%, P = 0·0078). Hospitalisation rates were higher (41·7% vs. 29·3%, P < 0·001), as was progression to HUS (8·5% vs. 1·1%, P < 0·001). PT2 was also associated with more severe disease than PT8, including vomiting (43·8% vs. 34·2%, P

= 0·015), hospitalisation (43·8%, P < 0·001), and progression to HUS (7·6%, P < 0·001). Clinical data were available for 42/44 non-O157 STEC cases and eighteen (42.8%) were reported to have developed HUS; Progression to HUS was therefore reported ten times more frequently in non-O157 cases than O157 cases (P < 0·001). Clinical presentation in non-O157 cases did not differ significantly by presence of eae although numbers were small. Isolates from all cases of confirmed STEC-HUS encoded *stx2*; 139 *stx2* only, and 23 *stx1* and *stx2*.

In Paper 5, 97 non-O157 isolates from 84 cases were analysed. The most common non-O157 STEC serogroup detected was O26 (23.2%). Forty-five (46.4%) isolates encoded *stx2* only, 28 (28.3%) carried *stx1* and *stx2*, and the remaining 24 (24.2%) *stx1* only. The most frequently detected *stx2* subtype was stx2a (64.0%). The *eae* gene was detected in 50 (51.5%) non-O157 STEC isolates. Six strains of STEC O104 had *aggR*, another pathogenicity factor, but it was not detected in any other STEC strains. Clinical data were available for 74 cases and 66 reported diarrhoea including HC in 35 cases (47.3%). A total of 28 (37.8%) of cases were hospitalised, similar to that reported in Paper 1. A quarter of cases (n=18) developed HUS. Children had the greatest odds of developing HUS (OR 4.4, p=0.005), but there was no difference by gender (OR 1.14, p=0.800). The development of HUS was significantly associated with STEC strains possessing *eae* (OR: 5.845, p=0.023) and/or *stx2a* (OR: 9.56, p=0.003).

In Paper 1, an additional 115 cases of HUS were identified in NESSS without evidence of STEC, including three who died: 87 cases had only submitted serum samples which were negative for antibodies to *E. coli* O157 and no specimen was received by GBRU for 13 cases. For 15 HUS cases stx-negative *E. coli* were isolated, including four strains possessing *eae*. A gender disparity was also observed in these HUS cases with 55·7% being female. Compared to STEC confirmed HUS cases, a significantly higher proportion of cases were adults aged over 60 years (20·0% vs. 7·9%, P = 0·007), with a mean age of 25·5 years (95% CI 20·5–30·5).

Discussion

In Paper 1, reporting of both microbiological and questionnaire data was complete for over 97% of STEC infections reported, providing a large, comprehensive dataset of STEC cases for analysis while the analysis of additional data on non-O157 strains (Paper 5), allowed further exploration of strain pathogenicity for these serotypes, rarely detected in England. The clinical burden of STEC infection was high with 61% of symptomatic cases experiencing HC, over a third of cases hospitalised and 6.4% cases reported as developing HUS. The greatest burden in terms of severity lay in children consistent with the literature. An excess of female cases under 15 and aged over 60 developed HUS. A number of previous studies have reported an excess of female HUS cases [9, 18, 19, 73-75], and in some a statistically significant association with female gender and developing STEC-HUS was found. [19, 73, 74]. However, an increased risk of HUS and female gender is unexplained, and may suggest host factors placing females at greater risk of developing HUS. In other studies no association between female gender and developing HUS has been demonstrated after adjusting for other factors, [7, 23, 75-77] and female gender did not correlate with HUS in Paper 5, although the study size was much smaller.

Among cases of STEC O157, severity of illness varied by PT. Vomiting, hospitalisation and progression to HUS were all significantly more frequently reported among cases of PT 21/28 and PT2, compared to PT8 and all other PT's. As described in Paper 2, PT and *Stx* are interrelated; most (93.2%) PT8 strains possess *stx1+2*, whereas PT2 and PT21/28 usually possessed *stx2* only. STEC O157 strains possessing *stx2* only were therefore associated with increased progression to HUS. Among cases with HUS, all of the strains isolated encoded either *stx2* only (85.8%) or *stx1* and *stx2* (14.2%). Cases of PT 21/28 and PT2 cases appear to be largely domestically acquired, suggesting the greatest burden of severe disease is due to domestic acquisition and is where interventions should be targeted.

The frequency of specific GI symptoms of STEC infection did not differ among non-O157 cases compared to O157 cases but hospitalisation rates and development of HUS were significantly higher in Non-O157 cases in Paper 1. This was expected as specimens are only sent for testing for Non-O157 STEC if there is clinical suspicion of

HUS, as per the national guidelines. When the pathogenicity of non-O157 strains was assessed in Paper 5 which included additional isolates detected by a GI PCR (Paper 5) development of HUS occurred less often, with no HUS cases infected with strains detected via the PCR. This study was based on just three laboratories undertaking PCR and was too small to draw conclusions on the overall public health impact of non-O157 STEC in England. However, developing HUS was significantly associated with *eae* positive and/or strains encoding *stx2a*, consistent with other studies. (74-76)

A further 115 cases of HUS, including three fatalities, were reported to NESSS without microbiological evidence of STEC Infection, representing 35% HUS cases captured by the system. Most (75.7%) had only submitted serum samples for testing which were negative for antibodies to *E.coli* O157. The gender composition of these HUS cases (55.7% females) was similar to all STEC cases. However, the mean age of these cases was significantly higher with a high proportion of adults aged over 60 years. This may be due to low clinical suspicion of STEC-HUS in the adult population or a different prodrome reducing the opportunity for sampling and testing of faecal specimens and warrants further investigation. If a gender disparity exists and is specific to adults, it may be that an under-ascertainment of adult HUS cases is obscuring the association between females and development HUS in some studies and further investigations are needed.

The data used are subject to some limitations. It is likely that HUS, hospitalisation and death were under-reported through NESSS; ESQs are completed to ensure that public health action is taken and represent the status of the patient at the time of completion. Half of cases were still ill when the questionnaire was administered, and were not followed up to determine subsequent outcome of their infection. The extent of the under-estimation of HUS through NESSS is unknown.

Impact and implications

Papers 1 and 5 have been referenced in the literature, with 20 citations each. They have provided a foundation for further work examining factors influencing disease severity; a subsequent multivariable analysis was undertaken on the data included in

Paper 1, for which I was second and corresponding author.[78] This study demonstrated that being aged 1-4 years, female, infected with PT21/28 and PT2 strains, and prescription of antibiotics were significant predictors for developing HUS. To explore further the role of *stx* subtypes I undertook a multivariable analysis to assess which STEC O157 lineages and *stx* subtypes are most associated with severe disease.[79]

Further monitoring of non-O157 STEC is needed to better ascertain their impact in England and identify specific strains of public health concern. It is also important to monitor STEC O157 subtypes over time as changes in circulating strains may impact on the overall clinical burden of disease. Ascertainment of HUS using current systems is limited and better data capture would enable more accurate estimates of the burden of disease, including further studies to examine gender and HUS and to estimate HUS in the adult population.

An HUS surveillance system or outcome surveillance as employed in a small number of countries would facilitate this. This could involve a bespoke system and/or linkage of NESSS with routinely collected data such as Hospital Episode Statistics. Such surveillance activities would be timely given the implementation of frontline PCR to detect non-O157 STEC and would allow better characterisation of the pathogenicity of non-O157 STEC strains, ultimately aiding with public health prioritisation of specific circulating STEC strains.

Objective 3: Evaluating the impact of enhanced detection and typing methods on our understanding of the epidemiology and public health impact of STEC

Introduction

This section considers the impact of newer technologies to subtype O157 STEC, to detect non-O157 STEC, and to subtype non-O157.

Differentiating strains through typing methods is important for outbreak detection and investigation.

In England, subtyping of STEC O157 includes differentiation by PT and Stx type. Since the early 2000's the majority (~60%) of STEC O157 isolates detected in England belong to one of two phage types (Papers 1 and 2). For Non-O157 isolates, Serogroup and Stx type only were determined and strains weren't further differentiated. The ability to discriminate between cases from the same or a separate source is therefore low, and reliant upon applying the person-place-time paradigm to detect outbreaks and inform case definitions. It is likely that outbreaks of common types go undetected, particularly when cases are dispersed geographically and/or in time. This under-ascertainment is supported by the data, with just two national outbreaks detected in the thirty-years, 1983-2012.

Between 2012 and 2014, molecular methods for detecting Non-O157 strains (GI PCR) were partly implemented, while MLVA was used to type isolates of STEC O157 and WGS was evaluated for STEC. The reconciliation of the resultant microbiological data sets with NESSS facilitated an evaluation of these methods and their impact upon our knowledge of clinical infection, as presented in Papers 3-6.

Methods

From 1st May 2012, all confirmed STEC O157 isolated in England were subject to MLVA typing and I undertook a study (Paper 3) on the first six months of data to retrospectively assess the impact of this approach on the detection of clusters of

disease, which could represent outbreaks. Cases were assigned as being part of a community cluster, household cluster or sporadic infection using MLVA and compared to the same cases categorised using existing methods for defining cases as related- a combination of monitoring PT and geographical and temporal links between cases.. The epidemiological data supporting a common link between cases in community clusters detected by MLVA were assessed.

In Paper 4, whole-genome sequencing (WGS) was assessed for its ability to identify linked cases. Isolates were randomly selected from a sample of STEC O157 strains processed by GBRU in 2012 and 2013, as well as historical isolates selected from 1990 and 2011 to represent the diversity of PTs seen in England and to provide population context. The SNP differences between isolates with a known epidemiological link were calculated to determine a threshold for defining related cases, termed the Common Source Threshold (CST).Clustering of cases over time within the CST were compared with clustering using an identical or SLV MLVA profile by survival analysis to compare the ability of the two methods at prospectively clustering cases together in time. Survival represented not clustering and failure represented clustering with another case. Both linking with a single other case and complete resolution of a cluster were assessed.

In Paper 5, I describe the impact of the introduction of a frontline PCR on detecting non-O157 STEC. WGS for typing of non-O157 STEC isolates to discriminating outbreaks from sporadic infection was used in Paper 6, the description of an outbreak of O26 alongside results from WGS and metagenomic analysis were included.

Results

In paper 3, isolates from 556 confirmed cases of STEC O157 were differentiated into 16 different PTs, most (62.8%) of which were PT21/28 or PT8 (n=349). Further typing by MLVA was undertaken on 539 isolates, and differentiated them into 341 unique (>2 locus variants) MLVA profiles. The concurrence between MLVA and PT was high with three instances where the same MLVA profile covered more than one PT.

Of the 341 unique MLVA profiles, 258 were uniquely observed for single cases only, representing sporadic infections. The other 83 were reported in clusters of >1 case:

42 due to known household transmission and 41 profiles shared amongst cases outside of the same household (community clusters). Of the 41 community clusters, 21 had already been detected and investigated through existing methods-recognition of shared epidemiological links and a common PT. Almost twice as many community clusters were detected therefore using MLVA as compared to standard practices to detect outbreaks. .This included re-categorisation of 101/267 apparently sporadic, domestic infections as part of community clusters, reducing the proportion of sporadic cases in the dataset from 69% to 43%. For the twenty clusters detected using MLVA only, most (n=18, 90%), of these clusters were of the two most common PT's, PT21/28 and PT8. The cluster sizes were small with a median of three cases. Additional evidence of a link between cases was found in 12/20 clusters, four with shared exposures identified on ESQs and eight comprising cases occurring contemporaneously. Eight clusters comprised cases with no additional linkage beyond MLVA and were dispersed geographically and in time, with cases spanning several months.

In Paper 4, 572 isolates from known outbreaks, linked household cases, multiple isolates from the same patient, and apparently sporadic cases were sequenced. All 183 isolates with a known epidemiological link to another isolate had <5 SNPs difference from their epidemiologically linked isolates. The mean difference was one SNP between isolates from the same household, a known outbreak or isolates from the same case. This led to the hypothesis that a threshold of 5 SNPs could be used to detect outbreaks of STEC O157 and was termed the CST. This threshold was applied to the dataset and identified 136 additional cases with no prior known epidemiological link, which fell within 5 SNPs of another case. The majority (87%) of these pairs were strains which had been isolated within 30 days of each other with a mean of 11 days. Twenty of 33 WGS clusters were not previously detected, but were those identified in Paper 3 using MLVA. WGS did not link any additional cases which MLVA hadn't and both methods were equally as sensitive.

Clustering isolates over time using the CST was compared to MLVA (an identical or SLV MLVA profile). There was no significant difference based on clustering a single isolate with another (log-rank test for equality of survivor function: $P = .101$; Cox hazard ratio = 0.89, $P = 0.198$), indicating no difference in timeliness of clustering

cases between the two methods. However, when time to completion of the cluster was considered (i.e. all cases in the cluster are identified), WGS was more timely than MLVA (Cox HR=1.44, p=0.001). This is because MLVA typing can rely on an intermediary isolate appearing in the dataset to link two isolates which may themselves be two or more loci apart.

In Paper 5, three frontline hospital laboratories implemented the GI PCR approach during the study period resulting in a significant increase in the detection of non-O157 STEC; In 2013, 42 (42.4 %) non-O157 STEC isolates were detected compared with a total of 57 in the preceding 4 years (2009–2012), (P<0.001). In total, 22/97 (22.7%) cases were detected by the three frontline laboratories using GI PCR. One hospital laboratory in London, the region with the lowest STEC incidence in England, reported 19 confirmed non-O157 cases during the first 12 months of adopting the GI PCR approach, while only five STEC O157 isolates were reported over that same period. This laboratory had previously reported three non-O157 cases between 2004 and 2012.

In Paper 6, 19 strains of *E. coli* O26:H11 were retrospectively sequenced including isolates from an outbreak to evaluate its utility in differentiating strains of this serogroup. WGS indicated that the nursery outbreak strains fell within 0-3 SNPs of each other. The outbreak included a case for whom faecal specimens were PCR positive for *stx2* and *eae* but STEC O26:H11 was not cultured. Metagenomics undertaken on the faecal specimen from that case and provided molecular evidence that the strain she was infected with was related to the nursery outbreak. The phylogenetic relationship of the 19 sequenced strains indicated two sequence types using MLST, one denoting *stx* positive strains and the other *stx* negative strains. Of the latter, 2/3 of the cases had severe disease and one developed HUS.

Discussion

In Paper 3, MLVA provided a more discriminatory typing tool than phage typing, identifying almost twice as many clusters than by PT, redefining the relative burden of sporadic infection and that to a shared source. The clusters only identified through MLVA were more geographically and temporally dispersed and most belonged to the most common PTs. It is in unsurprising they went undetected therefore and supports

the supposition that national outbreaks occur undetected in the absence of fine typing techniques, including MLVA and WGS. These more discriminatory typing schemes enable a more specific definition of the population of sporadic cases and analysis of exposures using these case definitions would provide a more accurate characterisation of risk factors for sporadic infection and better inform risk assessment and targeted public health intervention.

Retrospectively applying the person-place-time paradigm provided additional support to link the cluster cases in some of the MLVA identified clusters, but there were still those where no additional available evidence supported a link to a common source. The same clusters were identified using WGS in Paper 4, providing reassurance that MLVA was as specific as WGS for this dataset and that these weren't falsely detected clusters. The ESQ may not capture the detailed information often required to definitively link cases to a shared exposure. In Papers 3 & 4, the clusters were retrospectively reviewed and prospective investigation may have helped identify a common source through further interviewing. However, as most clusters were small the chance of identifying a source is reduced and analytical studies are not possible. Alternatively, clusters dispersed in time could reflect repeat sampling of the reservoir of STEC infection over time- indicative of attribution rather than outbreaks *per se*.

The survival analysis in Paper 4 indicated that there was no difference in clustering one case with another related case by method, but with WGS the cluster completes quicker. This is because all linked cases fall within the CST for all cases, whereas with MLVA several isolates will only be joined via an intermediate isolate (i.e. Double Locus Variants (DLVs) shared by a Single Locus Variant (SLV)) as shown graphically in paper 3. This means prospectively defining microbiological case definitions for outbreak investigations from the start is problematic using MLVA and what may appear as separate clusters or sporadic cases may later resolve as a single cluster if intermediaries are detected.

A further consideration in interpreting the results of Paper 3 is a later analysis comparing the number of locus variants against the CST determined that 42% of strains related by a DLV using MLVA did not fall within the CST (unpublished data,

2014). An identical or SLV MLVA profile was a proxy for the CST in absence of additional epidemiological data, in 99.1% of instances. It was due to these results that in Paper4, the survival analysis comparing clustering dynamics for WGS and MLVA included isolates within an SLV only. However, there was concurrence between the clusters detected and the cases comprising them between the two analyses in Papers 3 and 4 despite the different MLVA loci thresholds used. Paper 4 provided evidence that a threshold of within an SLV for clustering cases provides the same sensitivity as the WGS CST, giving confidence to laboratories not ready to adopt WGS technologies and that compared to PT, MLVA provides considerable advantages as a typing tool.

Evaluation of PCR (Paper 5) showed that almost four times as many Non-O157 isolates were detected via PCR than through culture and these were from laboratories in regions with the lowest STEC incidence in England. The ability for rapid detection of Non-O157 STEC is important for public health; In Paper 6, the first case, and subsequent outbreak, were detected through the referral of a stool specimen for further investigation to GBRU. Direct PCR on the stool identified *stx* genes and STEC O26 was subsequently isolated. Two weeks elapsed between onset of symptoms to reporting of STEC. If the PCR had been available at the frontline laboratory than diagnosis and management of STEC could have been more rapid, offering clear benefits in preventing onwards transmission in outbreaks.

However, this would require taking action based upon a PCR positive result alone which raises issues in itself; the increased sensitivity of the PCR leads to PCR-positive but culture-negative results which are difficult to interpret with respect to risk of transmission and clinical significance. STEC O26, the causative strain in the nursery outbreak, as well as serogroups O103, O111 and O145, were already associated with more severe disease and known to cause HUS. However, the potential resource impact for public health follow-up and investigation of cases following further expansion of the GI PCR more widely is clearly large and various other strains were found which may or may not be of significant public health concern.

The increase in detection of non-O157 strains highlighted a requirement to differentiate cases of non-O157 STEC. As described in papers 1-4, PT, MLVA and WGS

have been specifically used to type STEC O157 strains. Paper 6 trialled the retrospective use of WGS for typing Non-O157 isolates for an outbreak for the first time in England and demonstrated that WGS provided highly discriminatory confirmation of linkage at the molecular level. While the serotyping and *stx* typing alongside the epidemiological links were strong enough evidence at the time to regard the cases as linked, we demonstrated for the first time that WGS can provide rapid and robust confirmation of relationships between STEC O26 strains.

Impact and implications of the work

The increased detection of clusters of STEC has resource implications for local and national public health teams. Alongside further prospective evaluation, the analyses in Papers 3 & 4, were used to inform an algorithm developed and implemented by PHE, for responding to clusters. This involves using a threshold of detecting five or more linked cases before initiating further investigations beyond review of the ESQs. This is to balance the additional demands of detecting many more clusters and the likelihood of identifying a common source. This was implemented when routine MLVA was in use in 2014 and continued when MLVA was superseded by WGS in 2015. Adapted approaches have been adopted within PHE's GI department to respond to clusters of other GI pathogens. They are also of interest to international colleagues, and Papers 3 and 4 have been cited by 13 and 48 publications respectively. Paper 4 provided the forerunner to further investigating the interaction between phylogenetic and epidemiological relationships for STEC infections overall and in informing outbreaks, which I have continued to contribute to in my work.[65, 80]

The literature on the significance of Non-O157 STEC is relatively sparse and the numbers reported are often small. With 20 and 13 citations respectively, Papers 5 and 6 have contributed to this field. Building on this, I have also undertaken further work examining the role of the *stx* and *eae* genes as predictors of severity rather than serotype and guidance for a differential public health response by PHE to Non-O157 STEC based on host factors and the pathogenicity profile of the strains is currently under development.[79]

Conclusions

Together the papers included meet the aim of the research and provided a detailed update on the epidemiology of STEC infection in England, including an assessment of the impact of molecular methods (PCR, MLVA and WGS), on our understanding and control of STEC. In achieving this, I used the systematic collection of data by GBRU over three decades, augmented with the introduction of NESSS in 2009, outbreak surveillance since 1992 and the recent introduction of advanced detection and typing techniques in England to comprehensively describe the past and current epidemiology of STEC in England, evaluate clinical impact and assess the impact of the recent introduction of advanced detection and typing techniques in England. The body of work as a whole provided a future perspective on where research is needed as well as operational implications for detecting and managing sporadic cases and clusters of infection.

The epidemiology of STEC infections in England was described using current surveillance methods and provided an update on the incidence of STEC, burden, risk factors and trends over time. The risk of STEC infection in the population was unevenly distributed; women, children and those living in rural areas were at greater risk of infection. The analysis of outbreaks highlighted that microbiological risks in the food chain may change, and despite several interventions over the decades, persist to the extent that transmission of STEC over time was sustained. With regards to disease severity, an increased risk of HUS amongst females was highlighted while the greatest burden of severe disease was found to be due to domestic acquisition.

These studies were undertaken on routinely collected surveillance data. Analysis of such data provides a cost-effective and rapid solution to address research questions. The collection of standardised information on STEC through a national system, unique to the UK, provided a large and comprehensive dataset for analyses. The increased reliability of detailed, standardised enhanced data provided a more accurate reflection of cases attributable to outbreaks, secondary transmission and travel compared to laboratory surveillance alone. However, surveillance is subject to a number of ascertainment biases which could impact upon findings and their interpretation; cases detected are likely to reflect those with more severe disease which limits the generalisability of some results, and doesn't capture the extent of mild or asymptomatic infection. While it was possible to examine risk exposures

amongst cases, and undertake a case-case analysis, examining risk factors for STEC infection overall was not possible due to the absence of a control population. A case-control study would be a robust method to assess risks for STEC infection.

The assessments of disease severity made using surveillance data were limited in that clinical definitions were not used and the timeliness of patient interviews meant that clinical outcome could not be reliably assessed. Indeed, since these papers, I have undertaken a paediatric cohort study of HUS which demonstrated that HUS is under-ascertained in NESSS and development of HUS is actually reported in as high as 25% in some groups (manuscript in preparation). The publications did however provide a foundation for further work examining factors influencing disease severity; a subsequent multivariable analysis was undertaken on the data included in Paper 1, for which I was second and corresponding author.[78] This study demonstrated that being aged 1-4 years, female, infected with PT21/28 and PT2 strains, and prescription of antibiotics were significant predictors for developing HUS. I also collaborated with a PhD student who has undertaken a study using NESSS and the paediatric cohort data to examine risk factors for developing HUS, including markers of socioeconomic status (manuscript under submission).

To explore further the role of *stx* subtypes I undertook a multivariable analysis to assess which STEC O157 lineages and *stx* subtypes are most associated with severe disease. This along with Paper X have been used as an evidence base to inform an algorithm and national guidelines on public health follow up of cases of Non-O157 STEC, due to be published in 2018. In 2018, I have obtained access to Hospital Episodes Statistics (HES) data to link the data with NESSS to better understand the clinical burden of STEC infections by host and pathogen factors. This will be particularly important for gaining knowledge on the impact of Non-O157 strains and evaluating the public health guidelines which recommend reduced follow-up compared to O157.

.

Molecular techniques, which can rectify some of the limitations to surveillance (i.e. capturing Non-O157 data), were evaluated in this work .In summary together, Papers 3-6 demonstrate the utility of new molecular methods, including PCR, MLVA and WGS at enhancing STEC surveillance activities. PCR rapidly detects cases of STEC

Non-O157 which previously went undetected. The public health importance of these strains was difficult to ascertain due to limited size of the datasets but the dataset now contains over 1200 cases of Non O157 STEC which I will be analysing in 2018. In parallel, MLVA and subsequently WGS have provided both more sensitive and specific detection of outbreaks and case definitions and redefined previous thinking on the relative burden of disease attributed to sporadic infection and outbreaks. However, these enhanced methods present challenges in terms of interpretation and workload.. This work has informed an algorithm developed and implemented by PHE, for responding to WGS clusters of GI disease. This involves using a threshold of detecting five or more linked cases before initiating further investigations beyond review of the ESQs.

Paper 4 provided the forerunner to further investigating the interaction between phylogenetic and epidemiological relationships for STEC infections overall and in informing outbreaks, which I also continue to contribute to in my work at PHE as the lead surveillance scientist for STEC.

References

(1) **Konowalchuk J, Speirs JI, Stavric S.** Vero response to a cytotoxin of Escherichia coli. *Infect Immun* 1977; **18**(3): 775-779.

(2) **Frank C, et al.** Epidemic profile of Shiga-toxin-producing Escherichia coli O104:H4 outbreak in Germany. *N Engl J Med* 2011; **365**(19): 1771-1780.

(3) **Health Protection Surveillance Centre**. Epidemiology of Verotoxigenic E. coli in Ireland, 2015: Health and Safety Executive; 2016.

(4) **Cimolai N, Morrison BJ, Carter JE.** Risk factors for the central nervous system manifestations of gastroenteritis-associated hemolytic-uremic syndrome. *Pediatrics* 1992; **90**(4): 616-621.

(5) **Gerber A, et al.** Clinical course and the role of shiga toxin-producing Escherichia coli infection in the hemolytic-uremic syndrome in pediatric patients, 1997-2000, in Germany and Austria: a prospective study. *J Infect Dis* 2002; **186**(4): 493-500.

(6) **Decludt B, et al.** Haemolytic uraemic syndrome and Shiga toxin-producing Escherichia coli infection in children in France. The Societe de Nephrologie Pediatrique. *Epidemiol Infect* 2000; **124**(2): 215-220.

(7) **Dundas S, et al.** The central Scotland Escherichia coli O157:H7 outbreak: risk factors for the hemolytic uremic syndrome and death among hospitalized patients. *Clin Infect Dis* 2001; **33**(7): 923-931.

(8) **Elliott EJ, et al.** Nationwide study of haemolytic uraemic syndrome: clinical, microbiological, and epidemiological features. *Arch Dis Child* 2001; **85**(2): 125-131.

(9) **Micheletti MV, et al.** Hemolytic uremic syndrome: epidemiological and clinical features of a pediatric population in Tuscany. *Kidney Blood Press Res* 2010; **33**(5): 399-404.

(10) **Lynn RM, et al.** Childhood hemolytic uremic syndrome, United Kingdom and Ireland. *Emerg Infect Dis* 2005; **11**(4): 590-596.

(11) **Nathanson S, et al.** Acute neurological involvement in diarrhea-associated hemolytic uremic syndrome. *Clin J Am Soc Nephrol* 2010; **5**(7): 1218-1228.

(12) **Schifferli A, et al.** Hemolytic-uremic syndrome in Switzerland: a nationwide surveillance 1997-2003. *Eur J Pediatr* 2010; **169**(5): 591-598.

(13) **Adak GK, Long SM, O'Brien SJ.** Trends in indigenous foodborne disease and deaths, England and Wales: 1992 to 2000. *Gut* 2002; **51**(6): 832-841.

(14) **Ihekweazu C, et al.** Large outbreak of verocytotoxin-producing Escherichia coli O157 infection in visitors to a petting farm in South East England, 2009. *Epidemiol Infect* 2012; **140**(8): 1400-1413.

(15) **Frank C, Milde-Busch A, Werber D.** Results of surveillance for infections with Shiga toxin-producing Escherichia coli (STEC) of serotype O104:H4 after the large outbreak in Germany, July to December 2011. *Euro Surveill* 2014; **19**(14).

(16) **Yahata Y, et al.** Epidemiological analysis of a large enterohaemorrhagic Escherichia coli O111 outbreak in Japan associated with haemolytic uraemic syndrome and acute encephalopathy. *Epidemiol Infect* 2015; **143**(13): 2721-2732.

(17) **Soborg B, et al.** A verocytotoxin-producing E. coli outbreak with a surprisingly high risk of haemolytic uraemic syndrome, Denmark, September-October 2012. *Euro Surveill* 2013; **18**(2).

(18) **Proulx F, Sockett P.** Prospective surveillance of Canadian children with the haemolytic uraemic syndrome. *Pediatr Nephrol* 2005; **20**(6): 786-790.
(19) **Gould LH, et al.** Hemolytic uremic syndrome and death in persons with Escherichia coli O157:H7 infection, foodborne diseases active surveillance network sites, 2000-2006. *Clin Infect Dis* 2009; **49**(10): 1480-1485.
(20) **Griffin PM, et al.** Illnesses associated with Escherichia coli O157:H7 infections. A broad clinical spectrum. *Ann Intern Med* 1988; **109**(9): 705-712.
(21) **Pavia AT, et al.** Hemolytic-uremic syndrome during an outbreak of Escherichia coli O157:H7 infections in institutions for mentally retarded persons: clinical and epidemiologic observations. *J Pediatr* 1990; **116**(4): 544-551.
(22) **Cimolai N, et al.** Risk factors for the progression of Escherichia coli O157:H7 enteritis to hemolytic-uremic syndrome. *J Pediatr* 1990; **116**(4): 589-592.
(23) **Tserenpuntsag B, et al.** Hemolytic uremic syndrome risk and Escherichia coli O157:H7. *Emerg Infect Dis* 2005; **11**(12): 1955-1957.
(24) **Jenssen GR, et al.** The incidence and aetiology of acute kidney injury in children in Norway between 1999 and 2008. *Acta Paediatr* 2014; **103**(11): 1192-1197.
(25) **Chapman PA, Wright DJ, Norman P.** Verotoxin-producing Escherichia coli infections in Sheffield: cattle as a possible source. *Epidemiol Infect* 1989; **102**(3): 439-445.
(26) **Armstrong GL, Hollingsworth J, Morris JG, Jr.** Emerging foodborne pathogens: Escherichia coli O157:H7 as a model of entry of a new pathogen into the food supply of the developed world. *Epidemiol Rev* 1996; **18**(1): 29-51.
(27) **Ferens WA, Hovde CJ.** Escherichia coli O157:H7: animal reservoir and sources of human infection. *Foodborne Pathog Dis* 2011; **8**(4): 465-487.
(28) **Haas CN, et al.** Development of a dose-response relationship for Escherichia coli O157:H7. *Int J Food Microbiol* 2000; **56**(2-3): 153-159.
(29) **Tilden J, Jr., et al.** A new route of transmission for Escherichia coli: infection from dry fermented salami. *Am J Public Health* 1996; **86**(8): 1142-1145.
(30) **Paton AW, et al.** Molecular microbiological investigation of an outbreak of hemolytic-uremic syndrome caused by dry fermented sausage contaminated with Shiga-like toxin-producing Escherichia coli. *J Clin Microbiol* 1996; **34**(7): 1622-1627.
(31) **Teunis P, Takumi K, Shinagawa K.** Dose response for infection by Escherichia coli O157:H7 from outbreak data. *Risk Anal* 2004; **24**(2): 401-407.
(32) **O'Brien SJ, Adak GK, Gilham C.** Contact with farming environment as a major risk factor for Shiga toxin (Vero cytotoxin)-producing Escherichia coli O157 infection in humans. *Emerg Infect Dis* 2001; **7**(6): 1049-1051.
(33) **Parry SM, et al.** Risk factors for and prevention of sporadic infections with vero cytotoxin (shiga toxin) producing Escherichia coli O157. *Lancet* 1998; **351**(9108): 1019-1022.
(34) **Locking ME, et al.** Risk factors for sporadic cases of Escherichia coli O157 infection: the importance of contact with animal excreta. *Epidemiol Infect* 2001; **127**(2): 215-220.
(35) **Willshaw GA, et al.** Verocytotoxin-producing Escherichia coli (VTEC) O157 and other VTEC from human infections in England and Wales: 1995-1998. *J Med Microbiol* 2001; **50**(2): 135-142.
(36) **Gormley FJ, et al.** A 17-year review of foodborne outbreaks: describing the continuing decline in England and Wales (1992-2008). *Epidemiol Infect* 2011; **139**(5): 688-699.
(37) **Health Protection Agency.** The VTEC Operational manual: Operational guidance for HPA staff dealing with cases and incidentsof VTEC infection. 2010.

(38) **Gillespie IA, et al.** Foodborne general outbreaks of Shiga toxin-producing Escherichia coli O157 in England and Wales 1992-2002: where are the risks? *Epidemiol Infect* 2005; **133**(5): 803-808.
(39) **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**(1): 69-77.
(40) **Wheeler JG, et al.** Study of infectious intestinal disease in England: rates in the community, presenting to general practice, and reported to national surveillance. The Infectious Intestinal Disease Study Executive. *BMJ* 1999; **318**(7190): 1046-1050.
(41) **Tam CC, Rodrigues LC, O'Brien SJ.** The study of infectious intestinal disease in England: what risk factors for presentation to general practice tell us about potential for selection bias in case-control studies of reported cases of diarrhoea. *Int J Epidemiol* 2003; **32**(1): 99-105.
(42) **Vally H, et al.** Epidemiology of Shiga toxin producing Escherichia coli in Australia, 2000-2010. *BMC Public Health* 2012; **12**: 63.
(43) **Carroll AM, Gibson A, McNamara EB.** Laboratory-based surveillance of human verocytotoxigenic Escherichia coli infection in the Republic of Ireland, 2002-2004. *J Med Microbiol* 2005; **54**(Pt 12): 1163-1169.
(44) **Thomas A, et al.** Vero cytotoxin-producing Escherichia coli, particularly serogroup O157, associated with human infections in England and Wales: 1992-4. *Epidemiol Infect* 1996; **117**(1): 1-10.
(45) **Tuttle J, et al.** Lessons from a large outbreak of Escherichia coli O157:H7 infections: insights into the infectious dose and method of widespread contamination of hamburger patties. *Epidemiol Infect* 1999; **122**(2): 185-192.
(46) **Rangel JM, et al.** Epidemiology of Escherichia coli O157:H7 outbreaks, United States, 1982-2002. *Emerg Infect Dis* 2005; **11**(4): 603-609.
(47) **Wahl E, et al.** Investigation of an Escherichia coli O145 outbreak in a child day-care centre--extensive sampling and characterization of eae- and stx1-positive E. coli yields epidemiological and socioeconomic insight. *BMC Infect Dis* 2011; **11**: 238.
(48) **Snedeker KG, et al.** Primary and secondary cases in Escherichia coli O157 outbreaks: a statistical analysis. *BMC Infect Dis* 2009; **9**: 144.
(49) **Garvey P, et al.** Verotoxigenic Escherichia coli transmission in Ireland: a review of notified outbreaks, 2004-2012. *Epidemiol Infect* 2016; **144**(5): 917-926.
(50) **MacDonald E, et al.** Implications of screening and childcare exclusion policies for children with Shiga-toxin producing Escherichia coli infections: lessons learned from an outbreak in a daycare centre, Norway, 2012. *BMC Infect Dis* 2014; **14**: 673.
(51) **Fujii J, et al.** Risk of haemolytic uraemic syndrome caused by shiga-toxin-producing Escherichia coli infection in adult women in Japan. *Epidemiol Infect* 2016; **144**(5): 952-961.
(52) **Strachan NJ, et al.** Escherichia coli O157: burger bug or environmental pathogen? *Int J Food Microbiol* 2006; **112**(2): 129-137.
(53) **Innocent GT, et al.** Spatial and temporal epidemiology of sporadic human cases of Escherichia coli O157 in Scotland, 1996-1999. *Epidemiol Infect* 2005; **133**(6): 1033-1041.
(54) **Locking M, et al.** VTEC infections and livestock-related exposures in Scotland, 2004. *Euro Surveill* 2006; **11**(2): E060223 060224.
(55) **Pearce MC, et al.** Temporal and spatial patterns of bovine Escherichia coli O157 prevalence and comparison of temporal changes in the patterns of phage types associated with bovine shedding and human E. coli O157 cases in Scotland between 1998-2000 and 2002-2004. *BMC Microbiol* 2009; **9**: 276.

(56) **Money P, et al.** Cattle, weather and water: mapping Escherichia coli O157:H7 infections in humans in England and Scotland. *Environ Microbiol* 2010; **12**(10): 2633-2644.
(57) **Willshaw GA, et al.** Verocytotoxin-producing Escherichia coli infection and private farm visits. *Vet Rec* 2003; **152**(12): 365-366.
(58) Public Health England. Avoiding Infection on farm visits:advice for the public. In, 2014.
(59) Health and Safety Executive. Preventing or controlling ill health from animal contact at visitor attractions - Guidance on inspection and enforcement. In, 2011.
(60) **Launders N, et al.** A large Great Britain-wide outbreak of STEC O157 phage type 8 linked to handling of raw leeks and potatoes. *Epidemiol Infect* 2016; **144**(1): 171-181.
(61) **Perry N, et al.** Application of multilocus variable number tandem repeat analysis to monitor Verocytotoxin-producing Escherichia coli O157 phage type 8 in England and Wales: emergence of a profile associated with a national outbreak. *J Appl Microbiol* 2013; **115**(4): 1052-1058.
(62) **Wall PG, et al.** General outbreaks of vero cytotoxin producing Escherichia coli O157 in England and Wales from 1992 to 1994. *Commun Dis Rep CDR Rev* 1996; **6**(2): R26-33.
(63) **Butcher H, et al.** Whole genome sequencing improved case ascertainment in an outbreak of Shiga toxin-producing Escherichia coli O157 associated with raw drinking milk. *Epidemiol Infect* 2016; **144**(13): 2812-2823.
(64) **Sinclair C, et al.** Investigation of a national outbreak of STEC Escherichia coli O157 using online consumer panel control methods: Great Britain, October 2014. *Epidemiol Infect* 2017; **145**(5): 864-871.
(65) **Jenkins C, et al.** Public Health Investigation of Two Outbreaks of Shiga Toxin-Producing Escherichia coli O157 Associated with Consumption of Watercress. *Appl Environ Microbiol* 2015; **81**(12): 3946-3952.
(66) **Launders N, et al.** Outbreak of Shiga toxin-producing E. coli O157 associated with consumption of watercress, United Kingdom, August to September 2013. *Euro Surveill* 2013; **18**(44).
(67) **Anon.** Test all diarrhoeal stools for VTEC O157; sensitive methods are available if suspicion is high and yield likely to be low. *Commun Dis Rep CDR Wkly* 1999; **9**(46): 407, 410.
(68) **Ethelberg S, et al.** Virulence factors for hemolytic uremic syndrome, Denmark. *Emerg Infect Dis* 2004; **10**(5): 842-847.
(69) **Persson S, et al.** Subtyping method for Escherichia coli shiga toxin (verocytotoxin) 2 variants and correlations to clinical manifestations. *J Clin Microbiol* 2007; **45**(6): 2020-2024.
(70) **Luna-Gierke RE, et al.** Outbreaks of non-O157 Shiga toxin-producing Escherichia coli infection: USA. *Epidemiol Infect* 2014; **142**(11): 2270-2280.
(71) **Marejkova M, et al.** Enterohemorrhagic Escherichia coli as causes of hemolytic uremic syndrome in the Czech Republic. *PLoS One* 2013; **8**(9): e73927.
(72) **Brooks JT, et al.** Non-O157 Shiga toxin-producing Escherichia coli infections in the United States, 1983-2002. *J Infect Dis* 2005; **192**(8): 1422-1429.
(73) **Rowe PC, et al.** Haemolytic anaemia after childhood Escherichia coli O 157 .H7 infection: are females at increased risk? *Epidemiol Infect* 1991; **106**(3): 523-530.
(74) **Smith KE, et al.** Antibiotic treatment of Escherichia coli O157 infection and the risk of hemolytic uremic syndrome, Minnesota. *Pediatr Infect Dis J* 2012; **31**(1): 37-41.

(75) **Ikeda K, et al.** Predictors for the development of haemolytic uraemic syndrome with Escherichia coli O157:H7 infections: with focus on the day of illness. *Epidemiol Infect* 2000; **124**(3): 343-349.
(76) **Wong CS, et al.** Risk factors for the hemolytic uremic syndrome in children infected with Escherichia coli O157:H7: a multivariable analysis. *Clin Infect Dis* 2012; **55**(1): 33-41.
(77) **Bell BP, et al.** Predictors of hemolytic uremic syndrome in children during a large outbreak of Escherichia coli O157:H7 infections. *Pediatrics* 1997; **100**(1): E12.
(78) **Launders N, et al.** Disease severity of Shiga toxin-producing E. coli O157 and factors influencing the development of typical haemolytic uraemic syndrome: a retrospective cohort study, 2009-2012. *BMJ Open* 2016; **6**(1): e009933.
(79) **Dallman TJ, et al.** Applying phylogenomics to understand the emergence of Shiga-toxin-producing Escherichia coli O157:H7 strains causing severe human disease in the UK. *Microb Genom* 2015; **1**(3): e000029.
(80) **Byrne L, et al.** Epidemiological and Microbiological Investigation of an Outbreak of Severe Disease from Shiga Toxin-Producing Escherichia coli O157 Infection Associated with Consumption of a Slaw Garnish. *J Food Prot* 2016; **79**(7): 1161-1168.

APPENDIX I: The published work

Paper 1: The Epidemiology, microbiology and clinical impact of Shiga toxin-producing *Escherichia coli* in England, 2009-2012

Epidemiol. Infect., Page 1 of 13. © Cambridge University Press 2015
doi:10.1017/S0950268815000746

The epidemiology, microbiology and clinical impact of Shiga toxin-producing *Escherichia coli* in England, 2009–2012

L. BYRNE[1]*, C. JENKINS[2], N. LAUNDERS[1], R. ELSON[1] AND G. K. ADAK[1]

[1] *Public Health England Department of Gastrointestinal, Emerging and Zoonotic Infections, Centre for Infectious Disease Surveillance and Control, London, UK*
[2] *Public Health England Gastrointestinal Bacteria Reference Unit, London, UK*

Received 18 December 2014; Final revision 24 February 2015; Accepted 23 March 2015

SUMMARY

Between 1 January 2009 and 31 December 2012 in England, a total of 3717 cases were reported with evidence of Shiga toxin-producing *E. coli* (STEC) infection, and the crude incidence of STEC infection was 1·80/100 000 person-years. Incidence was highest in children aged 1–4 years (7·63/100 000 person-years). Females had a higher incidence of STEC than males [rate ratio (RR) 1·24, $P < 0·001$], and white ethnic groups had a higher incidence than non-white ethnic groups (RR 1·43, $P < 0·001$). Progression to haemolytic uraemic syndrome (HUS) was more frequent in females and children. Non-O157 STEC strains were associated with higher hospitalization and HUS rates than O157 STEC strains. In STEC O157 cases, phage type (PT) 21/28, predominantly indigenously acquired, was also associated with more severe disease than other PTs, as were strains encoding *stx2* genes. Incidence of STEC was over four times higher in people residing in rural areas than urban areas (RR 4·39, $P < 0·001$). Exposure to livestock and/or their faeces was reported twice as often in cases living in rural areas than urban areas ($P < 0·001$). Environmental/animal contact remains an important risk factor for STEC transmission and is a significant driver in the burden of sporadic STEC infection. The most commonly detected STEC serogroup in England was O157. However, a bias in testing methods results in an unquantifiable under-ascertainment of non-O157 STEC infections. Implementation of PCR-based diagnostic methods designed to detect all STEC, to address this diagnostic deficit, is therefore important.

Key words: *Escherichia coli*, foodborne zoonoses, gastrointestinal infections, infectious disease epidemiology, zoonotic foodborne diseases.

INTRODUCTION

Shiga toxin-producing *Escherichia coli* (STEC) are a group of bacteria associated with human disease and are defined by the presence of one or both phage-encoded Shiga toxin genes; *stx1* and *stx2*. While in England around 900 cases of STEC are reported annually compared to around 10 000 *Salmonella* and 60 000 *Campylobacter* cases, STEC are of significant public health concern due to the severity of disease. Symptoms can range from mild gastroenteritis through severe bloody diarrhoea, to haemolytic uraemic syndrome (HUS). HUS is most commonly seen in children aged <5 years, is recognized as the most common cause of acute kidney failure in children in the UK and can be fatal, particularly in infants, young children and the elderly [1].

* Author for correspondence: L. Byrne, Public Health England Department of Gastrointestinal, Emerging and Zoonotic Infections, Centre for Infectious Disease Surveillance and Control, 61 Colindale Avenue, London NW9 5EQ, UK.
(Email: lisa.byrne@phe.gov.uk)

Healthy cattle are the main reservoir of STEC although they are also carried by sheep and other animals [2, 3]. The infectious dose for human infection has been estimated to be <100 bacteria [4]. Transmission to humans occurs through either consumption of contaminated food or water, or exposure to a contaminated environment involving direct or indirect contact with animals or their faeces [2, 3]. The low infectious dose of STEC means that once in the population person-to-person spread is common [4–6]. STEC cause both sporadic and epidemic infections. While small outbreaks due to person-to-person spread are reported in closed settings, particularly childcare facilities [7, 8], large outbreaks are often associated with foodborne transmission [9, 10], and contact with ruminants, such as in open farms [11, 12].

The O157 STEC serogroup is most commonly associated with human disease in the UK. The majority of STEC O157 do not ferment sorbitol. Frontline diagnostic laboratories in England use standard methods, involving cefixine tellurite sorbitol MacConkey (CT-SMAC) agar [13, 14], to exploit this characteristic to preferentially detect STEC O157 from other faecal *E. coli*.

There are no simple, generally applicable culture-based tests for the detection of STEC other than serogroup O157 (non-O157 STEC) in faecal specimens, because of phenotypic diversity. However, over 400 serogroups of *E. coli* have been shown to produce Shiga toxins [15], and many non-O157 STEC have been associated with outbreaks [16–20]. The development of HUS has been shown to be associated with more than 100 serotypes of STEC [15]. In summer 2011, the largest outbreak of HUS ever recorded occurred in Germany and was found to be caused by an emergent strain of STEC, serotype O104:H4, not previously considered a significant pathotype to humans [18, 21–23].

While national laboratory report surveillance of STEC has been conducted in England for over 30 years [24, 25], on 1 January 2009, Public Health England [PHE; formerly the Health Protection Agency (HPA)] introduced a national enhanced surveillance scheme for STEC (NESSS) in England. Standardized microbiological, demographic, clinical and exposure data are collected and collated. These data are used to improve outbreak detection and elucidate the epidemiology of STEC in England. In this study we describe the microbiological characteristics and clinical impact of STEC infections, the demographic and geographical distribution of STEC infection and examine the associations between exposure to environmental factors and incidence of STEC infection by analysing data drawn from the first 4 years of enhanced surveillance (2009–2012). This is the first detailed report reconciling microbiological, epidemiological and clinical data for STEC cases in England.

METHODS

Microbiological investigations

Stool specimens from patients with suspected gastrointestinal infection were sent to local hospital laboratories where they were cultured for the presence of *E. coli* O157, following the UK Standards for Microbiology Investigations (http://www.hpa.org.uk/ProductsServices/MicrobiologyPathology/UKStandardsForMicrobiologyInvestigations/). Isolates identified locally as presumptive STEC O157, defined as non-sorbitol-fermenting *E. coli* agglutinating with *E. coli* O157 antisera, were sloped on nutrient agar and sent for confirmation and typing at the PHE Gastrointestinal Bacteria Reference Unit (GBRU) which provides the national reference service for STEC in England, Wales, and Northern Ireland. If STEC O157 was not isolated from a patient at the frontline laboratory, with clinical symptoms indicative of STEC (e.g. a child with HUS), faecal specimens were referred to GBRU and re-tested for the presence of both O157 and non-O157 STEC [26]. In cases where faecal specimens are unavailable, serum samples are referred for detection of antibodies to the lipopolysaccharide of *E. coli* O157 [27].

At GBRU, confirmation of STEC was performed using real-time PCR for the detection of *stx1, stx2, eae* (encoding intimin, associated with intimate attachment of the bacteria to the host gut mucosa) and *rfbE* O157 genes [26] on the sloped culture, which was then plated out onto sorbitol MacConkey (SMAC) agar, MacConkey agar and blood agar. Colonies from the plate found to be positive for either or both of the *stx* genes were identified as *E. coli* and serotyped using biochemical and serological tests [26]. Strains belonging to serogroup O157 were further differentiated by phage-typing [28].

Microbiological case definitions

Confirmed. STEC isolated from stool, serotyped and presence of *stx* genes confirmed.

Probable. Serum tested and antibodies to *E. coli* O157 or a limited range of other serotypes identified.

Negative. Specimen sent to GRBU but no evidence of STEC infection found

NESSS

Local laboratories reported presumptive isolates of STEC directly to PHE centres (PHEC), responsible for health protection. Each PHEC arranged for the STEC Enhanced Surveillance Questionnaire (ESQ) to be administered to patients. The ESQ collects data in the following categories: demographic details; risk status; clinical condition (including progression to HUS); household or other close contact details; exposures including travel, food and water consumption, contact with animals, and environmental exposures; case classification; outbreak status. Completed questionnaires were forwarded for inclusion in NESSS which is managed by the PHE Department of Gastrointestinal, Emerging and Zoonotic Infections (GEZI).

Epidemiological case definitions

Primary case. A symptomatic case with no history of close contact with a confirmed case in the 7 days prior to onset of illness.

Secondary case. Case with a date of onset >4 days after the primary case or where transmission is believed to be through exposure to a primary case.

Asymptomatic case. A person identified through contact screening procedures, with no symptoms consistent with STEC infection.

Unsure: It is not possible to determine whether the case is primary or secondary with the information available. This may be because the patient was lost to follow-up, is asymptomatic or in an outbreak where it is not possible to identify the primary case(s).

Travel-related case. Case who has reported any travel outside of the UK in the 7 days prior to their date of onset of illness.

Data handling

Data from laboratory referral forms and microbiological results were entered and stored in an electronic laboratory database. Data from each final report generated by GBRU was exported into NESSS and reconciled with data from ESQs based on patient identifiable information.

Data were exported into an Access database for coding and analysis. Incidence rates were calculated using the Office for National Statistics (ONS) 2010 mid-year population estimates as the denominator [29]. For incidence by ethnicity the 2009 population estimates were used for the denominator as the latest figures available [30]. Cases were assigned to rural, town/fringe or urban using case home postcodes. The Department for Environment, Food and Rural Affairs rurality classifications (http://www.defra.gov.uk/rural/ruralstats/rural-defn/rural-urban-method.pdf) were aggregated from six categories into three: urban, town & fringe, and rural. Incidence rates and rurality were visualized in ArcGIS at the middle super output level boundaries defined by ONS. Ethnic groups collected in five categories (white, Asian/Asian British, black/black British, mixed, Chinese) were recoded as white or non-white for analyses. Data were extracted from the laboratory database to compare age and gender composition of cases of STEC with cases of *Campylobacter* and *Salmonella*.

Incidence rate ratios (RR) were calculated in Stata v. 12·0 (Stata Corporation, USA) to compare incidence in different groups. Comparisons were made by broad geographical regions based on former HPA regions. Environmental exposures were examined for primary indigenous cases only. Rurality was treated as an ordered categorical variable and the χ^2 test for a linear trend used to assess the association between rurality and reported exposures. Comparisons were made between microbiological subtypes and disease severity using Fisher's exact test to test statistical significance ($P < 0{\cdot}05$).

RESULTS

Cases reported to NESSS

Between 1 January 2009 and 31 December 2012, 4792 suspected STEC cases were reported to NESSS (Fig. 1). This included 3939 (82·2%) patients for whom both an ESQ and one or more laboratory reports were received. Laboratory reports were received for 674 individuals for whom no questionnaire was received; most ($n = 561$, 83·2%) were STEC negative and 113 (3·0%) questionnaires for STEC-infected individuals were lost to follow-up. The number of questionnaires lost to follow-up was greatest ($n = 88$) in 2009, but declined to a total of 15 cases throughout the following 3 years.

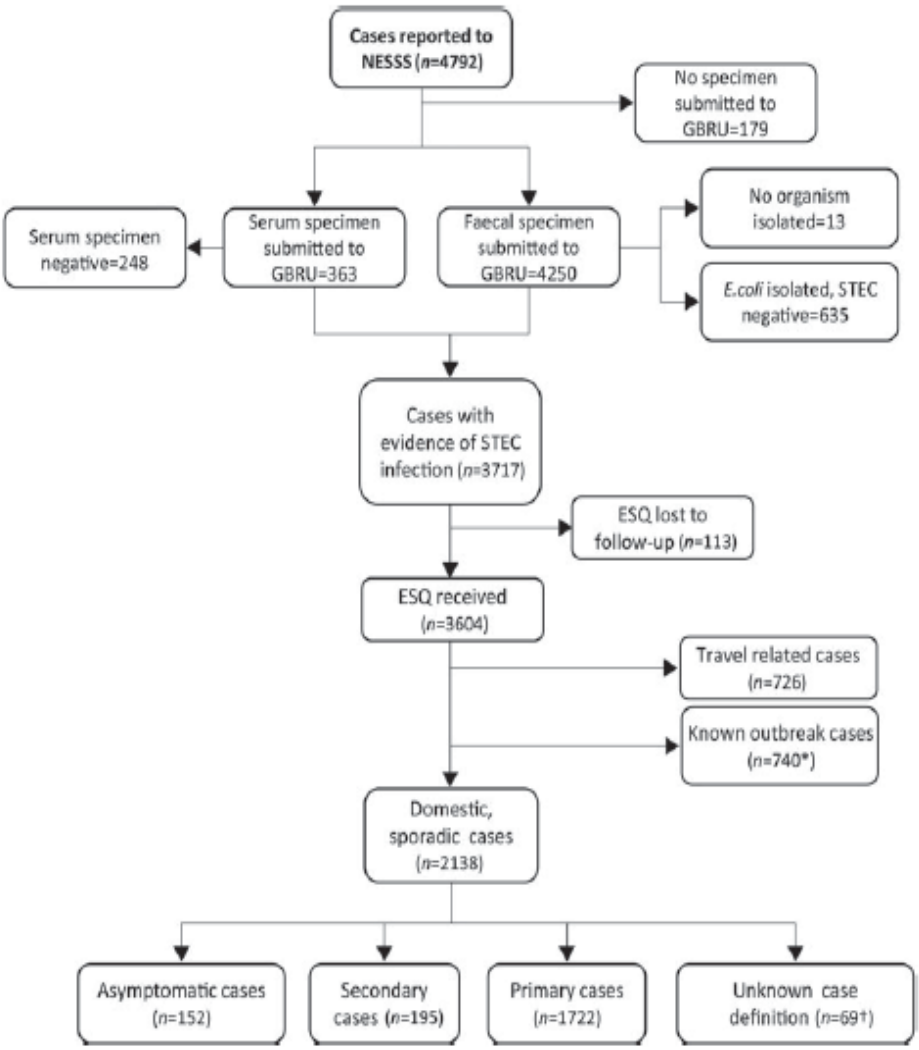


Fig. 1. Flowchart of all cases reported to the national enhanced surveillance scheme for STEC (NESSS), 2009–2012. GBRU, Gastrointestinal Bacteria Reference Unit; STEC, Shiga toxin-producing *E. coli*; ESQ, enhanced surveillance questionnaire. * Thirty-four cases were attributed to both travel and outbreaks. † It was not possible to determine an epidemiological case definition with the information available.

There was evidence of STEC infection for 3717 (80·6%) patients tested (Fig. 1) including 3602 patients with confirmed and 115 probable cases (patients with serological evidence of STEC infection). ESQs were received for 3604 confirmed (3512) and probable (92) cases. Of those, 740 were designated to known outbreaks and 726 were reported as travel cases; 34 of whom were attributed to travel-associated outbreaks. The most frequently visited countries for travel cases were Turkey ($n = 153$), followed by Egypt ($n = 101$) and Spain ($n = 91$).

A total of 2138 domestic, sporadic cases were reported; primary cases accounted for 80·5%, secondary 9·1% and 7·1% were asymptomatic. For 69 (3·3%) cases it was not possible to determine an epidemiological case definition with the available information.

Microbiological subtypes

The most commonly detected serogroup was O157, accounting for 3558 (98·8%) of cases (Table 1). Nineteen different non-O157 serogroups between 44 cases were detected. Foreign travel was reported significantly more frequently in non-O157 cases compared to O157 cases (52·3% *vs.* 20·6%, $P = 0·021$), while a similar proportion of O157 cases (21·1%) and non-O157 cases (20·5%) were associated with outbreaks. Serogroup O26 was the most common non-O157 serogroup with 15 cases (seven in 2012), five of whom were travel related. This was followed by O104 with six cases all linked to a large outbreak in Germany in summer 2011 [9, 18, 21]. Three cases

Table 1. *Serogroups of confirmed cases of STEC by travel status* in England, 2009–2012*

STEC serogroup	Travel status			
	Non-travel	Travel	Not known*	Total
O157	2738	734	86	3558
O26	8	5	2	15
O104	1	5	0	6
O145	2	1	0	3
O unidentifiable	1	0	1	2
O113	2	0	0	2
O146	2	0	0	2
O27	0	1	0	1
O103	0	1	0	1
O111AC	0	1	0	1
O118	0	1	0	1
O156	1	0	0	1
O161	1	0	0	1
O165	1	0	0	1
O172	1	0	0	1
O186	0	1	0	1
O49	1	0	0	1
O71	0	1	0	1
O92	1	0	0	1
O117	1	0	0	1
O159	0	0	1	1
All serogroups	2761	751	90	3602

STEC, Shiga toxin-producing *E. coli*.
* Cases where no enhanced surveillance questionnaire (ESQ) was provided and travel status was therefore unknown.

of O145 were reported, including one case with STEC O26 also isolated from their specimen.

Most STEC strains carried *stx2* genes; 2390 (66·3%) carried *stx2* genes only, 1194 (33·2%) encoded both *stx1* and *stx2*. Just 20 (0·5%) cases were infected with strains carrying *stx1* genes only. A higher proportion of non-O157 strains carried *stx1* genes only than O157 strains (15·9% *vs.* 0·36%, $P < 0·001$). All O157 STEC strains encoded intimin, whereas almost half (21/44) non-O157 STEC strains encoded intimin. No O157 strains were sorbitol fermenting, compared to 33 (75·0%) non-O157 strains.

Thirty-six different phage types (PTs) were isolated from cases with STEC O157 infection (Table 2). The most common overall were PT21/28 and PT8 with 1076 and 1069 cases, respectively. For each year of the study these two PTs contributed between 54·1% and 58·0% of STEC O157 cases except for 2011 (69·2%) when indigenous cases of PT8 were elevated due to 232 (25·4%) cases associated with a single national outbreak spanning December 2010 to July 2011 [31]. There was greater diversity in travel-related cases with 53·9% of cases being of a PT other than PT21/28 or PT8 compared to 36·0% of indigenous cases ($P < 0·0001$). In travel-related cases PT8 was the most common (41·6%) and PT21/28 comprised a very small number of infections (4·5%), whereas in indigenous cases it was the predominant PT (36·9%).

Faecal specimens were referred to GBRU for 646 STEC-negative cases. Presumptive STEC was isolated from specimens from 633 (98·3%) patients but were found to be negative by PCR for the presence of *stx* genes (i.e. *stx*-negative *E. coli* was isolated). In these isolates, 78 different *E. coli* serogroups were detected; O157 most frequently ($n = 150$), of which PT1 was the most common PT ($n = 71$). A higher proportion of *stx*-negative *E. coli* O157 strains fermented sorbitol than STEC O157 strains (14·8% *vs.* 0·25%, $P < 0·0001$). Other frequently detected ($n \geqslant 10$) serogroups were O145, O25, O148, O6, O2 and O49. Intimin was encoded by 146 (23·1%) *stx*-negative *E coli* strains including 74 serogroup O157 strains.

Incidence

The crude incidence of confirmed and probable STEC was 1·80/100 000 person-years [95% confidence interval (CI) 1·74–1·86]; however, this varied by age, gender, ethnicity and geography.

Age, gender and ethnicity

Incidence of confirmed and probable STEC increased from infants aged <1 year to its highest in those aged 1–4 years (7·63/100 000 person-years, 95% CI 7·11–8·18) (Fig. 2), it then declined with each subsequent age group to reach a trough in the 20–59 years (1·15/100 000 person-years, 95% CI 1·09–1·22) and ⩾60 years (1·16/100 000 person-years, 95% CI 1·07–1·86) age groups. Over 40% of reported STEC cases were children aged <15 years (Fig. 3*a*), compared to 7·0% for cases of campylobacteriosis (Fig. 3*b*) and 17·8% for cases of salmonellosis (Fig. 3*c*).

Females were overrepresented in STEC cases with 55·6% of cases being female and a higher incidence than males (RR 1·24, $P < 0·001$). The gender difference was most apparent in those aged 20–59 years (RR 1·73, $P < 0·001$) and ⩾60 years (RR 1·43, $P < 0·001$). This gender disparity was not observed for cases of campylobacteriosis (53·3% male) or salmonellosis (50·1% male).

Ethnicity was reported for 1430 (38·5%) cases and 1300 (90·9%) were of white ethnicity. Incidence in

Table 2. *Phage types of confirmed STEC O157 cases in England by travel status: 2009–2012*

STEC O157 PT	Travel-related cases, *n* (%)	Indigenous cases, *n* (%)	All cases*, *n* (%)
PT21/28	33 (4·50)	1014 (37·03)	1047 (30·16)
PT8	305 (41·55)	751 (27·43)	1056 (30·41)
PT32	93 (12·67)	197 (7·20)	290 (8·35)
PT2	13 (1·77)	181 (6·61)	194 (5·59)
PT4	13 (1·77)	102 (3·73)	115 (3·31)
PT14	57 (7·77)	51 (1·86)	108 (3·11)
PT34	40 (5·45)	61 (2·23)	101 (2·91)
PT54	38 (5·18)	57 (2·08)	95 (2·74)
PT31	19 (2·59)	58 (2·12)	77 (2·22)
PT33	4 (0·54)	64 (2·34)	68 (1·96)
Other phage types	49 (6·68)	116 (4·24)	165 (4·75)
Uncharacterized†	70 (9·54)	86 (3·14)	156 (4·49)
Total	734 (100·00)	2738 (100·00)	3472 (100·00)

PT, Phage type; STEC, Shiga toxin-producing *E. coli*.
* Excludes 86 cases where no enhanced surveillance questionnaire (ESQ) was received and travel status was therefore unknown.
† Includes untypable isolates and PT results which did not conform to a known PT.

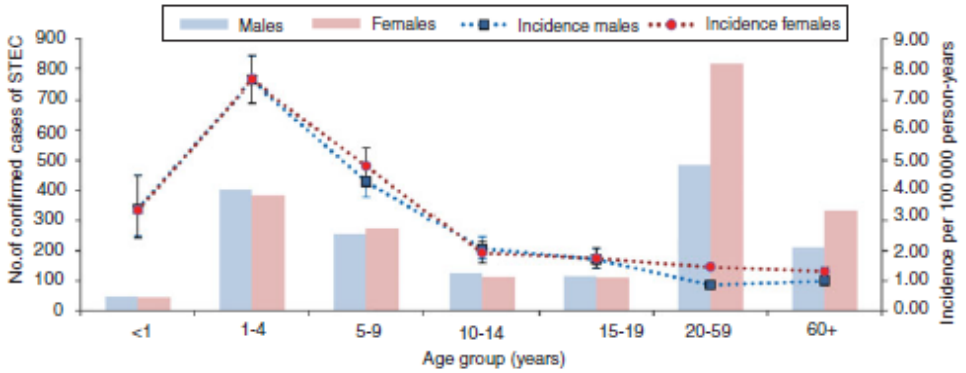


Fig. 2. No. confirmed and probable cases of Shiga toxin-producing *E. coli* (STEC) and incidence of STEC/100 000 person-years by age group and gender reported to the national enhanced surveillance scheme for STEC (NESSS), 2009–2012.

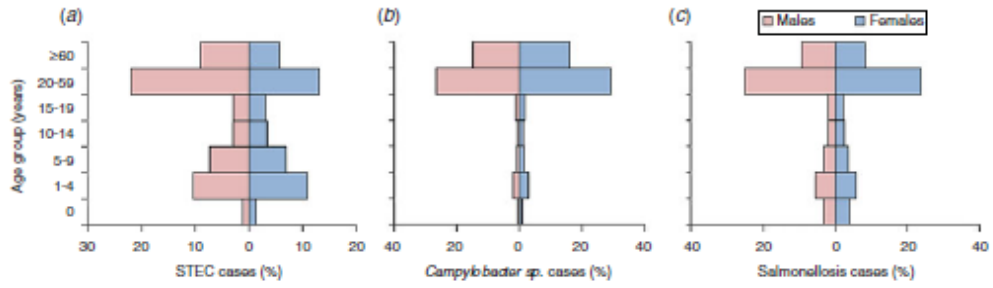


Fig. 3. Age and sex distributions for cases of (*a*) Shiga toxin-producing *E. coli* (STEC), (*b*) campylobacteriosis, and (*c*) salmonellosis and reported to national surveillance systems in England, 2009–2012.

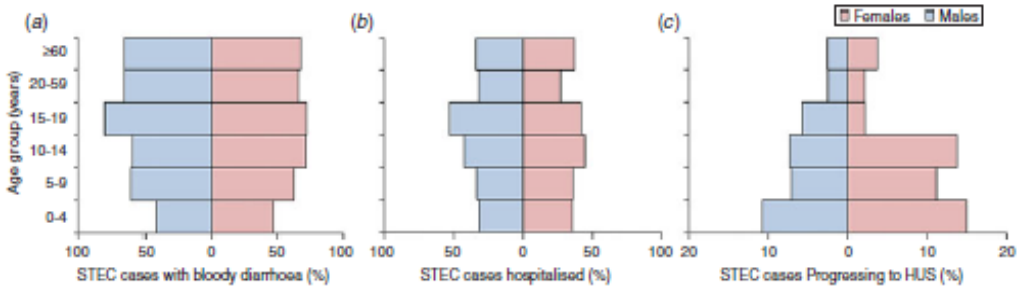


Fig. 4. Age and sex distribution for symptomatic confirmed and probable Shiga toxin-producing *E. coli* (STEC) cases in England reporting (*a*) bloody diarrhoea, (*b*) hospitalization and (*c*) progression to HUS, 2009–2012.

the white population was significantly higher than those of non-white ethnicity (RR 1·43, $P < 0·001$).

Clinical features

ESQs were available and symptomatic infection reported for 3267 (90·7%) confirmed and 92 (80·0%) probable cases of STEC. A history of diarrhoea was indicated for 92·8% (3117) cases, including bloody diarrhoea for 61·0% (2050). Abdominal pain was frequently reported (79·2%, 2662), while fever (32·5%, 1092), nausea (46·6%, 1564) and vomiting (37·3%, 1252) were reported less often. Half (1745) of cases were still ill when the ESQ was administered, the median duration of illness for the remaining cases was 6 days [interquartile range (IQR) 1–92]. Hospitalization was reported for 34·3% (1151) of cases and the median duration of hospitalization was 3 days (IQR 1–21). Cases were reported as STEC-HUS either on laboratory referral forms or ESQs for 6·4% (215) of cases. Thirteen deaths were reported, although cause of death was not provided. These included eight STEC O157 cases, one O26 case, and four fatalities with serological evidence of O157 infection only. Nine were female and all but one were adult cases with a mean age of 65 years (95% CI 52·7–77·3).

For most age groups, bloody diarrhoea and hospitalization in confirmed and probable STEC cases was similarly reported for both sexes (Fig. 4*a*, *b*). Both were reported most frequently in males aged 15–19 years; however, increased progression to HUS in this group was not observed (Fig. 4*c*). Three quarters of STEC-HUS cases were in children aged 0–14 years, with a mean age of 14·6 years (95% CI 11·7–17·5). STEC-HUS was more frequently reported in females than males in both children and adults aged ⩾60 years (Fig. 4*c*). The highest proportion of STEC-HUS cases was in females aged 1–4 years at 15·0%. Although serum specimens only were submitted for 92 probable STEC cases, a diarrhoeal prodrome was reported in most (80) cases, including bloody diarrhoea in over half (47). The majority (86) were hospitalized and over half (50) developed HUS.

Severity of illness also varied by microbiological subtype and specimen type (Table 3). In STEC O157 cases, vomiting was more frequently reported in cases of PT21/28 than PT8 (40·1% *vs.* 34·2%, $P = 0·0078$). Hospitalization rates were higher (41·7% *vs.* 29·3%, $P < 0·001$), as was progression to HUS (8·5% *vs.* 1·1%, $P < 0·001$). PT2 was also associated with more severe disease than PT8, including vomiting (43·8% *vs.* 34·2%, $P = 0·0153$), hospitalization (43·8%, $P = 0·0001$), and progression to HUS (7·6%, $P < 0·0001$).

There were 18 HUS cases with non-O157 STEC serogroups; ten cases of O26, two cases of O104 (related to the outbreak in Germany in 2011) and one case each of O113, O145, O159, O165, O172, and O186. Clinical features in non-O157 cases did not differ significantly by presence of intimin although numbers were small. Hospitalization rates were higher in non-O157 cases compared to O157 cases (Table 3, $P = 0·0121$). Progression to HUS was reported ten times more frequently in non-O157 cases than O157 cases ($P = 0·0001$). Two cases of non-O157 STEC-HUS, including one fatal O26 infection, were reported but no ESQ was received so information on clinical prodrome was not available. Isolates from all cases of confirmed STEC-HUS encoded *stx2*; 139 were *stx2* only, and 23 were *stx1* and *stx2*.

Clinical data were available for 302 cases infected with *stx*-negative *E. coli* [89 carried the intimin gene (*eae*)]. Of these cases, diarrhoea, vomiting, abdominal pain, fever, and nausea were reported by similar proportions as STEC cases. Bloody diarrhoea, however, was reported by only 60 (19·9%) and hospitalization by 49 (16·2%) of these *stx*-negative *E. coli* cases.

Table 3. *Clinical features of confirmed STEC cases in England, by microbiological subtype, 2009–2012*

Clinical feature	Serotype O157					Non-O157 serotypes	All serotypes
	PT21/28	PT8	PT2	Other PTs	All O157		
No. of cases*	967	1002	185	1077	3231	36	3267
Diarrhoea	878 (90·8)	957 (95·5)	176 (95·1)	996 (92·5)	3007 (93·1)	30 (83·3)	3037 (93·0)
Bloody diarrhoea	665 (68·8)	667 (66·6)	137 (74·0)	513 (47·6)	1982 (61·3)	21 (58·3)	2003 (61·3)
Abdominal pain	742 (76·7)	857 (85·5)	155 (83·8)	822 (76·3)	2576 (79·7)	25 (69·4)	2601 (79·6)
Vomiting	388 (40·1)‡	343 (34·2)	81 (43·8)‡	356 (33·0)	1168 (36·2)	15 (41·7)	1183 (36·2)
Fever	297 (30·71)	340 (33·9)	64 (34·6)	696 (64·6)	1036 (32·1)	10 (27·8)	1046 (32·0)
Hospitalizations	403 (41·7)‡	294 (29·3)	81 (43·8)‡	268 (24·9)	1046 (32·4)	19 (52·8)§	1065 (32·6)
HUS†	82 (8·5)‡	11 (1·1)	14 (7·6)‡	35 (3·3)	142 (4·4)	16 (44·4)§	158 (4·8)
Deaths†	6 (0·6)	0 (0·0)	0 (0·0)	2 (0·2)	8 (0·3)	0 (0·0)	8 (0·24)

Values given are *n* (%).
STEC, Shiga toxin-producing *E. coli*; PT, phage type.
* Cases where an enhanced surveillance questionnaire (ESQ) was received and reported as symptomatic only (excludes 80 cases where ESQs were lost to follow-up and 245 cases reported as asymptomatic).
† Three additional haemolytic uraemic syndrome (HUS) cases (O159 *eae*-, O157 *eae*+, O26 *eae*+), including one death (O26 strain) were reported but no ESQ was received so are not included.
‡ Denotes statistically significant difference compared to PT8.
§ Denotes a statistically significant difference compared to all O157 cases.

There were no significant differences in reporting of symptoms between *stx*-negative *E. coli* depending on the presence of intimin (data not shown). However cases with *stx*- and *eae*-negative strains were significantly more likely to report hospitalization than those with intimin-positive strains (24·07% *vs.* 8·53, $P = 0·001$).

An additional 115 cases of HUS were reported to NESSS without evidence of STEC, including three who died: 87 cases had only submitted serum samples which were negative for antibodies to *E. coli* O157 and no specimen was received by GBRU for 13 cases. For 15 HUS cases *stx*-negative *E. coli* was isolated, including one unidentifiable serogroup and three O157 strains encoding intimin. A gender disparity was also observed in these HUS cases with 55·7% being female with a mean age of 25·5 years (95% CI 20·5–30·5), which was significantly higher than 14·6 years for STEC-HUS cases ($P = 0·0001$). Compared to STEC-HUS cases, a significantly higher proportion of cases were adults aged ⩾60 years (20·0% *vs.* 7·9%, $P = 0·007$).

Seasonality

Two-thirds of STEC cases were reported in the summer months of May–September, with the highest frequency of cases in August of each year (data not shown). Similar seasonal fluctuations were observed in both domestic and travel cases. While the majority (48/72, 66·7%) of outbreaks occurred in the summer months, outbreaks also occurred in the winter months. Between December 2010 and June 2011, the largest recorded national outbreak of STEC in the UK occurred with 167 associated cases in England [31].

Geography

There were regional variations in the incidence of STEC. Incidence was lowest in the London region (1·04/100 000 person-years) and highest in the Yorkshire & Humber region (2·48/100 000 person-years). Rates were similarly high in the North East, North West and South West regions. The age and gender distribution of cases was similar across all regions. Ethnicity, however, varied regionally, with the highest non-white British population comprising 40·5% of the population in London and the lowest in the North East (7·6%) [18].

The incidence of primary domestic sporadic STEC cases declined with decreasing rurality (Fig. 5). Compared to cases living in urban areas, those living in town and fringe areas had a significantly higher incidence rate (RR 1·27, $P < 0·001$), while the incidence of STEC was over four times higher in people living in rural areas than those from urban areas (RR 4·39, $P < 0·001$).Many areas had zero incidence (i.e. no cases) of STEC infection during the study period,

Fig. 5. (*a*) Rurality and (*b*) geographical distribution of Shiga toxin-producing *E. coli* (STEC) incidence in England, 2009–2012.

and although the highest incidence was in rural areas, not all rural areas had high incidence and some had no STEC cases during the study period.

Geographically associated factors

Of 1722 primary, domestic, sporadic cases for whom an ESQ was received, one or more environmental exposures were reported for 681 (39·5%) (Table 4). These included direct or indirect contact with farm animals and/or their faeces (649, 37·7%), drinking water from private water supplies (74, 4·3%) and recreational exposure to open-freshwater sources, e.g. paddling, swimming or fishing in ponds, lakes, rivers or streams (63, 3·7%).

While contact with farm animals was reported in a farm setting for 301 (17·5%) cases, an additional 298 (17·3%) cases only reported contact with farm animals or their faeces in other settings, including at agricultural shows, livery yards and in paddocks or fields where animals were grazing. The species of farm animal was not specified for 298 (45·9%) cases who reported farm animal/faeces contact. Where one or more animal species was reported ($n = 351$), cattle were most frequently reported ($n = 173$), followed by horses ($n = 148$), sheep ($n = 133$), poultry ($n = 127$), goats ($n = 57$) and pigs ($n = 54$). Exposure to more than one species was reported for 187 cases.

Reporting of any environmental exposure increased with increasing rurality (Table 4) and was reported by twice the percentage of cases in rural areas compared to urban areas ($P < 0·001$). Reported consumption of water from private water supplies, open freshwater exposure, direct and indirect contact with farm animals/faeces were significantly higher in rural and town and fringe cases compared to urban cases ($P < 0·001$). Where farm settings were reported, the majority (84·5%) of urban cases reported visiting an open farm or holidaying on a farm, whereas the majority (58·9%) of rural cases reported living on, working at, or having access to a farm through relatives or neighbours (data not shown).

DISCUSSION

In England the highest burden of STEC infection is in children, consistent with previous reports from the UK and elsewhere [25, 32, 33]. Children with infectious intestinal disease are more likely than adults to present to primary care and have clinical specimens taken [34], meaning laboratory reporting emphasizes the impact of disease in children more than adults. However, the skew towards children in the age distribution of STEC cases is not observed for *Salmonella enterica* or *Campylobacter* sp. Transmission routes are largely environmental for STEC and less so for

Table 4. *Reported environmental exposures of 1722 primary, domestic, sporadic STEC cases in England, 2009–2012, by urbanicity of residential postcode*

Exposure/urbanicity group	Exposed *n* (%)	χ^2 test for trend *P* value
Any environmental exposure*		
Urban	340 (29·9)	
Town and fringe	113 (53·6)	
Rural	228 (61·3)	<0·0001
All cases	681 (39·5)	
Private water supplies		
Urban	22 (1·9)	
Town and fringe	5 (2·4)	
Rural	47 (12·6)	<0·0001
All cases	74 (4·3)	
Open freshwater exposure†		
Urban	33 (2·9)	
Town and fringe	14 (6·6)	
Rural	16 (4·3)	<0·0001
All cases	63 (3·6)	
Direct/indirect contact with farm animals/faeces on farms‡		
Urban	168 (14·7)	
Town and fringe	51 (24·2)	
Rural	82 (22·0)	<0·0001
All cases	301 (17·5)	
Direct/indirect contact with farms animals/faeces in non-farm setting§		
Urban	156 (13·7)	
Town and fringe	57 (27·0)	
Rural	135 (36·3)	0·0002
All cases	348 (20·2)	

STEC, Shiga toxin-producing *E. coli*.

* Includes any reported exposures to private water supplies, open freshwater sources and any contact with farm animals and/or their faeces.

† Includes paddling, swimming, boating or fishing in a freshwater pond, stream, lake or river.

‡ Includes direct or indirect contact with farm animals including cattle, sheep, goats, horses, pigs and poultry at private or open farms.

§ Includes direct or indirect contact with farm animals including cattle, sheep, goats, horses, pigs and poultry *not* at farms but in other settings including walking in fields where those animals graze or exposure to manure.

S. enterica or *Campylobacter* sp., and as children have poorer hygiene practices than adults this may contribute to the higher incidence in children. The low infectious dose and propensity for person-to-person transmission of STEC in childcare facilities also increases risk in these age groups [4, 7, 8]. The nature of STEC as an environmentally acquired pathogen is reflected in the increased incidence from infants aged <1 year to those aged 1–4 years. While infants have more limited food consumption than older age groups, they are less mobile and therefore less exposed to contaminated environments.

A small number of reports have described a gender disparity in adults with higher infection rates reported in females [16, 18, 33, 35]. To our knowledge the reasons for this are not established. There may be host factors placing women at greater risk of developing severe symptoms. Alternatively, females might have a higher level of exposure to foodborne STEC as principal food handlers; however, a gender disparity was not apparent for *S. enterica* or *Campylobacter* sp. Females may have different consumption practices compared to males which increase their risk of STEC. Similarly, if adult females are more commonly primary carers within households, they may be expected to have higher levels of exposure to STEC excreted by primary cases, particularly children, increasing their risk of acquiring infection through person-to-person transmission. Additional questions have been added to the ESQ so that these factors can be explored in more detail. There are marked variations in the geographical distribution of ethnic minority populations in England which may in part explain the higher incidence rates in those of white ethnicity than non-white ethnicity. There may also be behavioural aspects reducing the risk of acquiring STEC infection in ethnic minorities, or differences in ethnic groups accessing primary care. Ethnicity was only reported for 38·5% of cases and ethnicity categories were very broad: collection of more detailed ethnicity data would enhance the ability to explore these differences in more detail.

The geographical distribution of reported STEC infection in the UK and Republic of Ireland is extremely heterogeneous. At the country level, Ireland and Scotland report rates of infection which are more than twice that of England (5·33 and 4·4 *vs.* 1·80 cases/100 000 population [5, 32, 36]. In England the highest observed rates of infection were reported in the Northern and Western regions, consistent with the finding that cases were more likely to live in rural areas than urban areas. While this observation has been reported in previous studies the risk ratio of rural to urban cases observed in England was over twice that reported from Grampian, Scotland [37]. This might be expected given the relative diversity of England's population, topography, land usage and climate. The increased risk associated with living in rural areas and the wider geographical variations in disease incidence observed across the UK and Ireland appear to be markers for environmental exposure to

STEC in areas where ruminants are raised [32, 37]. Cases living in rural areas were more than twice as likely than those from urban areas to have had contact with livestock. Farm exposures differed in nature between urban and rural cases, with the former being exposed more often to commercial premises and rural cases to private farms. Open farms have the potential to cause outbreaks, [11, 12] and guidelines exist around control of infection [38, 39]. Our findings demonstrate that environmental/animal contact remains an important risk factor for disease transmission, and that non-commercial farm premises are a significant driver in the burden of sporadic STEC infection.

The morbidity burden of STEC in England remains high; overall a third of cases were hospitalized and progression to HUS was reported for 6·4% (215) STEC cases in England. This compares with 9·7% for Scotland and 7·8% for the USA [5, 40]. A relatively large number of HUS cases were microbiologically undiagnosed due to no available faecal specimens and this will have contributed to under-ascertainment of STEC-HUS cases. In addition, collection of HUS data on the laboratory referral form and ESQ may occur prior to a case developing HUS leading to further under-ascertainment. A high proportion of patients who submitted serum samples had HUS and no specimens were submitted for a small number of reported HUS patients, although, many reported a diarrhoeal prodrome including bloody diarrhoea. The operational guidelines for control of STEC infection in England recommend that serum and faecal specimens should be submitted to GBRU for further testing for all cases of HUS [41]. The microbiologically uncharacterized HUS cases were older than STEC-HUS cases, many in adults, a group not commonly recognized as being at risk of developing STEC-HUS and it may be that they are not being investigated and/or managed for STEC infection early in their course of illness. It is well documented that children, particularly those aged <5 years are at greater risk of developing STEC-HUS, and our findings are consistent with this [1, 40, 42, 43]. However, progression to HUS was greater in females and children and those aged ⩾60 years. This may again be suggestive of a host factor placing females at greater risk of developing STEC-HUS, and further investigation is needed.

The low identification rate (1·2%) of non-O157 strains during the study period is consistent with previous surveillance reports in England (and Wales) [24, 25]. However, in other European countries, non-O157 STEC have been identified as a significant cause of disease [16, 18, 32] and reporting to Europe-wide surveillance in 2010 [36], indicates non-O157 STEC accounts for many more reported infections. This is unsurprising given local diagnostic laboratories in England do not routinely screen diarrhoeal stool specimens for non-O157 STEC. Higher detection rates of STEC O26, particularly in Scotland and the Republic of Ireland, further highlight the under-ascertainment of non-O157 STEC in England [32, 36]. Significantly higher hospitalization and HUS rates were reported for non-O157 cases. However, while some non-O157 strains can clearly cause severe disease, faecal samples from patients with severe disease suggestive of STEC infection are referred to GBRU for comprehensive testing using a combination of PCR and culture of faeces. Consequently, there is bias towards detection of non-O157 STEC and sorbitol-fermenting STEC O157 from patients with disease from the severe end of the clinical spectrum – which may contribute to the significantly higher proportion of HUS cases in non-O157 strains reported here. Thus the overall burden of non-O157 STEC in England is, largely, unobservable through current surveillance systems. It is important therefore to encourage the implementation and evaluation of PCR-based diagnostic methods designed to detect all STEC, to address this diagnostic deficit and to ultimately inform and improve the health protection response to these pathogens.

PT21/28 was the most frequently detected PT in STEC-HUS cases and PT21/28 and PT2 cases were significantly more likely to report vomiting, hospitalization and HUS. All STEC-HUS cases had strains encoding either Stx2 only or Stx1 and Stx2 toxins, consistent with previous studies indicating that both PT21/28 and PT2, and Stx2 toxin profiles are associated with progression to HUS [1, 42]. The most common PT in indigenous cases was PT21/28, while PT8 which caused less severe disease was largely travel acquired. Thus, the greatest burden in terms of both numbers and morbidity of STEC in England is due to domestic acquisition and is where interventions should be targeted.

Confirmatory tests conducted by GBRU on specimens from presumptive cases demonstrated that almost one fifth of cases reported to NESSS did not have STEC infection, but many were infected with *stx*-negative *E.coli* O157. Most were common serogroups of non-sorbitol-fermenting strains of *E. coli*, many of which exhibited cross-reactions with *E. coli* O157 antiserum (in-house data from GBRU). The

initial screening processes used by firstline diagnostic laboratories are not designed to distinguish between STEC O157 and *stx*-negative *E. coli* O157. These isolates are thus considered as presumptive STEC O157 and cases are rapidly referred to local public health specialists for management including administration of the ESQ. The symptoms reported in cases of *stx*-negative *E. coli* in this dataset were similar to that of STEC cases, with the exception of bloody diarrhoea and HUS, indicating less severe disease was associated with *stx*-negative strains. *stx*-negative strains have been shown to be isolated during the course of STEC infection and may be progenies of STEC strains that have lost the *stx*-enconding phage [44]. However, the less severe symptoms and lower hospitalization rates in this group, compared to STEC cases, along with the absence of known pathogenicity factors in most strains, indicate they are more likely to be commensal strains mistakenly identified as STEC at the frontline laboratories and that some other microbiological or non-microbiological cause of gastrointestinal symptoms was present but not identified. Almost a third of *stx*-negative strains had the *eae* gene but the *eae*-positive strains were less likely to cause hospitalization or be associated with bloody diarrhoea than the *eae*-negative strains, supporting the assumption that the *stx*-negative strains are unlikely to be progenies of STEC strains that have lost the *stx*-encoding phage.

Continuation of NESSS will allow monitoring and analysis of longer term trends including emerging subtypes and risk factors. The addition of increased characterization of STEC strains through the application of molecular methods such as multi-locus variable number tandem repeat and whole genome sequencing will supplement enhanced surveillance data. Thus information from ESQs should work in synergy with microbiological characterization data to build a detailed picture of STEC strains, their pathogenicity and impact.

ACKNOWLEDGEMENTS

The authors thank Neil Perry, Vivienne do Nascimento, Michael Wright and Marie Chattaway in GBRU for their dedication and hard work in the laboratory. Thanks are also due to Kirsten Glen, Natalie Adams, and Radha Patel for their contributions to the National Enhanced Surveillance System for STEC in England, and to all public health practitioners who report to the scheme.

This research received no specific grant from any funding agency, commercial or not-for-profit sectors.

DECLARATION OF INTEREST

None.

REFERENCES

1. **Lynn RM, *et al.*** Childhood hemolytic uremic syndrome, United Kingdom and Ireland. *Emerging Infectious Diseases* 2005; **11**: 590–596.
2. **Armstrong GL, Hollingsworth J, Morris Jr. JG.** Emerging foodborne pathogens: *Escherichia coli* O157: H7 as a model of entry of a new pathogen into the food supply of the developed world. *Epidemiologic Reviews* 1996; **18**: 29–51.
3. **Ferens WA, Hovde CJ.** *Escherichia coli* O157: H7: animal reservoir and sources of human infection. *Foodborne Pathogens and Disease* 2011; **8**: 465–487.
4. **Tuttle J, *et al.*** Lessons from a large outbreak of *Escherichia coli* O157: H7 infections: insights into the infectious dose and method of widespread contamination of hamburger patties. *Epidemiology and Infection* 1999; **122**: 185–192.
5. **Locking ME, *et al.*** *Escherichia coli* O157 infection and secondary spread, Scotland, 1999–2008. *Emerging Infectious Diseases* 2011; **17**: 524–527.
6. **Parry SM, Salmon RL.** Sporadic STEC O157 infection: secondary household transmission in Wales. *Emerging Infectious Diseases* 1998; **4**: 657–661.
7. **Rangel JM, *et al.*** Epidemiology of *Escherichia coli* O157: H7 outbreaks, United States, 1982–2002. *Emerging Infectious Diseases* 2005; **11**: 603–609.
8. **Wahl E, *et al.*** Investigation of an *Escherichia coli* O145 outbreak in a child day-care centre – extensive sampling and characterization of eae- and stx1-positive *E. coli* yields epidemiological and socioeconomic insight. *BMC Infectious Diseases* 2011; **11**: 238.
9. **Buchholz U, *et al.*** German outbreak of *Escherichia coli* O104: H4 associated with sprouts. *New England Journal of Medicine* 2011; **365**: 1763–1770.
10. **Michino H, *et al.*** Massive outbreak of *Escherichia coli* O157: H7 infection in schoolchildren in Sakai City, Japan, associated with consumption of white radish sprouts. *American Journal of Epidemiology* 1999; **150**: 787–796.
11. **Ihekweazu C, *et al.*** Large outbreak of verocytotoxin-producing *Escherichia coli* O157 infection in visitors to a petting farm in South East England, 2009. *Epidemiology and Infection* 2012; **140**: 1400–1413.
12. **Willshaw GA, *et al.*** Verocytotoxin-producing *Escherichia coli* infection and private farm visits. *Veterinary Record* 2003; **152**: 365–366.
13. **March SB, Ratnam S.** Sorbitol-MacConkey medium for detection of *Escherichia coli* O157: H7 associated with hemorrhagic colitis. *Journal of Clinical Microbiology* 1986; **23**: 869–872.

14. **Chapman PA, *et al.*** An improved selective medium for the isolation of *Escherichia coli* O157. *Journal of Medical Microbiology* 1991; **35**: 107–110.
15. **Garrity GM, *et al.* (eds.).** *Bergey's Manual of Systematic Bacteriology. Volume 2: The Proteobacteria*, 2nd edn. New York: Springer, 2005.
16. **Schimmer B, *et al.*** Outbreak of haemolytic uraemic syndrome in Norway caused by stx2-positive *Escherichia coli* O103: H25 traced to cured mutton sausages. *BMC Infectious Diseases* 2008; **8**: 41.
17. **Bradley KK, *et al.*** Epidemiology of a large restaurant-associated outbreak of Shiga toxin-producing *Escherichia coli* O111:NM. *Epidemiology and Infection* 2012; **140**: 1644–1654.
18. **Frank C, *et al.*** Epidemic profile of Shiga-toxin-producing *Escherichia coli* O104:H4 outbreak in Germany. *New England Journal of Medicine* 2011; **365**: 1771–1780.
19. **Rounds JM, *et al.*** Non-O157 Shiga toxin-producing *Escherichia coli* associated with venison. *Emerging Infectious Diseases* 2012; **18**: 279–282.
20. **Taylor EV, *et al.*** Multistate outbreak of *Escherichia coli* O145 infections associated with romaine lettuce consumption, 2010. *Journal of Food Protection* 2013; **76**: 939–944.
21. **Buchholz U, *et al.*** German outbreak of *Escherichia coli* O104: H4 associated with sprouts. *New England Journal of Medicine* 2011; **365**: 1763–1770.
22. **Karmali MA, *et al.*** Association of genomic O island 122 of *Escherichia coli* EDL 933 with verocytotoxin-producing *Escherichia coli* seropathotypes that are linked to epidemic and/or serious disease. *Journal of Clinical Microbiology* 2003; **41**: 4930–4940.
23. **European Food Safety Authority (EFSA).** Panel on Biological Hazards (BIOHAZ). Scientific opinion on STEC seropathotype and scientific criteria regarding pathogenicity assessment. *EFSA Journal* 2013; **11**: 3138.
24. **Thomas A, *et al.*** Vero cytotoxin-producing *Escherichia coli*, particularly serogroup O157, associated with human infections in England and Wales: 1992–4. *Epidemiology and Infection* 1996; **117**: 1–10.
25. **Willshaw GA, *et al.*** Verocytotoxin-producing *Escherichia coli* (VTEC) O157 and other VTEC from human infections in England and Wales: 1995–1998. *Journal of Medical Microbiology* 2001; **50**: 135–142.
26. **Jenkins C, *et al.*** Assessment of a real-time PCR for the detection and characterization of verocytotoxigenic *Escherichia coli*. *Journal of Medical Microbiology* 2012; **61**: 1082–1085.
27. **Chart H, *et al.*** Analysis of saliva for antibodies to the LPS of *Escherichia coli* O157 in patients with serum antibodies to *E. coli* O157 LPS. *Journal of Medical Microbiology* 2003; **52**: 569–572.
28. **Khakhria R, Duck D, Lior H.** Extended phage-typing scheme for *Escherichia coli* O157: H7. *Epidemiology and Infection* 1990; **105**: 511–520.
29. **Office for National Statistics.** 2010 mid-year population estimates (http://www.ons.gov.uk/ons/rel/pop-estimate/population-estimates-for-uk–england-and-wales–scotland-and-northern-ireland/mid-2010-population-estimates/index.html). Accessed 10 September 2012.
30. **Office for National Statistics.** Statistical bulletin: population estimates by ethnic group 2002–2009 (http://www.ons.gov.uk/ons/rel/pop-estimate/population-estimates-for-uk–england-and-wales–scotland-and-northern-ireland/mid-2010-population-estimates/index.html). Accessed 10 September 2012.
31. **Perry N, *et al.*** Application of multilocus variable number tandem repeat analysis to monitor Verocytotoxin-producing Escherichia coli O157 phage type 8 in England and Wales: emergence of a profile associated with a national outbreak. *Journal of Applied Microbiology* 2013; **115**: 1052–1058.
32. **Carroll AM, Gibson A, McNamara EB.** Laboratory-based surveillance of human verocytotoxigenic *Escherichia coli* infection in the Republic of Ireland, 2002–2004. *Journal of Medical Microbiology* 2005; **54**: 1163–1169.
33. **Vally H, *et al.*** Epidemiology of Shiga toxin producing *Escherichia coli* in Australia, 2000–2010. *BMC Public Health* 2012; **12**: 63.
34. **Wheeler JG, *et al.*** Study of infectious intestinal disease in England: rates in the community, presenting to general practice, and reported to national surveillance. The Infectious Intestinal Disease Study Executive. *British Medical Journal* 1999; **318**: 1046–1050.
35. **Vrbova L, *et al.*** A descriptive study of reportable gastrointestinal illnesses in Ontario, Canada, from 2007 to 2009. *BMC Public Health* 2012; **12**: 970.
36. **European Food Safety Authority, European Centre for Disease Prevention and Control.** The European Union summary report on trends and sources of zoonoses, zoonotic agents and food-borne outbreaks in 2010. *EFSA Journal* 2012 **10**(3): 2597 [442 pp] (www.efsa.europa.eu/efsajournal).
37. **Strachan NJ, *et al.*** Escherichia coli O157: burger bug or environmental pathogen? *International Journal of Food Microbiology* 2006; **112**: 129–137.
38. **Access to Farms Partnership**. Preventing or controlling ill health from animal contact at visitor attractions: Industry Code of Practice. June 2012.
39. **Public Health England.** Avoiding infection on farm visits: advice for the public. March 2014.
40. **Walker CL, Applegate JA, Black RE.** Haemolytic-uraemic syndrome as a sequela of diarrhoeal disease. *Journal of Health Population and Nutrition* 2012; **30**: 257–261.
41. **Health Protection Agency.** The VTEC operational manual; Operational guidance for HPA staff dealing with cases and incidents of VTEC infection, 2011.
42. **Kleanhouse H, *et al.*** Haemolytic uraemic syndromes in the British isles, 1985–8: association with Verocytotoxin producing *Escherichia coli*. Part 2: microbiological aspects. *Archives of Disease in Childhood* 1990; **65**: 722–727.
43. **Milford DV, *et al.*** Haemolytic uraemic syndromes in the British isles, 1985–8: association with Verocytotoxin producing *Escherichia coli*. Part 1: clinical and epidemiological aspects. *Archives of Disease in Childhood* 1990; **65**: 716–721.
44. **Bielaszewska M, *et al.*** Shiga toxin-negative attaching and effacing *Escherichia coli*: distinct clinical associations with bacterial phylogeny and virulence traits and inferred in-host pathogen evolution. *Clinical Infectious Diseases* 2008; **47**: 208–217.

Paper 2: Shiga toxin-producing *Escherichia coli* O157 in England and Wales, 1983-2012.

SYNOPSIS

Shiga Toxin–Producing *Escherichia coli* O157, England and Wales, 1983–2012

Natalie L. Adams, Lisa Byrne, Geraldine A. Smith, Richard Elson, John P. Harris, Roland Salmon, Robert Smith, Sarah J. O'Brien, Goutam K. Adak, Claire Jenkins

We evaluated clinical Shiga toxin–producing *Escherichia coli* O157 infections in England and Wales during 1983–2012 to describe changes in microbiological and surveillance methods. A strain replacement event was captured; phage type (PT) 2 decreased to account for just 3% of cases by 2012, whereas PT8 and PT21/28 strains concurrently emerged, constituting almost two thirds of cases by 2012. Despite interventions to control and reduce transmission, incidence remained constant. However, sources of infection changed over time; outbreaks caused by contaminated meat and milk declined, suggesting that interventions aimed at reducing meat cross-contamination were effective. Petting farm and school and nursery outbreaks increased, suggesting the emergence of other modes of transmission and potentially contributing to the sustained incidence over time. Studies assessing interventions and consideration of policies and guidance should be undertaken to reduce Shiga toxin–producing *E. coli* O157 infections in England and Wales in line with the latest epidemiologic findings.

Shiga toxin–producing *Escherichia coli* (STEC) serogroup O157 emerged as a pathogen of public health concern during the early 1980s and was first isolated in the United Kingdom in July 1983 (Figure 1) from 3 cases linked to an outbreak of hemolytic uremic syndrome (HUS) (*1*). After this emergence, the Gastrointestinal Bacteria Reference Unit (GBRU), Public Health England (PHE) (then the Public Health Laboratory Service), reviewed a large archive of isolates and concluded that, before 1983, STEC O157 was not a major cause of gastrointestinal disease in England and Wales (*2*).

Author affiliations: Public Health England, London, UK (N.L. Adams, L. Byrne, G.A. Smith, R. Elson, J.P. Harris, G.K. Adak, C. Jenkins); National Institute for Health Research Health Protection Research Unit in Gastrointestinal Infections, Liverpool, UK (N.L. Adams, J.P. Harris, S.J. O'Brien, G.K. Adak); National Institute for Health Research Health Protection Research Unit in Gastrointestinal Infections, Colindale, UK (N.L. Adams, G.K. Adak, C. Jenkins); Public Health Wales, Cardiff, UK (R. Salmon, R. Smith)

DOI: http://dx.doi.org/10.3201/eid2204.151485

Despite being relatively rare in comparison with other gastrointestinal infections, STEC O157 is of public health concern because of its potential for severity. Symptoms of infection include abdominal cramps, nausea, and bloody diarrhea. In 5%–14% of cases, infection leads to HUS, a severe and potentially fatal systemic condition primarily affecting the kidneys (*3*). The primary STEC virulence factor is Shiga toxin (Stx), which targets cells expressing the glycolipid globotriaosylceramide, disrupting host protein synthesis and causing apoptotic cell death (*4*). Children and elderly persons are most susceptible to severe illness, and HUS is recognized as the most common cause of acute renal failure among children in the United Kingdom (*5*).

Cattle and other ruminants are natural reservoirs for STEC O157, and transmission to humans occurs through direct or indirect contact with the animals or their feces or through ingestion of contaminated food or water. A low infectious dose and propensity for person-to-person spread means transmission in households and closed settings such as schools is common (*6*,*7*), as is the potential for large outbreaks (*8*–*12*). We describe changes in the epidemiology of STEC O157 in England and Wales during a 30-year period (1983–2012) against a background of changing microbiological and surveillance methods over time.

Methods

Case Ascertainment, 1983–2012

Beginning in 1983, only fecal specimens from patients with HUS or hemorrhagic colitis were referred for STEC O157 testing; before 1989, few specimens were referred for testing. Beginning in 1997 in England and Wales, referral for testing was extended to all patients with symptoms of gastrointestinal infection, including vomiting, diarrhea, or bloody feces.

Microbiology Methods, 1983–2012

GBRU provides the national reference service for STEC in England and Wales. Beginning in 1983, individual colonies were tested for toxin production by using the verocytotoxin cell assay, and positive colonies were identified biochemically and serotyped (*13*). In 1987 at GBRU, the verocytotoxin

cell assay was replaced with a molecular probe assay to detect the *stx* gene (*14*) and sorbitol MacConkey culture medium. Later, sorbitol MacConkey culture medium containing cefixime and tellurite was developed, facilitating isolation of STEC O157 from fecal specimens, but testing was not implemented for all patients with symptoms of gastrointestinal infection in all local hospital laboratories until 1997 (*15*). Isolates of *E. coli* O157 identified locally are sent for confirmation and typing at GBRU.

Detection and confirmation of STEC at GBRU includes biochemical identification and serotyping of bacterial isolates. Since 1989, strains belonging to *E. coli* O157 have been further differentiated by using a phage typing scheme developed in Canada (*16*). Retrospective phage typing was undertaken for all viable strains collected before 1989. During 1994–2011, detection of *stx1* or *stx2* used a block-based PCR (*17*), which was replaced in 2012 with real-time PCR targeting *stx1* or *stx2* and the intimin (*eae*) gene, associated with intimate attachment of the bacteria to the host gut mucosa (*18*).

Data Collection Methods, 1983–2012

The amount of epidemiologic and microbiological data increased during the study period. During 1983–2003, a

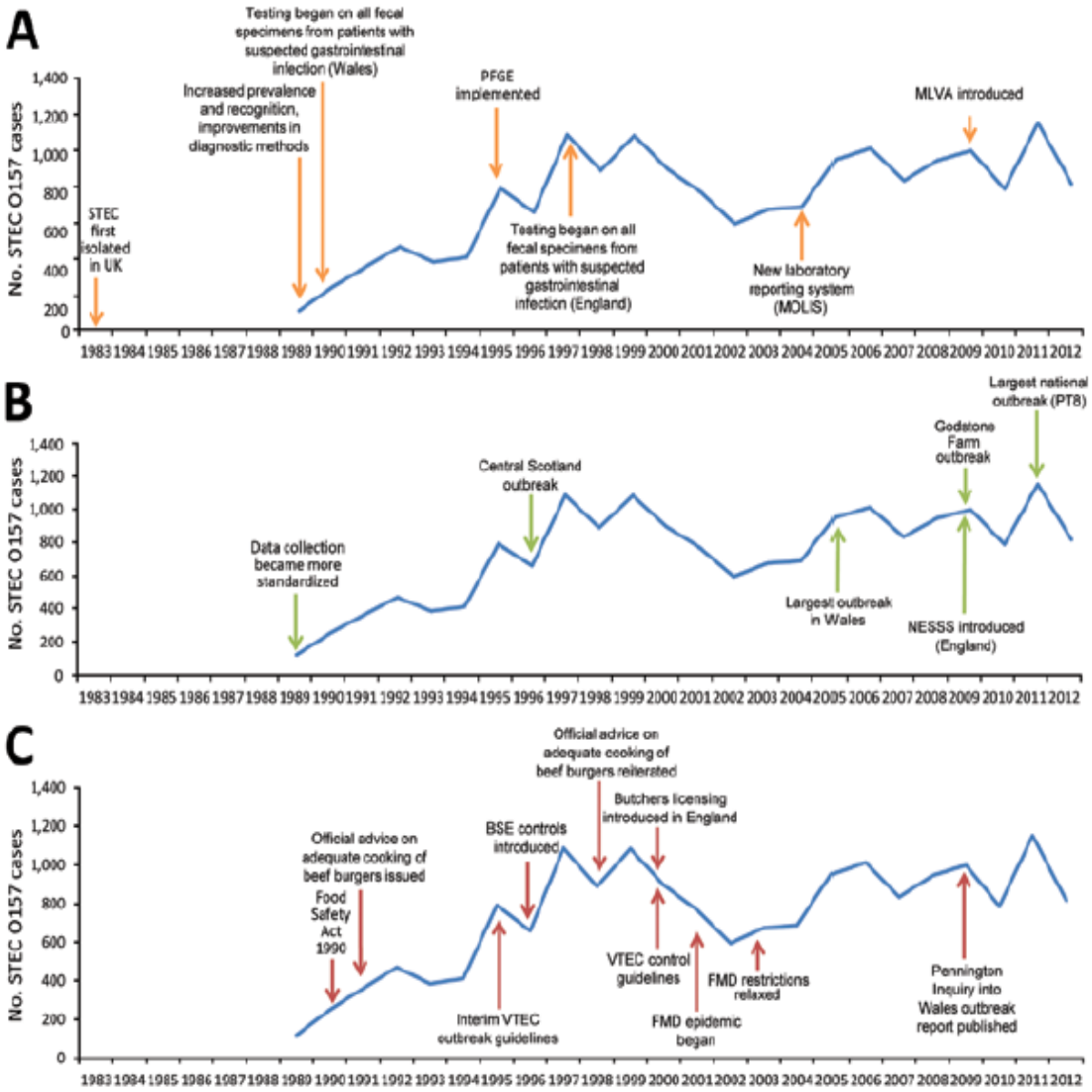


Figure 1. Timeline of key events influencing the epidemiology (A), microbiology (B), and guidance and control (C) of STEC O157, England and Wales, 1983–2012. Numbers before 1989 are available only as an aggregate for that period and therefore cannot be presented by year. BSE, bovine spongiform encephalopathy; FMD, foot and mouth disease; MLVA, multilocus variable-number tandem-repeat analysis; MOLIS, Modular Open Laboratory Information System; NESSS, National Enhanced Surveillance Scheme for STEC; PFGE, pulsed-field gel electrophoresis; PT, phage type; STEC, Shiga toxin–producing *Escherichia coli*; VTEC, verocytotoxin–producing *E. coli*.

dedicated laboratory database was used to record patient and microbiological data. In 2004, a new laboratory reporting system, Modular Open Laboratory Information System, was implemented (Figure 1). These laboratory databases captured microbiological results and demographic details of cases, as well as limited epidemiologic data (HUS diagnosis, outbreak association, recent history of foreign travel).

In January 2009, PHE introduced the National Enhanced Surveillance Scheme for STEC (Figure 1) (*19*). This scheme captured epidemiologic information through standardized questionnaires administered to all persons with STEC and linked to microbiological data in the Modular Open Laboratory Information System.

Detection of outbreaks relied on detecting unusual increases in STEC activity or reporting of shared exposures among cases of the same phage type (PT). Outbreaks were recorded on paper before 1992. In 1992, PHE began standardized surveillance of outbreaks of gastrointestinal disease where ≥2 persons with the same infection are linked, or probably linked, to the same source. In brief, local PHE units report standardized epidemiologic data on all outbreaks of gastrointestinal diseases, including source of infection and microbiological data.

Data Analyses

Data on STEC O157 patients in England and Wales were analyzed in 3 time periods, 1983–1988, 1989–1996, and 1997–2012, to account for periods of differing case ascertainment and data collection. Case numbers for 1989–1996 were small and represent biased sampling toward severe STEC O157 infections therefore calculation of incidence and interpretation of trends would be meaningless, and these were calculated only for 1997–2012.

We performed descriptive analyses in Microsoft Excel 2010 (Microsoft Corporation, Redmond, WA, USA). Crude incidence rates were calculated by using the Office of National Statistics mid-year population estimates (*20*). Crude incidence rate ratios (RR) and 95% CIs were calculated in Stata version 13.0 (StataCorp LP, College Station, TX, USA) for comparison among groups.

Results

Microbiology of STEC O157, 1983–2012

In England and Wales during 1983–1988, a total of 279 patients were infected with STEC O157, including 110 from 3 outbreaks. Of the 169 non–outbreak-related isolates, 155 were retrospectively phage typed; the most common types were PT2 (49 [31.6%] cases), PT1 (38 [24.5%]), and PT49 (22 [14.2%]).

During 1989–1996, the number of cases increased (3,448 total cases), and the proportions of common PTs changed annually (Figure 2). In 1996, a new PT was designated PT21/28 after reexamination of the lysis profiles of PT21 and PT28 isolates (*21*). By 1996, PT2 (244 [37%] isolates), PT8 (85 [12.9%] isolates), and PT21/28 (92 [13.9%] isolates) were the most common PTs, and the proportion of PT1 (28 [4.2%] isolates) and PT49 (42 [6.4%] isolates) had declined.

During 1997–2012, the decline in these PTs continued, and PT1 and PT49 were rarely observed. PT2 also declined to just 28 (3.4%) isolates by 2012 from a peak of 430 (54.3%) isolates in 1995, a significant decrease

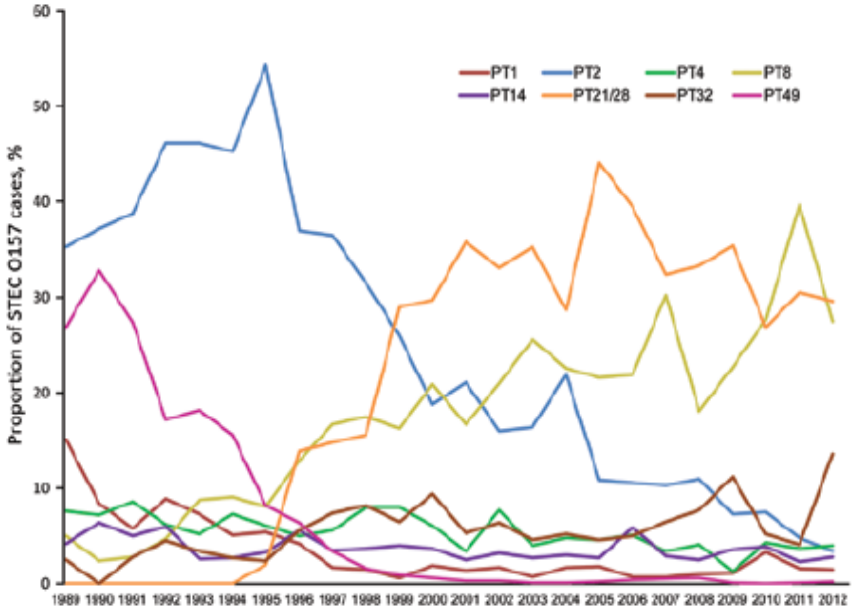


Figure 2. Proportions of common phage types (PTs) of Shiga toxin–producing *Escherichia coli* O157 identified, England and Wales, 1989–2012.

for this period (p<0.001) (Figure 2). Concurrently, numbers of PT21/28 rapidly increased, accounting for 420 (44.1%) cases by 2005, a significant increase for the period (p<0.001), and thereafter remaining the most frequently detected PT (Figure 2). PT8 also increased significantly, from 182 (16.7%) to 225 (27.5%) of cases by 2012 (p<0.001).

Strains encoding Stx1 only were rare (81 [0.6%] isolates), and most (45 [55.6%] isolates) were PT8. Strains encoding Stx2 only were most frequent (10,182 [71.8%] isolates), followed by Stx1+2 (3,921 [27.6%] isolates). Stx type and PT are interrelated; most PT8 strains (3,040 [93.2%] isolates) possessed *stx1+2*, whereas PT2 and PT21/28 usually possessed *stx2* only (2,100 [92.8%] and 4,340 [99.7%] isolates, respectively).

Epidemiology of STEC O157, 1997–2012

Case Numbers and Crude Incidence

A total of 14,184 laboratory-confirmed STEC O157 cases were identified in England and Wales; the mean was 887 (95% CI 802–972) cases per year. Crude incidence was 1.65 (95% CI 1.49–1.81) cases/100,000 person-years but varied by year, geography, and patient age and sex (Figure 3). Identifiable peaks in case numbers corresponded to reported outbreaks (Figure 1; Table, http://wwwnc.cdc.gov/EID/article/22/4/15-1485-T1.htm). Crude incidence decreased from 1999, reaching its lowest in 2002 (1.1 cases/100,000 person-years [595 cases]), but returned to previous levels in 2005 and was sustained thereafter.

STEC O157 infections demonstrated a distinct seasonality. Cases began to increase in April and declined beginning in September (data not shown).

Patient Age and Sex

Patient age was reported for 13,015 (91.8%) cases. Children <15 years of age constituted 5,867 (45.1%) cases; the greatest proportion (2,970 [22.8%] cases) occurred among those 1–4 years of age. Crude incidence decreased with increasing age; incidence was lowest for persons >60 years of age (0.98 [95% CI 0.82–1.12] cases/100,000 person-years) (Figure 3). Crude incidence was significantly higher for children 1–4 years of age (7.21 [95% CI 6.34–8.04] cases/100,000 person-years) than for those 20–59 years of age (RR 7.16, p<0.001) and >60 years of age (RR 7.36, p<0.001).

Sex was reported for 13,947 (98.3%) patients. Female patients accounted for 7,717 (55.3%) cases, and crude incidence was significantly higher for female than for male patients (RR 1.19, p<0.001; 1.76 [95% CI 1.59–1.93] cases/100,000 person-years , vs. 1.48 [95% CI 1.34–1.62] cases/100,000 person-years). Age and sex were reported for 12,848 (90.6%) patients. Sex disparity was highest for those 20–59 years of age (RR 1.60 for women vs. men; p<0.001).

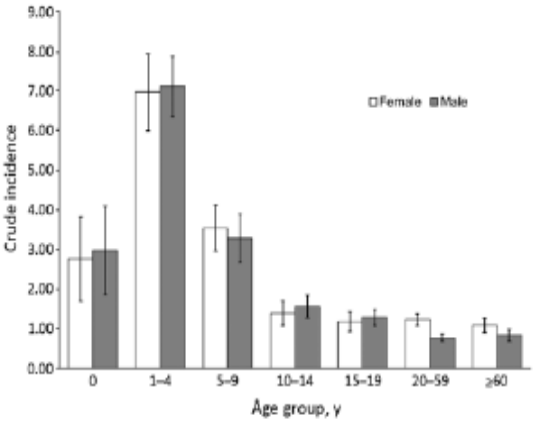


Figure 3. Crude incidence (cases per 100,000 person-years) of Shiga toxin–producing *Escherichia coli* O157, by patient age group and sex, England and Wales, 1997–2012. Error bars indicate 95% CIs.

The proportion of Stx2-only strains decreased with increasing age. Most (288 [81.8%]) children 1–4 years of age were infected with strains carrying Stx2 only, compared with 1,352 (65.4%) of persons ≥60 years of age (p<0.001). In parallel, the proportion of Stx1+2 profiles increased with age; 548 (16.1%) 1–4-year-olds were reported to have Stx1+2, compared with 698 (33.7%) of persons ≥60 years of age (p<0.001). We found no differences in sex by PT or Stx.

Geography

Annual crude incidence was highest in Cumbria and Lancashire (North West England) (3.70 [95% CI 2.70–4.70] cases/100,000 person-years), followed by Yorkshire and Humber (North East England, 2.75 [95% CI 2.37–3.13] cases/100,000 person-years) and Devon, Cornwall, and Somerset (South West England, 2.71 [95% CI 2.35–3.07] cases/100,000 person-years), whereas annual crude incidence was lowest in London (0.99 [95% CI 0.83–1.14] cases/100,000 person-years). Cases were almost 4 times more likely to be reported in Cumbria and Lancashire (RR 3.72) than in London (p<0.001). Within areas, crude incidence remained stable over time, other than peaks associated with outbreaks. We found no notable differences by geography in age, sex, PT, or seasonality.

Outbreaks, 1983–2012

During 1983–2012, a total of 335 outbreaks were reported, ranging from 0 to 25 outbreaks annually (Table). These outbreaks constituted 3,107 (17.4%) cases (median 5 cases, range 2–257 cases).

Large outbreaks caused peaks in annual crude incidence (Figure 1). For example, in 1995, eleven outbreaks comprising 141 cases occurred (Table), including a large nursery outbreak in Wales affecting 49 children (*6*). In

1999, nineteen outbreaks (236 cases) occurred, causing incidence to peak. Nine were attributed to contaminated food vehicles, including 3 caused by milk pasteurization failures, 1 affecting 88 persons (*22*). Outbreaks caused by postpasteurization contamination of milk also occurred in 2000 and 2002, as did 2 outbreaks associated with drinking raw milk in 2000, but no milk-related outbreaks were observed during the remainder of the study period.

Food vehicles contributed the highest number of outbreaks (101 [30.3%]) and outbreak cases (1,418 [45.9%]) (Table). These outbreaks included 38 attributed to eating contaminated meat; 16 to eating undercooked meat, such as burgers at barbecues; and 22 to cross-contamination of cooked meats. The cross-contamination outbreaks were larger; the largest meat-related outbreak occurred in Wales when meat from a butcher supplied to institutions infected 118 persons with STEC O157 in 2005 (*12*). After that, meat-related outbreaks were infrequent; 7 meat-related outbreaks (compared with 31 before this outbreak) were reported in the subsequent 7 years.

The first implicated food vehicle in this study was raw potatoes in a 1985 outbreak, and outbreaks associated with eating vegetables were reported throughout the years. The largest national outbreak in Great Britain (252 cases) caused by STEC O157 PT8, linked to handling raw leeks and potatoes, was reported in 2011 and led to the highest incidence during the period (Figure 1) (*23*).

Person-to-person spread in institutional settings accounted for 29.1% of outbreaks and more than one quarter of outbreak cases (825 cases). Twenty-six outbreaks occurred in institutional settings: care homes (16 outbreaks), prisons (4 outbreaks), and hospitals (6 outbreaks). No outbreaks in these settings occurred after 2007. Seventy-two outbreaks, which resulted in 808 cases, occurred in child-care facilities. Each year, 1–7 outbreaks in child-care facilities occurred, but outbreaks increased in frequency in later years; during 1983–2003, a total of 25 outbreaks (313 cases) were reported, whereas 47 outbreaks (495 cases) were reported in the subsequent 9 years.

Direct or indirect contact with animals through the environment accounted for 22.4% of outbreaks and 17.3% of patients linked to outbreaks. The number of petting farm outbreaks increased during the study period. During 1983–2002, a total of 12 petting farm outbreaks were reported; during 2003–2012, a total of 31 outbreaks on petting farms were reported, including, in September 2009, the largest reported farm outbreak, which affected 93 persons (*9*).

Most outbreaks were caused by the most frequently detected STEC O157 PTs, including PT21/28 (117 [34.9%] outbreaks), PT2 (79 [23.6%] outbreaks), and PT8 (42 [12.5%] outbreaks). In accordance with the general trends in PT, PT2 outbreaks declined over time, whereas PT8 and PT21/28 outbreaks increased. For outbreaks attributed to contact with animals or their environments, almost half (28 [47.5%] outbreaks) were caused by PT21/28 strains, a further 16 (27.1%) by PT2 strains, and only 4 (6.8%) by PT8 strains. Ten outbreaks attributed to contaminated water were caused by PT2 (5 outbreaks), PT21/28 (4 outbreaks), and PT4 (1 outbreaks); none were caused by PT8. In foodborne outbreaks, 25 (28.1%) were caused by PT2, 32 (36.1%) by PT21/28, and 20 (22.5%) by PT8.

Discussion

Our review provides a historical perspective contributing to the evidence of the evolving epidemiology of STEC O157. The data capture a strain replacement event showing the dramatic decline in PT2 and the increase and dominance of PT8 and PT21/28. Outbreak settings and vehicles also changed during the study period; prison, hospital, and care-home outbreaks decreased, and outbreaks in childcare facilities increased. Additionally, outbreaks associated with meat and milk decreased, and outbreaks attributed to petting farms increased. These data support previous reports that PT21/28 is indigenous to Great Britain and PT8 is largely imported, because most PT8 outbreaks were foodborne and a greater proportion of PT21/28 were attributed to environmental or animal contact (*19*,*21*).

The reasons for the decline in STEC incidence during 2000–2004 are unknown and cannot be attributed to any particular event or intervention, although several possible explanations exist. After a large STEC O157 outbreak in central Scotland in 1996 (*8*), specific interventions were implemented throughout the entire United Kingdom in catering, retail, and meat hygiene sectors to reduce the risk for infection. These included butchers' licensing, legislation, and enforcement of Hazard Analysis and Critical Control Point systems; amendment of the Food Standards Agency Code of Practice; and introduction of the Clean Livestock Policy, which aimed to reduce contamination by feces or mud on the coats and fleeces of animals for slaughter (*24*). The effectiveness of these policies was apparent through the shift in causes of outbreaks presented in this study; after their implementation, outbreaks caused by cross-contamination from raw meat clearly declined.

Why the decline in STEC incidence was not sustained beyond 2004 is unclear; however, declining numbers were observed in the United States in 2003 and 2004, followed by increases beginning in 2005. The decline coincided with industry measures aimed at reducing contamination of ground beef; however, as in the United Kingdom, the reason for the subsequent increase is unknown (*25*). Apparent changes in sources and outbreak settings might indicate changes in food vehicles or transmission routes among all cases, and although earlier interventions successfully controlled transmission of STEC infection, other effective transmission routes have taken hold in more recent years.

Also, in this study, outbreak detection relied on the classic epidemiologic triad of person, place, and time, along with PT. Any outbreaks dispersed over time, or of a common PT, might have gone undetected. Data collected on outbreaks and sources—and therefore trends—will be incomplete.

Farming methods and destruction of animal populations changed considerably during the study period after concerns about bovine spongiform encephalopathy in 1996 and foot and mouth disease in 2001. The decline of PT1, PT2, and PT49, and the corresponding emergence of PT21/28, was mirrored in Scotland (*26*) and suggests a strain replacement event. The destruction and restocking of UK cattle herds after concerns about bovine spongiform encephalopathy and foot and mouth disease might have been a causative factor. In Ireland, PT32 is the most commonly reported PT (*27*); PT21/28 is rarely detected outside the British Isles (*19*).

Improvements in data collection during our study led to increased ascertainment of epidemiologic data during the 30-year period alongside important developments in microbiological methods. Thus, the sustained incidence of infection could be a surveillance artifact, masking the success of interventions through increasing case ascertainment, a potential bias when datasets spanning many years, such as this one, are analyzed. In England and Wales, although surveillance of clinical STEC infections is routine, no surveillance programs are ongoing to monitor the prevalence of STEC in cattle or other animals. Efforts by the agricultural, veterinary, and food industries to monitor STEC incidence and strain types would inform the success of interventions and provide insight into the ecology of the pathogen. However, STEC rarely causes disease in animals, and funding is limited for such programs in England and Wales. In Europe, current monitoring information is generated from outbreak investigations and ad hoc studies skewed toward foodborne transmission of STEC O157 and might be limited in assessing the role of environmental transmission.

As described previously, infection is highest in children and females (*5*,*19*,*28*). Children 1–4 years of age had 7 times the risk of persons ≥60 years of age, probably because of a complex interplay of various factors, such as host immunity or reporting artifacts, with children more likely to seek care at healthcare settings (*29*). Additionally, the propensity for household transmission of STEC O157 (*30*) might be exacerbated by children having poorer hygiene practices that increase exposure to STEC O157 from the environment (*19*), and the potential for prolonged excretion in children (*7*). Children were more often infected with STEC O157 Stx2-only strains, associated with more severe disease (*4*,*5*,*21*,*31*), which might in part explain why cases occurred more often in children, because they were more likely to require care at healthcare settings. The higher crude incidence rates for female than for male patients has been reported previously (*19*,*28*); the reasons are unknown but might reflect biologic host factors, differences in health-seeking behavior, or other behaviors placing women at increased risk for infection, such as having contact with children or being primary household food handlers (*19*,*32*).

In our review, crude incidence was higher in the north and southwest than in the central and southeastern areas of England. Crude incidence in Scotland is consistently higher still (*19*). Previous studies have described such geographic variation and demonstrated that differences reflect differences in weather, land use, or environmental exposure between persons living in or visiting rural areas and those in urban areas, fitting with environmental transmission of STEC O157 (*19*,*33*).

Our 30-year review captures the emergence of a clinically significant zoonotic pathogen in a well-characterized population sample and documents the effectiveness of improvements in epidemiologic and microbiological methods on ascertaining STEC O157. However, despite interventions that successfully shifted outbreak settings, these organisms persist in causing illness in England and Wales, and the crude incidence of STEC O157 has remained relatively stable. Robust studies are required to assess the effectiveness of interventions, which currently remain unclear, and to consider future policies and guidance to reduce STEC O157 infection in England and Wales in the context of the complex interaction between the organism, reservoir, food chain, and transmission pathway.

Acknowledgments

We thank Marie Anne Chattaway, Vivienne DoNascimento, and Neil Perry for their microbiology expertise. We also acknowledge the roles of Bernard Rowe, Brian Jiggle, Tom Cheasty, Sylvia Scotland, Andrea Thomas, Jude Evans, and especially the late Henry Smith for their major contribution to the field. We thank Naomi Launders, Kirsten Glen, and Dilys Morgan for their contribution to STEC epidemiology. Finally, we are grateful to the microbiologists and health protection and environmental health specialists who have contributed data and reports to national surveillance systems and the epidemiologists and information officers who have worked on the national surveillance of intestinal infectious diseases for Centre for Infectious Disease Surveillance and Control and Health Protection Services Colindale.

This work was supported by the National Institute for Health Research Health Protection Research Unit in Gastrointestinal Infections.

This research was funded by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Gastrointestinal Infections at the University of Liverpool in partnership with Public Health England (PHE), University of

East Anglia, University of Oxford, and the Institute of Food Research. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health, or Public Health England.

Miss Adams is an epidemiologist at Public Health England in London and is working on a PhD thesis on health inequalities in gastrointestinal infections at the University of Liverpool. Her research interests include gastrointestinal disease surveillance.

References

1. Taylor CM, White RH, Winterborn MH, Rowe B. Haemolytic-uraemic syndrome: clinical experience of an outbreak in the West Midlands. BMJ. 1986;292:1513–6. http://dx.doi.org/10.1136/bmj.292.6534.1513
2. Day NP, Scotland SM, Cheasty T, Rowe B. *Escherichia coli* O157:H7 associated with human infections in the United Kingdom. Lancet. 1983;321:825. http://dx.doi.org/10.1016/S0140-6736(83)91887-1
3. Tarr PI, Gordon CA, Chandler WL. Shiga-toxin–producing *Escherichia coli* and haemolytic uraemic syndrome. Lancet. 2005;365:1073–86.
4. Ethelberg S, Olsen KE, Scheutz F, Jensen C, Schiellerup P, Enberg J, et al. Virulence factors for hemolytic uremic syndrome, Denmark. Emerg Infect Dis. 2004;10:842–7.
5. Lynn RM, O'Brien SJ, Taylor CM, Adak GK, Chart H, Cheasty T, et al. Childhood hemolytic uremic syndrome, United Kingdom and Ireland. Emerg Infect Dis. 2005;11:590–6. http://dx.doi.org/10.3201/eid1104.040833
6. Al-Jader L, Salmon RL, Walker AM, Williams HM, Willshaw GA, Cheasty T. Outbreak of *Escherichia coli* O157 in a nursery: lessons for prevention. Arch Dis Child. 1999;81:60–3. http://dx.doi.org/10.1136/adc.81.1.60
7. Swerdlow DL, Griffin PM. Duration of faecal shedding of *Escherichia coli* O157:H7 among children in day-care centres. Lancet. 1997;349:745–6. http://dx.doi.org/10.1016/S0140-6736(05)60196-1
8. Cowden JM, Ahmed S, Donaghy M, Riley A. Epidemiological investigation of the central Scotland outbreak of *Escherichia coli* O157 infection, November to December 1996. Epidemiol Infect. 2001;126:335–41. http://dx.doi.org/10.1017/S0950268801005520
9. Ihekweazu C, Carroll K, Adak B, Smith G, Pritchard GC, Gillespie IA, et al. Large outbreak of verocytotoxin-producing *Escherichia coli* O157 infection in visitors to a petting farm in South East England, 2009. Epidemiol Infect. 2012;140:1400–13. http://dx.doi.org/10.1017/S0950268811002111
10. Michino H, Araki K, Minami S, Takaya S, Sakai N, Miyazaki M, et al. Massive outbreak of *Escherichia coli* O157:H7 infection in schoolchildren in Sakai City, Japan, associated with consumption of white radish sprouts. Am J Epidemiol. 1999;150:787–96. http://dx.doi.org/10.1093/oxfordjournals.aje.a010082
11. Public Health Agency of Canada. Waterborne outbreak of gastroenteritis associated with a contaminated municipal water supply, Walkerton, Ontario, May–June 2000. Can Commun Dis Rep. 2000;26:170–3.
12. Pennington TH. The public inquiry into the September 2005 outbreak of *E. coli* O157 in South Wales. Aberdeen: HMSO; 2009 [cited 2015 Jul 15]. http://gov.wales/docs/dhss/publications/150618ecoli-reporten.pdf
13. Scotland SM, Day NP, Rowe B. Production of a cytotoxin affectin Vero cells by strain of *Escherichia coli* belonging to the traditional entreopathogenic serogroups. FEMS Microbiol Lett. 1980;7:15–7. http://dx.doi.org/10.1111/j.1574-6941.1980.tb01567.x
14. Willshaw GA, Smith HR, Scotland SM, Field AM, Rowe B. Heterogeneity of *Escherichia coli* phages encoding Vero cytotoxins: comparison of cloned sequences determining VT1 and VT2 and development of specific gene probes. J Gen Microbiol. 1987;133:1309–17.
15. Smith HR, Rowe B, Gross RJ, Fry NK, Scotland SM. Haemorrhagic colitis and Vero-cytotoxin–producing *Escherichia coli* in England and Wales. Lancet. 1987;1:1062–5. http://dx.doi.org/10.1016/S0140-6736(87)90485-5
16. Khakhria R, Duck D, Lior H. Extended phage-typing scheme for *Escherichia coli* O157:H7. Epidemiol Infect. 1990;105:511–20. http://dx.doi.org/10.1017/S0950268800048135
17. Thomas A, Jiggle B, Smith HR, Rowe B. The detection of Vero cytotoxin–producing *Escherichia coli* and *Shigella dysenteriae* type 1 in faecal specimens using polymerase chain reaction gene amplification. Lett Appl Microbiol. 1994;19:406–9. http://dx.doi.org/10.1111/j.1472-765X.1994.tb00968.x
18. Jenkins C, Lawson AJ, Cheasty T, Willshaw GA. Assessment of a real-time PCR for the detection and characterization of verocytotoxigenic *Escherichia coli*. J Med Microbiol. 2012;61:1082–5. http://dx.doi.org/10.1099/jmm.0.041517-0
19. Byrne L, Jenkins C, Launders N, Elson R, Adak GK. The epidemiology, microbiology and clinical impact of Shiga toxin–producing *Escherichia coli* in England, 2009–2012. Epidemiol Infect. 2015;143:3475–87. http://dx.doi.org/10.1017/S0950268815000746
20. Office for National Statistics Mid-1971 to mid-2012 population estimates: England and Wales; quinary age group; estimated resident population. 2013 [cited 2014 Mar 17]. http://www.ons.gov.uk/ons/rel/pop-estimate/population-estimates-for-uk--england-and-wales--scotland-and-northern-ireland/population-estimates-timeseries-1971-to-current-year/index.html
21. Dallman TJ, Ashton PM, Byrne L, Perry NT, Petrovska L, Ellis R, et al. Applying phylogenomics to understand the emergence of Shiga-toxin-producing Escherichia coli O157:H7 strains causing severe human disease in the UK. Microbial Genomics. 2015;1(3). http://dx.doi.org/10.1099/mgen.0.000029
22. Goh S, Newman C, Knowles M, Bolton FJ, Hollyoak V, Richards S, et al. *E. coli* O157 phage type 21/28 outbreak in North Cumbria associated with pasteurized milk. Epidemiol Infect. 2002;129:451–7. http://dx.doi.org/10.1017/S0950268802007835
23. Launders N, Locking ME, Hanson M, Willshaw G, Charlett A, Salmon R, et al. A large Great Britain–wide outbreak of STEC O157 phage type 8 linked to handling of raw leeks and potatoes. Epidemiol Infect. 2015;144:171–81.
24. Pennington TH. *E. coli* O157 outbreaks in the United Kingdom: past, present, and future. Infect Drug Resist. 2014;7:211–22. http://dx.doi.org/10.2147/IDR.S49081
25. Centers for Disease Control and Prevention. Preliminary FoodNet data on the incidence of infection with pathogens transmitted commonly through food—10 states, 2006. MMWR Morb Mortal Wkly Rep. 2007;56:336–9.
26. Pearce MC, Chase-Topping ME, McKendrick IJ, Mellor DJ, Locking ME, Allison L, et al. Temporal and spatial patterns of bovine *Escherichia coli* O157 prevalence and comparison of temporal changes in the patterns of phage types associated with bovine shedding and human *E. coli* O157 cases in Scotland between 1998–2000 and 2002–2004. BMC Microbiol. 2009;9:276. http://dx.doi.org/10.1186/1471-2180-9-276
27. Carroll AM, Gibson A, McNamara EB. Laboratory-based surveillance of human verocytotoxigenic *Escherichia coli* infection in the Republic of Ireland, 2002–2004. J Med Microbiol. 2005;54:1163–9. http://dx.doi.org/10.1099/jmm.0.46147-0
28. Vally H, Hall G, Dyda A, Raupach J, Knope K, Combs B, et al. Epidemiology of Shiga toxin producing *Escherichia coli* in Australia, 2000–2010. BMC Public Health. 2012;12:63. http://dx.doi.org/10.1186/1471-2458-12-63
29. Wheeler JG, Sethi D, Cowden JM, Wall PG, Rodrigues LC, Tompkins DS, et al. Study of infections intestinal disease in

England: rates in the community, presenting to general practice and reported to national surveillance. BMJ. 1999;318:1046–50. http://dx.doi.org/10.1136/bmj.318.7190.1046

30. Parry SM, Salmon RL, Willshaw GA, Cheasty T. Risk factors for and prevention of sporadic infections with vero cytotoxin (Shiga toxin) producing *Escherichia coli* O157. Lancet. 1998;351:1019–22. http://dx.doi.org/10.1016/S0140-6736(97)08376-1
31. Milford DV, Taylor CM, Guttridge B, Hall SM, Rowe B, Kleanthous H. Haemolytic uraemic syndromes in the British Isles 1985–8: association with verocytotoxin producing *Escherichia coli*. Part 1: clinical and epidemiological aspects. Arch Dis Child. 1990;65:716–21. http://dx.doi.org/10.1136/adc.65.7.716
32. Lake AA, Hyland RM, Mathers JC, Rugg-Gunn AJ, Wood CE, Adamson AJ. Food shopping and preparation among the 30-somethings: whose job is it? (The ASH30 study). Br Food J. 2006;108:475–86. http://dx.doi.org/10.1108/00070700610668441
33. Innocent GT, Mellor DJ, McEwen SA, Reilly WJ, Smallwood J, Locking ME, et al. Spatial and temporal epidemiology of sporadic human cases of *Escherichia coli* O157 in Scotland, 1996–1999. Epidemiol Infect. 2005;133:1033–41. http://dx.doi.org/10.1017/S0950268805003687

Address for correspondence: Natalie L. Adams, Gastrointestinal Infections Department, National Infection Service, Public Health England, 61 Colindale Ave, Colindale NW9 5EQ, UK; email: natalie.adams@phe.gov.uk

Paper 3: Evaluating the use of multilocus variable number tandem repeat analysis of Shiga toxin-producing *Escherichia coli* O157 as a routine public health tool in England.

OPEN ACCESS Freely available online

PLOS ONE

Evaluating the Use of Multilocus Variable Number Tandem Repeat Analysis of Shiga Toxin-Producing *Escherichia coli* O157 as a Routine Public Health Tool in England

Lisa Byrne[1]*, Richard Elson[1], Timothy J. Dallman[2], Neil Perry[2], Philip Ashton[2], John Wain[3], Goutam K. Adak[1], Kathie A. Grant[2], Claire Jenkins[2]

1 Gastrointestinal Emerging and Zoonotic Infections Department, Health Protection Services, Public Health England, Centre for Infectious Disease Surveillance and Control, London, United Kingdom, **2** Gastrointestinal Bacteria Reference Unit, Microbiology Services, GBRU, Public Health England, Centre for Infectious Disease Surveillance and Control, London, United Kingdom, **3** Faculty of Medicine and Health Services, Norwich Medical School, University of East Anglia, Norwich, United Kingdom

Abstract

Multilocus variable number tandem repeat analysis (MLVA) provides microbiological support for investigations of clusters of cases of infection with Shiga toxin-producing *E. coli* (STEC) O157. All confirmed STEC O157 isolated in England and submitted to the Gastrointestinal Bacteria Reference Unit (GBRU) during a six month period were typed using MLVA, with the aim of assessing the impact of this approach on epidemiological investigations. Of 539 cases investigated, 341 (76%) had unique (>2 single locus variants) MLVA profiles, 12% of profiles occurred more than once due to known household transmission and 12% of profiles occurred as part of 41 clusters, 21 of which were previously identified through routine public health investigation of cases. The remaining 20 clusters were not previously detected and STEC enhanced surveillance data for associated cases were retrospectively reviewed for epidemiological links including shared exposures, geography and/or time. Additional evidence of a link between cases was found in twelve clusters. Compared to phage typing, the number of sporadic cases was reduced from 69% to 41% and the diversity index for MLVA was 0.996 versus 0.782 for phage typing. Using MLVA generates more data on the spatial and temporal dispersion of cases, better defining the epidemiology of STEC infection than phage typing. The increased detection of clusters through MLVA typing highlights the challenges to health protection practices, providing a forerunner to the advent of whole genome sequencing as a diagnostic tool.

Citation: Byrne L, Elson R, Dallman TJ, Perry N, Ashton P, et al. (2014) Evaluating the Use of Multilocus Variable Number Tandem Repeat Analysis of Shiga Toxin-Producing *Escherichia coli* O157 as a Routine Public Health Tool in England. PLoS ONE 9(1): e85901. doi:10.1371/journal.pone.0085901

Editor: A. Mark Ibekwe, U. S. Salinity Lab, United States of America

Received August 1, 2013; **Accepted** December 6, 2013; **Published** January 17, 2014

Funding: The MLVA component of this study, and the authors PA and NP were funded by the Public Health England Strategic Research and Development Fund: Grant #108061. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: lisa.byrne@phe.gov.uk

Introduction

Shiga toxin-producing *Escherichia coli* (STEC) are associated with human illness and are defined by the presence of the phage-encoded Shiga toxin genes, *stx1* and/or *stx2*. Symptoms of STEC infection range from mild gastroenteritis through to severe bloody diarrhoea and approximately 6% of cases develop haemolytic uraemic syndrome (HUS);[1] HUS is a serious condition where shiga toxins affect the blood and kidneys. It most frequently affects children and is recognised as the most common cause of kidney failure in children.

The main reservoir of STEC in England is cattle although it is carried by other animals, mainly ruminants.[2] Transmission to humans occurs through direct or indirect contact with animals or their environment, consumption of contaminated food or water and through person-to-person contact.[3–6] STEC may cause both sporadic and epidemic infections and several large outbreaks have been recorded.[7–9]Traditionally, outbreaks of STEC are identified through routine investigation of cases, either through identifying common features between cases in terms of exposures, the appearance of microbiological subtypes (i.e. serogroups and phage types) among cases which are temporally or geographically linked, or indications that the number of cases in a particular location or of a particular subtype is higher than expected. Between 2009 and 2012, there were 67 reported outbreaks, affecting 737 individuals in England and virtually all of these were identified following investigation either locally by the Public Health England Centre (PHEC) or nationally by the Department of Gastrointestinal, Emerging and Zoonotic Infections (GEZI) (unpublished data). The remaining 2,982 cases not attributed to outbreaks were either sporadic cases or due to person-to-person transmission within households.

Each year, over 1000 isolates of presumptive *E. coli* O157 (the most frequently detected serogroup in England) are submitted to the Gastrointestinal Bacteria Reference Unit (GBRU), (the national reference laboratory for gastrointestinal pathogens in

England) and on average approximately 900 are confirmed as STEC O157. All confirmed STEC O157 isolates are phage typed,[10] and the most frequently reported phage types each year are PT 21/28 and PT 8. Standard Enhanced Surveillance Questionnaires (ESQ) data are routinely collected from presumptive cases of STEC, often before isolates are confirmed as STEC by the GBRU. These data are scrutinised and identification of common features between cases is most often the first indication of an outbreak. Where enhanced surveillance data identifies cases which may be linked, strains are selected for additional molecular typing by multilocus variable number tandem repeat (VNTR) analysis (MLVA).[11] While routine phage typing provides a rapid, robust and cost effective screen, MLVA, provides a more discriminatory, portable typing technique. MLVA thus provides more sensitive microbiological support for epidemiological links and, since implementation of its selective use in 2006, has made a valuable contribution to outbreak investigations in England and elsewhere. [12;13]However, MLVA typing of all isolates has resource implications because of the high numbers of isolates received by GBRU during peak periods.

From the 1st May 2012, all confirmed STEC O157 isolates submitted to GBRU were typed using MLVA, with the aim of evaluating this typing approach for prospective outbreak identification and assessing how it would impact on epidemiological investigations. Data and findings from the first six months of this strategy are presented herein.

Materials and Methods

Microbiological characterisation

Detection and confirmation of STEC at GBRU included biochemical identification and serotyping of bacterial isolates. Real time polymerase chain reaction (PCR)was performed to determine the presence of Shiga toxin genes *stx1* and/or *stx2*.[14] Strains belonging to serogroup O157 were further differentiated by phage typing.[10]

From 1st May 2012, all STEC O157 strains submitted to GBRU were typed using MLVA. Amplification of eight MLVA loci was performed in 20 ul reactions volumes on all isolates in two quadruplex PCR reactions. [11;15] VNTR-10 locus was not tested. The method was modified to use the following dye labels in the forward primers in the amplifications: NED in VNTR3 and VNTR17; 6-FAM in VNTR34, VNTR19, VNTR9 and VNTR36 and VIC in VNTR25 and VNTR37. Sizing of the amplified products was on an ABI 3730 Genetic Analyzer with 600 LIZ (Applied Biosystems) as size standard and data were analysed with Peakscanner software. Fragment sizes were imported into Bionumerics software via an algorithm that calculated the tandem repeat numbers for each locus. A minimum spanning tree was constructed to visually compare MLVA profiles with phage types.

Enhanced surveillance of STEC in England

Local laboratories report presumptive isolates of STEC directly to PHE Centres, who arrange for an ESQ to be administered to patients either by local health protection practitioners or environmental health professionals. The ESQ collects demographic details; risk status; clinical condition (including progression to HUS); household or other close contact details; laboratory results; exposures including travel, food and water consumption, contact with animals and environmental factors; epidemiological case classification; and outbreak/cluster status. Completed questionnaires are forwarded for inclusion in the National Enhanced Surveillance System for STEC in England (NESSS) which is managed by the PHE GEZI department.

Case categorisation and comparison

Case Definitions. Domestic case: An STEC case with no travel outside of the UK reported in the seven days prior to onset of illness.

Travel case: An STEC case with travel outside of the UK reported in the seven days prior to onset of illness.

Household case: An STEC case epidemiologically and microbiologically linked to one or more cases in the same household.

Cluster case: An STEC case epidemiologically and microbiologically linked to one or more case outside of the same household.

Sporadic case: An STEC case with no known links to any other STEC cases.

Case definitions were applied first on the basis of ESQ data and results of standard microbiological methods (i.e. phage typing/*stx* profiles). Cases were re-categorised on the basis of MLVA results. Comparisons were made between case categories based on phage typing and MLVA. The discriminatory ability of both phage typing and MLVA typing were calculated using Simpson's index of diversity as previously described.[16]

Categorisation and evaluation of clusters. Clusters were grouped into three categories:

(i) Household clusters
(ii) Known Clusters- detected through routine public health follow-up of STEC cases identifying linked cases
(iii) MLVA identified clusters – newly identified through MLVA typing

For clusters newly identified through MLVA, the ESQ's for cases were retrospectively reviewed for epidemiological links including:

(i) Temporal linkage: Cases with onset of illness occurring within seven days of each other.
(ii) Exposure linkage: Cases reporting shared exposures, including geographical links (all cases resided or visited a location within one PHE region), as reported on the ESQ.

Results

Case categorisation and comparison

Between 1st May and 31st October 2012, 556 confirmed cases of STEC O157 in England were reported and 16 different PT's were identified amongst cases; the most frequently reported PT overall was PT 8 (n = 189, 33.9%), followed by PT 21/28 (n = 160, 28.8%). STEC O157 isolates from 539 cases (96.9%) were further typed by MLVA and a total of 341 unique (>2 SLV's) MLVA profiles were identified. Over three-quarters (n = 258) of profiles were unique to one case. Of the remaining profiles (24%), half (n = 41) occurred more than once due to known household transmission and half (n = 42) occurred more than once outside of households (i.e. community clusters). The concurrence between MLVA profiles and PT was high, with just three instances where the same MLVA profile covered more than one PT (figure 1). Simpson's diversity index for the phage typed isolates was 0.782, and for MLVA typing was 0.996, indicating the increased discriminatory power of this method.

ESQ data indicated a history of foreign travel for 150 cases (27.8%), including 20 cases comprising eight household clusters and 42 cases linked to 13 travel associated clusters (Figure 2a).

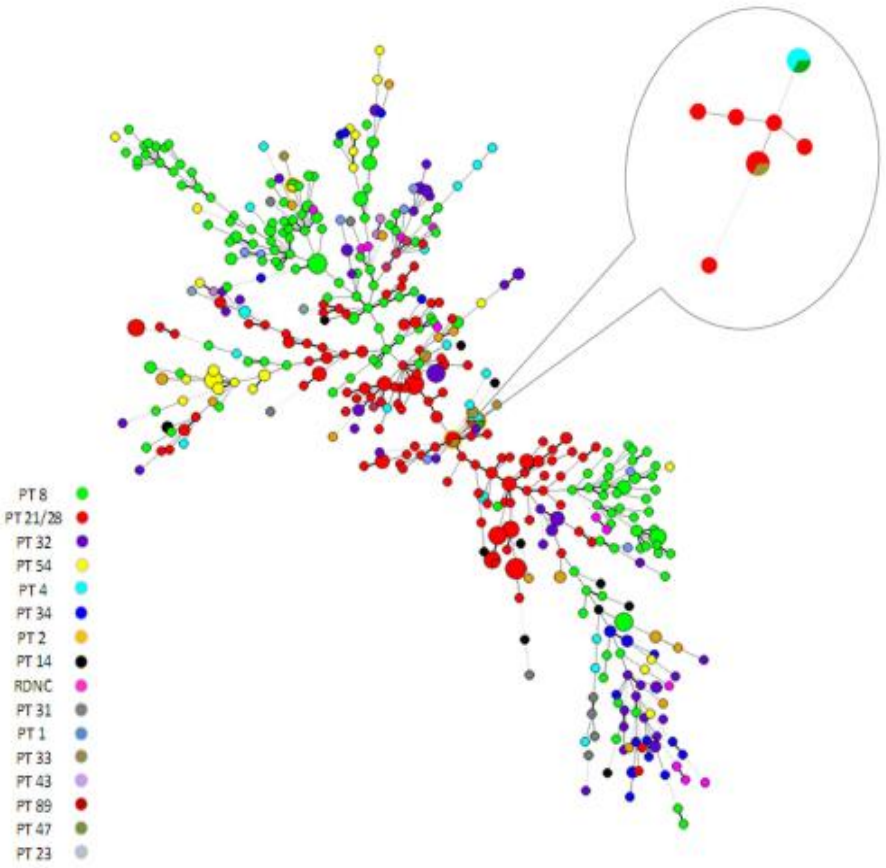


Figure 1. Minimum spanning tree of STEC O157 isolates MLVA profiles categorised by phage type[1]. 1. Includes 539 confirmed cases of STEC O157 in England. The size of nodes is proportional to the number of instances of that unique profile. Join lines represent locus variants: Related SLV's and DLV's are represented through a solid line while >two locus variants are denoted through a dotted join.In three clusters MLVA profiles spanned two phage types, these are highlighted in the figure. Inset presents two of these instances where the same MLVA profile was reported in two different phage types.
doi:10.1371/journal.pone.0085901.g001

Among 389 domestic cases, eight community clusters comprising 37 (8.9%) cases and 34 household clusters comprising 86 (22.1%) cases were reported. The remaining 267 cases (68.6%) were not linked to any known clusters and were classed as sporadic through ESQ data and phage typing. However, unique MLVA profiles were reported for less than half of domestic cases (42.9%, n = 167) and 101 were re-categorised as community cluster cases (Figure 2b). MLVA typing indicated 28 separate non-travel community clusters: 20 newly identified through MLVA typing. Compared to phage typing, the proportion of domestic sporadic cases was thus reduced by 27.8% through MLVA.

Cluster categorisation and evaluation

(i) Household clusters. A fifth (n = 106) of all cases comprised 42 different household clusters, eight of which were travel-related. Household clusters ranged between two and seven cases in size, with a median of 2 cases (IQR: 2, 3). Of the 42 clusters, in 24 households all cases had the same MLVA profile, 11 exhibited Single locus variants (SLVs), five double locus variants (DLVs) and two had >2 locus variants.

(ii) Known Clusters. Twenty-one of the 41 community clusters defined through MLVA typing were known clusters, previously identified as epidemiologically linked through surveillance data from ESQ's and phage typing. This included 13 clusters of cases associated with travel to Malta (3), Turkey (3), Cape Verde (2), Tunisia (1), Morocco (1), Egypt (1), Spain (1) and Israel (1). Eight non-travel community clusters comprising 37 cases (Table 1) were detected through routine follow-up of cases and were investigated. Exposures included attendance at nurseries and schools (4), recent consumption of minced beef products outside the home (1), contact with ruminants in a park or farm (2) and attendance at a wedding (no food identified) (1). All eight clusters comprised cases which were temporally and geographically linked (all cases resided or visited a location within one PHEC).Inclusion of cases was informed through MLVA typing in all eight clusters. One cluster of PT 21/28 *stx2* was first identified when three cases reported consuming beef burgers at the same food outlet. Investigations initially focused on the food outlet, however, six additional cases not linked to this food outlet were identified by GBRU as having the same MLVA profile and investigations were widened. Raw minced beef contaminated with STEC O157 during the supply chain was suspected as the source of infection.

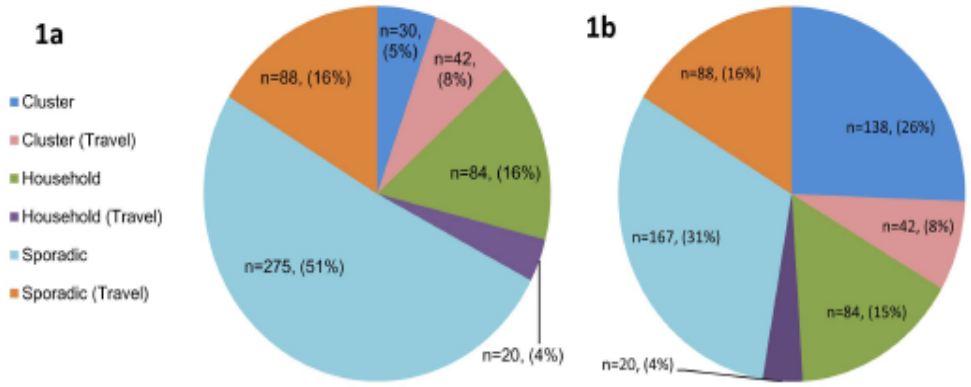


Figure 2. a) Case classification of STEC O157 cases categorised through VESQ data and phage typing, and 2b) Case classification of STEC O157 cases categorised through VESQ data and MLVA typing: May-October 2012[1,2]. 1.Includes 539/556 confirmed cases who had isolated typed by MLVA. 2.Travel cases reported travel outside of the UK in the seven days prior to onset of illness
doi:10.1371/journal.pone.0085901.g002

MLVA did not identify additional cases in the six other clusters. However, in Cluster #3 (Nursery cluster A), MLVA indicated that despite different PT's (figure 1), the cases were linked and that the cluster was seeded by a case who was part of a separate cluster associated with visiting a farm (Cluster #6). The outbreak associated with exposure to animals in a country park comprised strains with MLVA profiles that had DLVs, whereas the other six outbreaks were comprised of strains that shared identical profiles or were SLVs.

(iii) MLVA identified clusters. An additional 20 community clusters not detected through ESQ data and phage typing were identified through MLVA. These clusters ranged in size between two and 12 cases, with a median of three cases (IQR: 3, 7). Eighteen of the 20 clusters (92 cases) had associated strains which were either PT 21/28 *stx*2 or PT8 *stx*1+2. ESQ data for cases from the clusters were reviewed for links between cases, and epidemiological links between cases were identified in twelve clusters (Table 2).

Clusters with temporal and exposures linkage. Four MLVA clusters had cases both occurring close together in time and additional evidence of linkage between cases. Definitive sources of infection were not identified but cases were linked geographically and/or through exposures: Three of the clusters consisted of cases from different English regions who had travelled to the same areas on holiday and some had visited the same attractions or undertaken the same activities. The fourth cluster was of two cases who lived in the same town. One cluster of two in a household comprised cases with an identical MLVA profile but one case had a PT54 strain and the other PT8.

One MLVA cluster of seven cases comprised cases from different regions in England. Reviewing the ESQ data revealed that two cases had visited the same village, where another case was resident. Four cases were re-interviewed and all reported visiting the same public house. There were no common consumption patterns between the cases, and two cases had visited the premises outside of the seven day exposure period. The potential source could have been a person or the environment or it may that contamination was taken into the public house. Three other cases did not report travel to the same location and no obvious shared exposures were reported on the ESQ. Because of the time that had elapsed between notification and onset no further follow-up of these cases was undertaken.

Clusters with temporal linkage only. Eight MLVA clusters consisted of 45 cases with onset dates close together in time (maximum seven days between cases) but no common geography or exposures were reported on the ESQ's. Five clusters

Table 1. Summary of community clusters identified through routine follow-up of cases: May- October 2012.

Cluster	PT/*stx*	Total cases	Cases identified through MLVA[1]
1. minced beef	PT 21/28 *stx*2	9	6
2. wedding reception	PT 8 *stx*1+2	2	0
3. nursery/school A[2]	PT 4/47 *stx*2	2	0
4. nursery/school B	PT 21/28 *stx*2	6	0
5. nursery/school C	PT 8 *stx*1+2	4	0
6. nursery/school D	PT 32 *stx*2	5	0
7. Petting farm[2]	PT 4 *stx*2	2	0
8. Country park	PT 54 *stx*2	8	0
Total		38	6

[1]Cases not linked with recognised clusters through phage typing and VESQ data.
[2]One case from the farm seeded the outbreak in nursery A.
doi:10.1371/journal.pone.0085901.t001

Table 2. Summary of evidence from NESSS for linkage between cases for community clusters of STEC O157 identified through MLVA between 1st May and 31st October 2012.

Evidence of linkage between cases	No. clusters	Total No. cases	Min. cases/cluster	Max. cases/cluster
MLVA, temporal[1] & shared exposures[2]	4	29	2	12
MLVA & temporally related cases[1]	8	45	2	11
MLVA linked only	8	27	2	7
All clusters	20	101	2	13

[1]Cases with onset dates between zero and seven days apart.
[2]Includes residing in or travel within the UK to the same area, shared direct or indirect contact with animals and/or their environment.
doi:10.1371/journal.pone.0085901.t002

each had two to three cases, but one larger cluster comprised eight cases (one with a different PT, figure 1), and two clusters 11 cases each. The epidemic curves, lack of unusual exposures and the wide geographical distribution of cases were suggestive of widely distributed food-borne sources of infection. For one cluster of 11, nine cases were re-interviewed. The only common exposure among all nine cases was the consumption of pre-packed salad from different branches of one major supermarket chain. No further cases were reported and it was not possible to undertake further analytical epidemiology.

Clusters with no additional linkage. No additional evidence for a cluster could be found from the ESQ for eight MLVA clusters, totalling 27 cases. The size of these clusters varied from two to seven cases. Within these clusters, cases did not appear to be linked geographically, were dispersed over time, spanning several months, and no obvious shared exposures were reported on the ESQ's.

Discussion

A number of studies have described the ease of use and timeliness of MLVA and demonstrated the discriminatory power of this approach over phage typing and pulse-field gel electrophoresis.[17-19] This study also clearly demonstrates the increased discriminatory power of MLVA over phage typing, with three-quarters of MLVA profiles being unique to one case. The addition of MLVA typing identified twenty additional clusters among domestic cases, reducing the proportion of sporadic cases by almost 30%. This approach thus better defines the epidemiology of STEC infection in England, providing a more accurate description of sporadic and cluster cases. Analysis of more accurately defined sporadic cases over time will facilitate better characterisation of the population at risk and enable a more accurate description of the important exposures that may be leading to these infections.

In England, the most prevalent STEC O157 PT's are PT21/28 and PT8. Five of the eight known clusters identified through public health investigation and supplemented by phage typing, were of a PT other than PT 21/28 or PT 8. Detection of rarer PT's facilitates identification of clusters. In addition, all but one of the known clusters were linked to a precise location, and all cluster cases were temporally linked. In contrast, only a fifth of the MLVA identified clusters were linked geographically, and not to precise locations. In addition, most were of PT 21/28 or PT8, as per all STEC O157 cases, making it extremely difficult to detect such clusters with phage typing alone. There may be under-reporting of PT 8 cases as symptoms appear less severe which may increase the likelihood of PT8 clusters occurring undetected. In contrast, PT 21/28 is associated with more severe disease and progression to HUS than other PT's. [1] As the greatest burden of domestic infection is attributed to this PT, it is important to detect clusters, identify sources of infection and establish control measures.

MLVA was useful in informing investigation of known clusters. The additional data can be used to elucidate previously undisclosed links, such as in the farm cluster seeding the nursery cluster, thus highlighting gaps in control measures. MLVA data had a substantial impact on investigations of Cluster#1 (associated with minced beef outbreak), where it captured additional cases and widened the focus of the outbreak control team. Through increased ascertainment of linked cases, both the likelihood and accuracy of determining the source of infection was improved.

Although MLVA identified additional clusters, analyses of data did not lead to an increased chance of a public health intervention. For most of the 20 clusters newly detected through MLVA, reviewing the ESQ data revealed temporal clustering of cases, but additional evidence of shared exposures was indicated in just four clusters. In such clusters where cases are linked to a relatively broad geographic area and the potential source of infection is environmental, it is difficult to pinpoint the source of infection from ESQ data. For eight clusters, no temporal, geographic or exposure linkage was detected. The ability to establish epidemiological links between cases is in part diminished due to the small sample size of the clusters, making it difficult to confidently establish commonality of exposures. ESQ data were retrospectively reviewed when some time had passed since most cases were ill, the clusters were small in size and were temporally contained. Therefore the ability to undertake further analytical study was limited.

There were three larger clusters comprising nationally distributed cases with apparent temporal links suggestive of a food-borne source of infection. While, the ESQ, introduced in England in 2009, has made a major contribution to STEC surveillance, it was not designed to capture the highly detailed information often required to definitively link cases to a shared exposure. The ESQ is not a trawling questionnaire so the vehicle of infection, for example, could be a food-stuff not included in the questionnaire or included but not in enough detail to identify it. Additionally, similarities in consumption between cases may reflective of common population behaviours, such as in the cluster of cases reporting consumption of pre-packed salad.

While this study demonstrated the added value of MLVA in increasing recognition of clusters and providing insight into the microbiology and epidemiology of STEC O157 in England, it raises questions in terms of how to respond in terms of public health. Most STEC clusters are small so the chance of proof of source and hence intervention is reduced, while analytical studies are not possible. The resource involved in investigating clusters is not inconsiderable; input from a wide range of multi-disciplinary professionals is required and in nationally dispersed clusters, co-

ordinating local investigations is challenging. In real-time surveillance the eventual size of a cluster is unknown, and difficulties lie in deciding at what point hypothesis generation exercises should be initiated. The delay in obtaining MLVA data (although MLVA is more rapid than PFGE, the MLVA data is often available more than eight days after the patient becomes symptomatic), and thereby identifying linked cases, may also impede investigation. The increased detection of clusters through the additional sensitivity of MLVA typing provides a forerunner for the advent of next generation sequencing of clinical isolates and highlights both the advantages to, and the challenges in public health practice, in utilising molecular methods as routine, which need to be addressed.

PFGE remains the method of choice for outbreak investigation in many countries, although the use of MLVA is increasing as the contribution of the MLVA approach to overall surveillance of STEC O157, in addition to outbreak investigation, is recognised. Despite the challenges described above, the universal MLVA approach has revealed a more accurate picture of the extent of linked cases in England and demonstrated that while STEC O157 clusters in England are small, they contribute a significant proportion of cases. Continuation of MLVA typing on all confirmed STEC O157 isolates submitted to the GBRU will generate more data on the spatial and temporal dispersion of cases of STEC O157. Furthermore, in conjunction with enhanced surveillance data and timely public health investigation of clusters, MLVA offers the potential to provide a more accurate measure of clusters, insight into routes of this infection in England and evidence to inform health protection responses at local and national levels.

Acknowledgments

We would like to thank Kirsten Glen, Natalie Adams, Naomi Launders and Radha Patel for their contributions to the National Enhanced Surveillance System for STEC in England. Thanks to all public health practitioners who report to the scheme. We would also like to acknowledge Marie Chattaway, Vivienne do Nascimento and Michael Wright in GBRU for their laboratory expertise.

Author Contributions

Conceived and designed the experiments: LB RE CJ TD JW KG GA. Performed the experiments: NP CJ. Analyzed the data: LB RE NP CJ TD PA. Wrote the paper: LB CJ RE. Critically reviewed the manuscript and approved the final version for submission: TD NP PA JW GA KG.

References

1. Lynn RM, O'Brien SJ, Taylor CM, Adak GK, Chart H, et al. (2005) Childhood hemolytic uremic syndrome, United Kingdom and Ireland. Emerg Infect Dis 11(4):590–596.
2. Armstrong GL, Hollingsworth J, Morris JG Jr (1996) Emerging foodborne pathogens: Escherichia coli O157:H7 as a model of entry of a new pathogen into the food supply of the developed world. Epidemiol Rev 18(1):29–51.
3. Gillespie IA, O'Brien SJ, Adak GK, Cheasty T, Willshaw G (2005) Foodborne general outbreaks of Shiga toxin-producing Escherichia coli O157 in England and Wales 1992–2002: where are the risks? Epidemiol Infect 133(5):803–808.
4. Locking ME, O'Brien SJ, Reilly WJ, Wright EM, Campbell DM, et al. (2001) Risk factors for sporadic cases of Escherichia coli O157 infection: the importance of contact with animal excreta. Epidemiol Infect 127(2):215–220.
5. O'Brien SJ, Adak GK, Gilham C. (2001) Contact with farming environment as a major risk factor for Shiga toxin (Vero cytotoxin)-producing Escherichia coli O157 infection in humans. Emerg Infect Dis 7(6):1049–1051.
6. Pritchard GC, Smith R, Ellis-Iversen J, Cheasty T, Willshaw GA. (2009) Verocytotoxigenic Escherichia coli O157 in animals on public amenity premises in England and Wales, 1997 to 2007. Vet Rec 164(18):545–549.
7. Cowden JM, Ahmed S, Donaghy M, Riley A. (2001) Epidemiological investigation of the central Scotland outbreak of Escherichia coli O157 infection, November to December 1996. Epidemiol Infect 126(3):335–341.
8. Ihekweazu C, Carroll K, Adak B, Smith G, Pritchard GC, et al. (2012) Large outbreak of verocytotoxin-producing Escherichia coli O157 infection in visitors to a petting farm in South East England, 2009. Epidemiol Infect 140(8):1400–1413.
9. Salmon R. (2005) Outbreak of verotoxin producing E.coli O157 infections involving over forty schools in south Wales, September 2005. Euro Surveill 10(10):E051006.
10. Khakhria R, Duck D, Lior H. (1990) Extended phage-typing scheme for Escherichia coli O157:H7. Epidemiol Infect 105(3):511–520.
11. Hyytia-Trees E, Smole SC, Fields PA, Swaminathan B, Ribot EM. (2006) Second generation subtyping: a proposed PulseNet protocol for multiple-locus variable-number tandem repeat analysis of Shiga toxin-producing Escherichia coli O157 (STEC O157). Foodborne Pathog Dis 3(1):118–131.
12. Perry N, Cheasty T, Dallman T, Launders N, Willshaw G. (2013) Application of multi-locus variable number tandem repeat analysis to monitor Verocytotoxin-producing Escherichia coli O157 phage type 8 in England and Wales: emergence of a profile associated with a national outbreak. J Appl Microbiol:10.
13. Underwood AP, Dallman T, Thomson NR, Williams M, Harker K, et al. (2013) Public health value of next-generation DNA sequencing of enterohemorrhagic Escherichia coli isolates from an outbreak. J Clin Microbiol 51(1):232–237.
14. Jenkins C, Lawson AJ, Cheasty T, Willshaw GA. (2012) Assessment of a real-time PCR for the detection and characterization of verocytotoxigenic Escherichia coli. J Med Microbiol 61 (Pt 8):1082–1085.
15. Hyytia-Trees E, Lafon P, Vauterin P, Ribot EM. (2010) Multilaboratory validation study of standardized multiple-locus variable-number tandem repeat analysis protocol for shiga toxin-producing Escherichia coli O157: a novel approach to normalize fragment size data between capillary electrophoresis platforms. Foodborne Pathog Dis 7(2):129–136.
16. Hunter PR, Gaston MA. (1988) Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J Clin Microbiol 26(11):2465–2466.
17. Dyet KH, Robertson I, Turbitt E, Carter PE. (2011) Characterization of Escherichia coli O157:H7 in New Zealand using multiple-locus variable-number tandem-repeat analysis. Epidemiol Infect 139(3):464–471.
18. Noller AC, McEllistrem MC, Pacheco AG, Boxrud DJ, Harrison LH. (2003) Multilocus variable-number tandem repeat analysis distinguishes outbreak and sporadic Escherichia coli O157:H7 isolates. J Clin Microbiol 41(12):5389–5397.
19. Pei Y, Terajima J, Saito Y, Suzuki R, Takai N, et al. (2008) Molecular characterization of enterohemorrhagic Escherichia coli O157:H7 isolates dispersed across Japan by pulsed-field gel electrophoresis and multiple-locus variable-number tandem repeat analysis. Jpn J Infect Dis 61(1):58–64.

Paper 4: Whole Genome Sequencing for National Surveillance of Shiga Toxin Producing *Escherichia coli* O157.

Clinical Infectious Diseases Advance Access published May 13, 2015

MAJOR ART

Whole-Genome Sequencing for National Surveillance of Shiga Toxin–Producing *Escherichia coli* O157

Timothy J. Dallman,[1,a] Lisa Byrne,[1,a] Philip M. Ashton,[1] Lauren A. Cowley,[1] Neil T. Perry,[1] Goutam Adak,[1] Liljana Petrovska,[2] Richard J. Ellis,[2] Richard Elson,[1] Anthony Underwood,[1] Jonathan Green,[1] William P. Hanage,[3] Claire Jenkins,[1] Kathie Grant,[1] and John Wain[4]

[1]Public Health England, London, and [2]Animal and Plant Health Agency, Addlestone, Surrey, United Kingdom; [3]Harvard School of Public Health, Boston, Massachusetts; and [4]University of East Anglia, Norwich, United Kingdom

Background. National surveillance of gastrointestinal pathogens, such as Shiga toxin–producing *Escherichia coli* O157 (STEC O157), is key to rapidly identifying linked cases in the distributed food network to facilitate public health interventions. In this study, we used whole-genome sequencing (WGS) as a tool to inform national surveillance of STEC O157 in terms of identifying linked cases and clusters and guiding epidemiological investigation.

Methods. We retrospectively analyzed 334 isolates randomly sampled from 1002 strains of STEC O157 received by the Gastrointestinal Bacteria Reference Unit at Public Health England, Colindale, in 2012. The genetic distance between each isolate, as estimated by WGS, was calculated and phylogenetic methods were used to place strains in an evolutionary context.

Results. Estimates of linked clusters representing STEC O157 outbreaks in England and Wales increased by 2-fold when WGS was used instead of traditional typing techniques. The previously unidentified clusters were often widely geographically distributed and small in size. Phylogenetic analysis facilitated identification of temporally distinct cases sharing common exposures and delineating those that shared epidemiological and temporal links. Comparison with multi locus variable number tandem repeat analysis (MLVA) showed that although MLVA is as sensitive as WGS, WGS provides a more timely resolution to outbreak clustering.

Conclusions. WGS has come of age as a molecular typing tool to inform national surveillance of STEC O157; it can be used in real time to provide the highest strain-level resolution for outbreak investigation. WGS allows linked cases to be identified with unprecedented specificity and sensitivity that will facilitate targeted and appropriate public health investigations.

Keywords. public health; whole-genome sequencing; Shiga toxin–producing *Escherichia coli* O157; national surveillance.

Gastrointestinal disease is an important public health problem in England, with up to 20% of the population experiencing at least 1 episode of acute gastroenteritis each year [1]. An effective national surveillance program for gastrointestinal diseases is imperative to identify cases with linked exposures; this is especially pertinent for pathogens that may enter nationally distributed food networks. Although conventional epidemiological investigation using detailed questionnaires and contact tracing is vital, to achieve optimal surveillance we must complement these activities with a rapid and robust molecular typing method to accurately discriminate between linked cases and sporadic infections.

With >1000 presumptive isolates submitted to the Gastrointestinal Bacteria Reference Unit (GBRU) annually [2], infections with Shiga toxin–producing *Escherichia coli* O157 (STEC O157) continue to exert a public health burden in England, both economically and in

Received 8 December 2014; accepted 4 April 2015.
[a]T. J. D. and L. B. contributed equally to this work.
Correspondence: Claire Jenkins, BSc, PhD, Public Health England, 61 Colindale Avenue, London NW9 5EQ, UK (claire.jenkins@phe.gov.uk).
Clinical Infectious Diseases®

DOI: 10.1093/cid/civ318

terms of morbidity and mortality. Symptoms of STEC infection range from mild to severe, but typically include bloody diarrhea. Approximately 6% of cases develop hemolytic uremic syndrome (HUS) [3].

The main reservoir of STEC O157 in England is cattle, although it is carried by other animals, mainly ruminants [4, 5]. Transmission to humans occurs through direct or indirect contact with animals or their environments, consumption of contaminated food or water, and person-to-person contact [6–8]. Contamination of the food supply can cause large-scale national and multinational outbreaks [9–11].

Outbreaks, involving ≥2 cases in different households or residential institutions, vary in number annually but since 2009 have contributed between 9% and 25% of isolates in England and Wales (GBRU/Department of Gastrointestinal Emerging and Zoonotic Infections, in-house data), with the majority of cases being apparently sporadic. All isolates received by GBRU are routinely phage typed [12], but in England, the majority (60%) of isolates are either PT8 or PT21/28, and so the ability of this method to discriminate between cases resulting from separate exposures is very low. Multi locus variable number tandem repeat analysis (MLVA) is used to provide higher levels of strain discrimination.

The utility of whole-genome sequencing (WGS) for the investigation of outbreaks has already been demonstrated for several bacterial pathogens [13, 14], and there is increasing evidence in the literature for the positive contribution of WGS to outbreaks involving gastrointestinal pathogens [15–19]. The aim of this study was to expand the use of WGS by evaluating a WGS approach to inform national surveillance of a major pathogen. By validating the WGS approach using clearly defined outbreak and sporadic cases of STEC O157 and by investigating the findings, WGS can provide additional insights into outbreak definition, transmission networks, and other aspects of the underlying epidemiology of this pathogen.

METHODS

Strain Selection

A total of 572 isolates were selected for sequencing: 334 isolates were randomly selected from 1002 STEC O157 culture-positive isolates received by GBRU from cases in England, Wales, and Northern Ireland during 2012; 147 isolates were randomly selected from 939 STEC O157 culture-positive isolates from cases in England, Wales, and Northern Ireland received by GBRU in 2013; and an additional 91 English historical isolates received between 1990 and 2011 were selected based on phage type diversity to provide context as a sample of the background population. The total collection contained strains from known outbreaks, household clusters, serial strains isolated from the same patient, and strains from apparently sporadic cases. A total of 18 phage types [20] were represented.

Genome Sequencing and Sequence Analysis

Genomic DNA was fragmented and tagged for multiplexing with Nextera XT DNA Sample Preparation Kits (Illumina) and sequenced using the Illumina GAII platform with 2 × 150 bp reads. Short reads were mapped to the reference STEC O157 strain Sakai [21] using BWA-SW [22]. The sequence alignment map output from BWA was sorted and indexed to produce a binary alignment map (BAM) using Samtools [23]. GATK2 [24] was used to create a variant call format file from each of the BAMs, which were further parsed to extract only single-nucleotide polymorphism (SNP) positions that were of high quality in all genomes (Mapping Quality >30, Depth >10, Genotype Quality >30, Variant Ratio >0.9). An alignment of polymorphic positions was used to create approximate maximum likelihood trees using FastTree [25] under the Jukes–Cantor model of nucleotide evolution. Pairwise SNP distances between the genomes of each strain were calculated. FASTQ sequences were deposited in the National Center for Biotechnology Information Short Read Archive under bioproject PRJNA248042.

Data Handling

STEC enhanced surveillance questionnaires (SESQs) are administered to all cases of STEC O157 in England. The SESQ collects demographic details; risk status; clinical condition (including progression to HUS); household or other close contact details; exposures including travel, food, and water consumption, contact with animals, and environmental factors; epidemiological case classification; and outbreak /cluster status. SESQ data were reviewed for each selected strain and strains classified in respect to known outbreak status, known household cluster status, or whether multiple isolates originated from the same patient. Any strains fulfilling these criteria were designated as having a known epidemiological link.

Pairwise SNP distances were calculated for all strains in this study. In previously reported outbreaks, onset of illness in cases occurs a median of 39 days from another linked case with a mode of 1 (Public Health England). Using specimen dates of isolates, temporality between isolates of different genetic distances were compared. The pairwise SNP distribution and temporal links between known linked cases was examined and a relatedness threshold determined accordingly. As related strains are likely to originate from a common source, the threshold was termed the common source threshold (CST). This threshold was then applied to all other strains in the dataset and evaluated for epidemiological context.

Related strains within the CST were classified into clusters on the basis of having at least 1 SNP distance within the CST to another isolate in the dataset. Clusters not previously identified

were designated WGS linked clusters. Temporal and geographic links between cases in clusters were examined and comparisons made between epidemiologically identified and WGS linked clusters.

Deeper phylogenetic relationships were also investigated to ascertain whether they provided epidemiologically useful information or associations. Clusters of 25 SNP genetic distance were constructed (herby referred to as phylogenetic clusters [PCs]) and those with >1 CST cluster within each PC were investigated for shared epidemiological associations.

All STEC O157 isolates reported between 1 May 2012 and 31 December 2013 that have been typed through both MLVA and WGS were used to investigate clustering dynamics for each method. Survival analysis was used to test the null hypothesis that there is no difference in timeliness and completeness of clustering-related isolates using the 2 methods. For survival analysis, an isolate clustering with another isolate based on ≤1 locus variant for MLVA or ≤CST for WGS represented a failure. Across the study period, isolates will enter at various time points based on laboratory report date. At that point, the isolate is at risk of clustering with other isolates already in the study population or isolates entering the study at a later date. Kaplan–Meier estimates of the survivor function were estimated for both methods and displayed as cumulative survival curves with accompanying tables presenting those at risk at specific time points. The proportional hazards assumption was tested by plotting the log cumulative hazard in both groups. Where the proportional hazards assumption applied, the survival function in the 2 groups was compared by calculating a hazard ratio using Cox regression.

RESULTS

Distribution of Pairwise Distance Between Closely Related Isolates

For 183 of 425 strains used in this study, an epidemiological link to at least 1 other case was known. This included 16 cases where multiple isolates were sequenced from the same person, 43 isolates that were part of 26 separate household clusters, and 124 cases that were part of 14 known outbreaks. The remaining 242 strains had no common link previously identified. The pairwise SNP distance distribution revealed that no pair of epidemiologically linked isolates had >5 SNP differences with a mean of 1 SNP in isolates from the same household (standard deviation [SD], 0.99) or known common source (SD, 1.04) and 0.3 SNPs (SD, 0.60) from isolates from the same person (Figure 1).

One hundred thirty-six cases with no known epidemiological link were within ≤5 SNPs to another case. The majority (87%) of pairs that fell within the 5 SNP threshold comprised strains isolated within 30 days of each other with a mean interval between pairs of samples being 11 days. Between a genetic distance of 5 and 10 SNPs, the mean interval between pairs of samples (n = 24) increased to 258 days (Figure 2). As all previously linked isolates fell within a 5 SNP threshold and the majority of pairs of cases within this threshold were temporally linked, we hypothesize a threshold of 5 SNPs to categorize isolates as related. In this context, the term "related" alludes to a common source of infection and strains that are within 5 SNPs of another are referred to as falling within the CST.

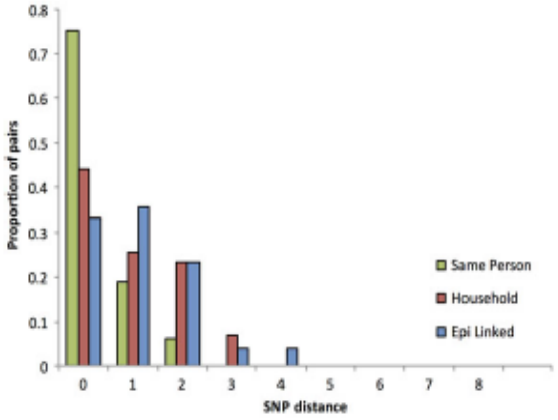


Figure 1. Histogram showing proportion of pairs against single-nucleotide polymorphism (SNP) distance of cases with a known epidemiological link.

Applying the CST

One hundred sixty strains isolated during 2012 fell within the CST. These strains can be formed into 53 clusters where members of the cluster must share at least 1 link within the CST. Twenty of the clusters (46 strains) represented either household outbreaks or multiple strains from the same patient. The remaining 33 clusters comprised 114 strains representing 34% of the dataset. Routine public health investigation previously undertaken had not identified 20 of 33 clusters, and these were designated WGS linked clusters. Of the 20 WGS linked clusters, 18 comprised between 2 and 4 cases, while 2 larger clusters comprised 12 and 7 cases. Overall, if we conclude that all cases within the CST are part of epidemiologically linked clusters, this corresponds to an increase in sensitivity of >50% in detecting linked cases outside the household setting when using WGS to supplement the current approach.

Epidemiology of WGS Linked Clusters

The 20 WGS linked clusters were statistically more geographically dispersed than the 13 epidemiologically linked clusters (Figure 3A), with a mean residential distance of 169 km (SD, 111 km) for the former and 29 km (SD, 34 km) for the latter ($P = .04$ [5], 1-tailed t test). Strains of STEC O157 associated with a large national foodborne PT8 outbreak from 2011

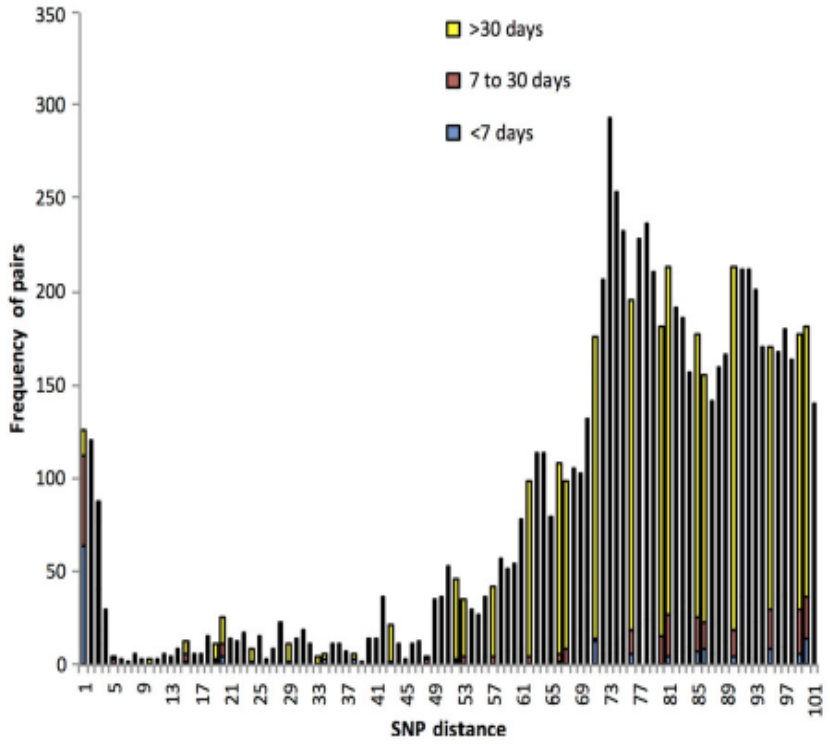


Figure 2. Histogram showing frequency of pairs against single-nucleotide polymorphism (SNP) distance. Each bar is colored as a proportion of pairs isolated within <7 days, 7–30 days, and >30 days.

[9] and a petting farm PT21/28 outbreak [26] were included for context (Figure 3*B*). The geographical dispersal of cases linked by WGS mirrors the distribution of the national PT8 outbreak as well as encompassing the distribution of a geographically restricted outbreak. Conversely, the epidemiologically linked clusters most closely mirrored the geographically restricted outbreak, highlighting the difficulty in recognizing national distributed cases without high-resolution strain discrimination such as WGS.

Retrospective epidemiological follow-up was undertaken for cases in the 2 larger WGS clusters. Cluster 1 comprised 12 nationally distributed cases with onset dates all within 15 days of each other. Following reinvestigation, the only common exposure identified was the consumption of a specific prepacked leafy salad from different branches of 1 major supermarket chain. Cluster 2 contained 7 cases, of which 4 cases were from separate English public health regions with onset dates spanning a 2-week period. Following reinvestigation, it was identified that 3 cases had visited the same village, where another case was resident, within the incubation period. All 4 cases had been hiking in the same national park, putting them at risk of environmental exposure. The remaining 3 cases shared no obvious exposures, suggesting that the cases were exposed to the same source of infection but via different routes and/or vehicles, highlighting the importance of using all available epidemiological investigations when interpreting WGS for outbreak investigations.

Outbreak Detection MLVA Versus WGS

Clustering based on the WGS defined CST increased sensitivity in identifying linked cases; however, it was also necessary to compare this approach to other fine-typing methods deployed for STEC O157 (eg, MLVA). Using a survival analysis of 481 samples typed by both methods in 2012–2013, survival (ie, not clustering with another isolate) showed no significant difference with MLVA vs WGS CST based on clustering a single isolate with another (log-rank test for equality of survivor function: $P = .101$; Cox hazard ratio = 0.89, $P = .198$) (Figure 4). This indicates there is no difference in timeliness of clustering between the 2 methods. However, when we consider the time to cluster completion (all cases of a cluster are clustered) from the initial cluster event (any 2 cases of a cluster are clustered), there is a significant speed increase in time to completion of clusters with WGS CST as opposed to MLVA (log-rank test for equality of survivor function: $P = .0006$; Cox hazard ratio = 1.44, $P = .001$) (Figure 5).

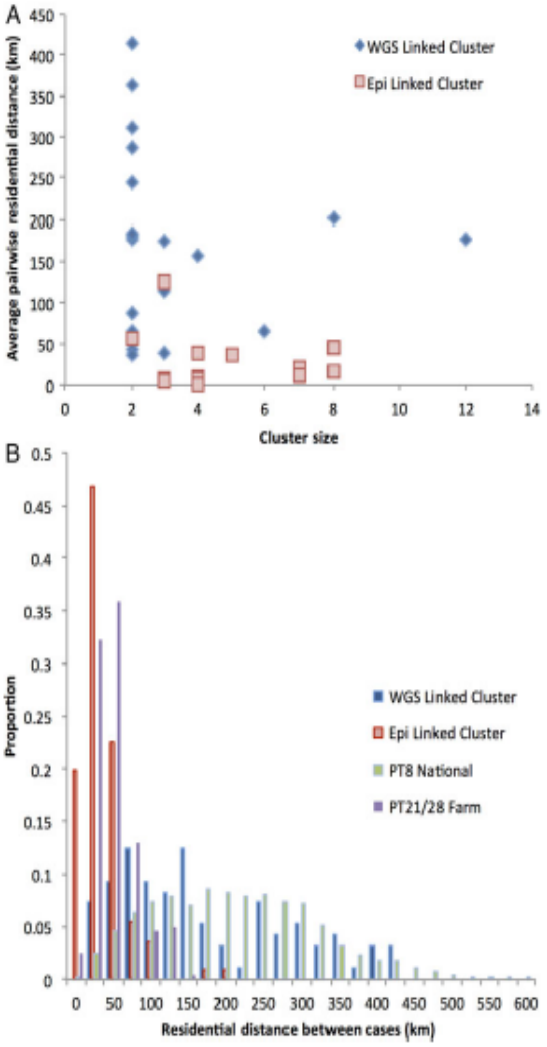


Figure 3. *A*, Scatter diagram showing the average pairwise residential distance of each close contact cluster against the size in number of cases. The coloring represents whether the cluster was already identified through epidemiological investigation or if identified by whole-genome sequencing (WGS) alone. *B*, Histogram showing the distribution of residential distance for WGS linked clusters and epidemiologically linked clusters. PT8 National and PT21/28 Farm represent distributed foodborne and point source outbreaks, respectively.

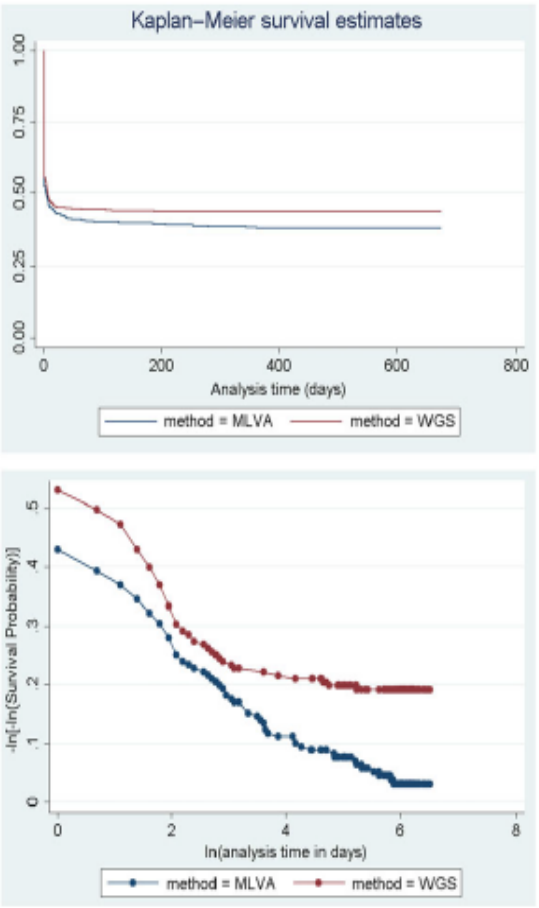


Figure 4. Kaplan–Meier failure estimates and proportional hazards assumption test showing that there is no difference in timeliness of clustering between whole-genome sequencing (WGS) and multi locus variable number tandem repeat analysis (MLVA).

Epidemiological Context of Phylogenetic Clusters

Cases within the CST represent temporally linked cases, and these have been shown to include cases with common epidemiological exposures. Although the temporal relationship between pairs quickly dissipated as the genetic distance moved outside the CST, we investigated whether deeper phylogenetic relationships also provided epidemiologically useful information or associations. Nineteen PCs (see "Methods" section) were identified, and 10 had no geographical association or common exposures between the CST clusters within as assessed through the SESQ. One PC contained 3 CST clusters sharing a common exposure to a national park in the Midlands (Figure 6); each cluster correlated with year of isolation, highlighting the potential to identify the persistence of strains in the environment over time.

Two PCs contained CST clusters where the majority of strains were of Northern Irish provenance. Those patients who were not resident in Northern Ireland reported travel to various parts of the province in their SESQ. Similarly, PCs were identified with cases associated with Wales and travel to the Middle East (Figure 7).

DISCUSSION

In this study, the potential of WGS in national surveillance of STEC O157 was assessed for its ability to improve outbreak detection and provide additional insights over conventional epidemiological investigations. WGS confirmed that strains from the

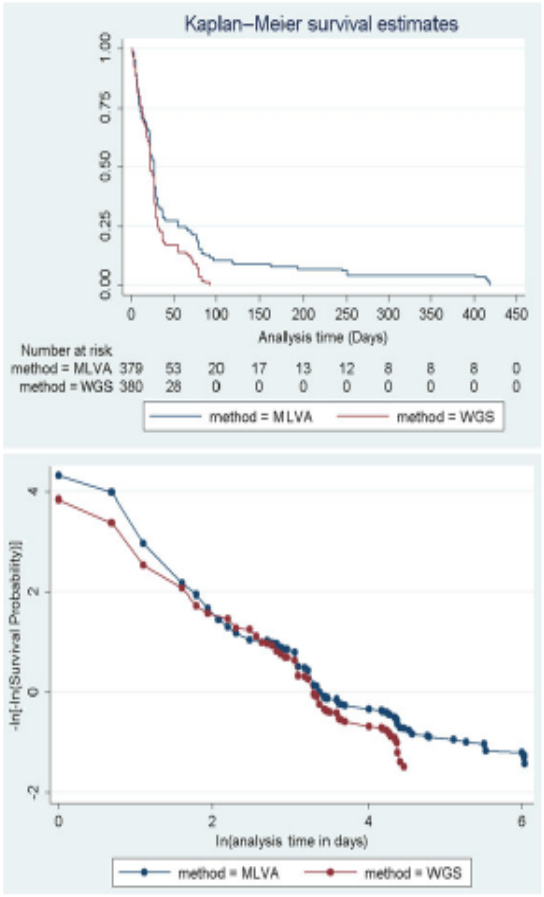


Figure 5. Kaplan–Meier survival estimates and proportional hazards assumption test showing that after isolates have clustered, time to completion of that cluster is significantly faster with whole-genome sequencing (WGS) than with multi locus variable number tandem repeat analysis (MLVA).

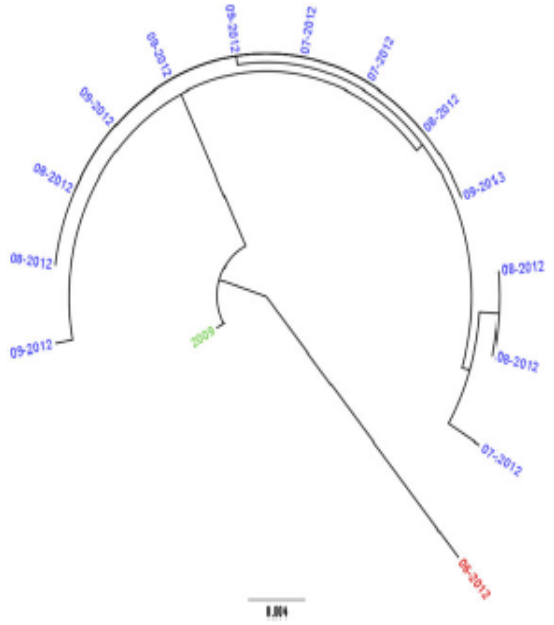


Figure 6. Maximum likelihood phylogeny of 15 isolates representing 27 single-nucleotide polymorphisms (SNPs) across 25 coding DNA sequence (2 noncoding SNPs) with a total core genome size of 4 915 463 bp associated with cases that visited the same national park. The clusters represent 3 different common source threshold clusters, colored red, blue, and green within a single phylogenetic cluster. The level of resolution allows the delineation of strains from different years. The strain in red was temporally related to the strains in blue but significantly different genomically to suggest a different source of Shiga toxin–producing *Escherichia coli* exposure.

same patient, from cases within the same household, and from cases with known epidemiological links had little or no difference in their core genomes. These cases fell within a 5 SNP threshold within which we found strong temporal correlations suggestive of epidemiological linkage. Using this empirically observed cutoff of 5 SNPs, we could determine with unprecedented clarity which strains of STEC O157 were likely to be epidemiologically linked. WGS detected linked cases of STEC O157 in 334 representative strains from an annual season with twice the sensitivity of current methods. This suggests that current outbreak detection is highly specific, but comparatively insensitive, and that the previous estimate of outbreaks, involving ≥2 cases in different households or residential institutions, contributing between 9% and 25% of isolates in England and Wales, is conservative. Previously elusive clusters were often more geographically dispersed than those identified using the traditional approach. It is suggested that these geographically dispersed outbreaks with no obvious common exposures are foodborne. This type of outbreak profiling will facilitate outbreak investigations through focusing hypothesis generation on foodborne exposures at an early stage.

In this study, we show that for identifying linked cases, the current threshold of ≤1 locus variant for clustering provides the same sensitivity as using the WGS CST. This is an important finding, as it not only gives confidence in the interpretation of MLVA to those public health laboratories not yet ready to adopt WGS methodologies but also allows cross-communication of results between practitioners of these 2 techniques. An important distinction between the 2 methods is the time it takes to resolve complete clusters of cases within an outbreak, with WGS CST completing clusters significantly faster than MLVA. This feature can be explained by the fact all linked cases tend to fall within the CST for all cases, whereas in a large MLVA cluster several isolates will only be joined via an intermediate isolate (ie, double-locus variants joined by a shared single-locus variant). This phenomenon has implications in accurately defining the microbiological case definition at the start of an outbreak investigation as outbreaks that resolve themselves to a single cluster

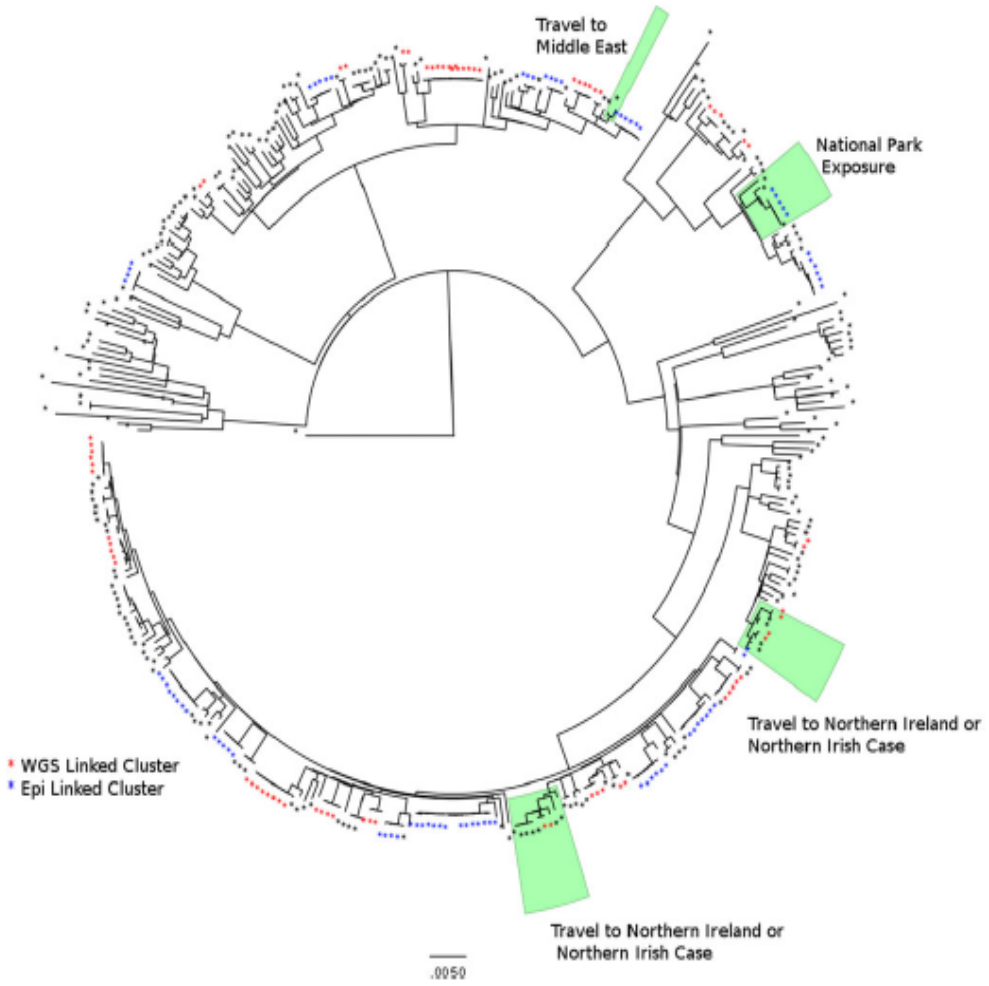


Figure 7. Maximum likelihood phylogeny of 374 isolates, representing 7756 single-nucleotide polymorphisms (SNPs), across 2902 coding DNA sequence (1166 noncoding SNPs) with a total core genome size of 3 808 948 bp. Common source threshold clusters identified through whole-genome sequencing (WGS) alone are colored red, and those identified through traditional methods are colored blue. Phylogenetic clusters that contained strains with related exposures are shaded green.

may appear as multiple clusters until intermediate isolates are sampled.

The phylogenetic context of common source clusters was analyzed to see if there was any epidemiological signal between separate but related common source events. Several regional or travel-associated PCs were identified, highlighting the geographical isolation of STEC O157 even within the British Isles. The geographical signal observed in the WGS of STEC O157 has been described previously [27] and has obvious implications in facilitating outbreak investigations. For example, isolates could be linked to food sourced from specific regions of the world, or cases could be ruled out of a point source outbreak by confirming their strain originated from further afield, given adequate sampling of potential source populations.

The primary aims of gastrointestinal disease surveillance are to identify outbreaks, monitor long-term trends, and inform the effectiveness of policy and other public health interventions. WGS demonstrates unparalleled sensitivity and accuracy in identifying linked cases coupled with phylogenetic clustering of how strains are related over time and space. Its ability to accurately define sporadic cases over time enables better characterization of the population at risk and to assess the relative importance of exposures leading to sporadic infections, which may differ from those leading to outbreaks.

Timely analysis and interpretation of WGS data will inform public health interventions by identifying linked cases (ie, early warning of outbreaks) as well as inferring epidemiological context through evolutionary relationships. Furthermore, the ability to unambiguously rule out associations will prevent inappropriate

public health actions from being taken, saving resources at the health protection and local authority level. Good communication and rapid sharing of real-time STEC O157 WGS data with colleagues working in the agriculture, veterinary, and food industries across international borders will allow evidence-based trace-back of isolates to their source and reveal specific risk factors in the food chain and environment, thus facilitating the targeting of resources and public health interventions to have maximum impact on reducing the burden of STEC O157 disease in England.

Notes

Acknowledgments. We would like to thank all the staff at Gastrointestinal Bacteria Reference Unit and Gastrointestinal Emerging and Zoonotic Infections who work on STEC O157.

Disclaimer. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health (NIH).

Financial support. This work was supported by the National Institute for Health Research scientific research development fund (grant number 108601). W. P. H. was supported by the National Institute of General Medical Sciences of the NIH (award number U54GM088558).

Potential conflicts of interest. J. W. is on the board of Discuva Ltd and has received funding from Hoffmann-La Roche. W. P. H. has received funds from Antigen Discovery Ltd. All other authors report no potential conflicts.

All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Wheeler JG, Sethi D, Cowden JM, et al. Study of infectious intestinal disease in England: rates in the community, presenting to general practice, and reported to national surveillance. The Infectious Intestinal Disease Study Executive. BMJ **1999**; 318:1046–50.
2. Jenkins C, Lawson A, Cheasty T, Bolton E, Smith G. Assessment of a real-time PCR for the detection and characterisation of verocytotoxigenic *Escherichia coli*. J Med Microbiol **2012**; 61:1082–5.
3. Lynn RM, O'Brien SJ, Taylor CM, et al. Childhood hemolytic uremic syndrome, United Kingdom and Ireland. Emerg Infect Dis **2005**; 11:590–6.
4. Pennington H. *Escherichia coli* O157. Lancet **2010**; 376:1428–35.
5. Ferens WA, Hovde CJ. *Escherichia coli* O157:H7: animal reservoir and sources of human infection. Foodborne Pathog Dis **2011**; 8:465–87.
6. Locking ME, O'Brien SJ, Reilly WJ, et al. Risk factors for sporadic cases of *Escherichia coli* O157 infection: the importance of contact with animal excreta. Epidemiol Infect **2001**; 127:215–20.
7. Gillespie IA, O'Brien SJ, Adak GK, Cheasty T, Willshaw G. Foodborne general outbreaks of Shiga toxin-producing *Escherichia coli* O157 in England and Wales 1992–2002: where are the risks? Epidemiol Infect **2005**; 133:803–8.
8. Pritchard GC, Smith R, Ellis-Iversen J, Cheasty T, Willshaw GA. Verocytotoxigenic *Escherichia coli* O157 in animals on public amenity premises in England and Wales, 1997 to 2007. Vet Rec **2009**; 164:545–9.
9. Perry N, Cheasty T, Dallman T, Launders N, Willshaw G. Application of multi-locus variable number tandem repeat analysis to monitor verocytotoxin-producing *Escherichia coli* O157 phage type 8 in England and Wales: emergence of a profile associated with a national outbreak. J Appl Microbiol **2013**; 115:1052–8.
10. Buchholz U, Bernard H, Werber D, et al. German outbreak of *Escherichia coli* O104:H4 associated with sprouts. N Engl J Med **2011**; 365:1763–70.
11. Bell BP, Goldoft M, Griffin PM, et al. A multistate outbreak of *Escherichia coli* O157:H7-associated bloody diarrhea and hemolytic uremic syndrome from hamburgers. The Washington experience. JAMA **1994**; 272:1349–53.
12. Khakhria R, Duck D, Lior H. Extended phage-typing scheme for *Escherichia coli* O157:H7. Epidemiol Infect **1990**; 105:511–20.
13. Walker TM, Ip CL, Harrell RH, et al. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. Lancet Infect Dis **2013**; 13:137–46.
14. Didelot X, Eyre DW, Cule M, et al. Microevolutionary analysis of *Clostridium difficile* genomes to investigate transmission. Genome Biol **2012**; 13:R118.
15. Gilmour MW, Graham M, Van DG, et al. High-throughput genome sequencing of two *Listeria monocytogenes* clinical isolates during a large foodborne outbreak. BMC Genomics **2010**; 11:120.
16. Mellmann A, Harmsen D, Cummings CA, et al. Prospective genomic characterization of the German enterohemorrhagic *Escherichia coli* O104:H4 outbreak by rapid next generation sequencing technology. PLoS One **2011**; 6:e22751.
17. Underwood AP, Dallman T, Thomson NR, et al. Public health value of next-generation DNA sequencing of enterohemorrhagic *Escherichia coli* isolates from an outbreak. J Clin Microbiol **2013**; 51:232–7.
18. McDonnell J, Dallman T, Atkin S, et al. Retrospective analysis of whole genome sequencing compared to prospective typing data in further informing the epidemiological investigation of an outbreak of *Shigella sonnei* in the UK. Epidemiol Infect **2013**; 141:2568–75.
19. Allard MW, Luo Y, Strain E, et al. High resolution clustering of *Salmonella enterica* serovar Montevideo strains using a next-generation sequencing approach. BMC Genomics **2012**; 13:32.
20. Willshaw GA, Smith HR, Cheasty T, Wall PG, Rowe B. Vero cytotoxin-producing *Escherichia coli* O157 outbreaks in England and Wales, 1995: phenotypic methods and genotypic subtyping. Emerg Infect Dis **1997**; 3:561–5.
21. Hayashi T, Makino K, Ohnishi M, et al. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. DNA Res **2001**; 8:11–22.
22. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics **2010**; 26:589–95.
23. Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools. Bioinformatics **2009**; 25:2078–9.
24. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res **2010**; 20:1297–303.
25. Price MN, Dehal PS, Arkin AP. FastTree 2—approximately maximum-likelihood trees for large alignments. PLoS One **2010**; 5:e9490.
26. Ihekweazu C, Carroll K, Adak B, et al. Large outbreak of verocytotoxin-producing *Escherichia coli* O157 infection in visitors to a petting farm in South East England, 2009. Epidemiol Infect **2012**; 140:1400–13.
27. Mellor GE, Sim EM, Barlow RS, et al. Phylogenetically related Argentinean and Australian *Escherichia coli* O157 isolates are distinguished by virulence clades and alternative Shiga toxin 1 and 2 prophages. Appl Environ Microbiol **2012**; 78:4724–31.

Paper 5: The epidemiology and microbiology of Shiga-toxin producing *Escherichia coli* other than serogroup O157 in England 2009-2013.

Journal of Medical Microbiology (2014), 63, 1181–1188 DOI 10.1099/jmm.0.075895-0

Epidemiology and microbiology of Shiga toxin-producing *Escherichia coli* other than serogroup O157 in England, 2009–2013

Lisa Byrne,[1] Gemma L. Vanstone,[2] Neil T. Perry,[3] Naomi Launders,[1] Goutam K. Adak,[1] Gauri Godbole,[3] Kathie A. Grant,[3] Robin Smith[2] and Claire Jenkins[3]

Correspondence
Claire Jenkins
clairejenkins1@phe.gov.uk

[1]Gastrointestinal, Emerging and Zoonotic Infections Department, Public Health England, 61 Colindale Avenue, London NW9 5EQ, UK

[2]Department of Medical Microbiology, Royal Free London NHS Foundation Trust, London NW3 2QG, UK

[3]Gastrointestinal Bacteria Reference Unit, Public Health England, 61 Colindale Avenue, London NW9 5HT, UK

Received 14 March 2014
Accepted 12 June 2014

The implementation of direct testing of clinical faecal specimens for gastrointestinal (GI) pathogens by PCR offers a sensitive and comprehensive approach for the detection of Shiga toxin-producing *Escherichia coli* (STEC). The introduction of a commercial PCR assay, known as GI PCR, for the detection of GI pathogens at three frontline hospital laboratories in England between December 2012 and December 2013 led to a significant increase in detection of STEC other than serogroup O157 (non-O157 STEC). In 2013, 47 isolates were detected in England, compared with 57 in the preceding 4 years (2009–2012). The most common non-O157 STEC serogroup detected was O26 (23.2 %). A total of 47 (47.5 %) STEC isolates had *stx2* only, 28 (28.3 %) carried *stx1* and *stx2*, and the remaining 24 (24.2 %) had *stx1* only. Stx2a (64.0 %) was the most frequently detected Stx2 subtype. The *eae* (intimin) gene was detected in 52 (52.5 %) non-O157 STEC isolates. Six strains of STEC O104 had *aggR*, but this gene was not detected in any other STEC serogroups in this study. Haemolytic ureamic syndrome was significantly associated with STEC strains possessing *eae* [odds ratio (OR) 5.845, P=0.0235] and/or *stx2a* (OR 9.56, P=0.0034) subtypes. A matched case–control analysis indicated an association between non-O157 STEC cases and contact with farm animals. Widespread implementation of the PCR approach in England will determine the true incidence of non-O157 STEC infection, highlight the burden in terms of morbidity and mortality, and facilitate the examination of risk factors to indicate whether there are niche risk exposures for particular strains.

INTRODUCTION

Shiga toxin-producing *Escherichia coli* (STEC), also known as verocytotoxin-producing *E. coli* (VTEC), is defined by the presence of the *stx* genes (*stx1*, *stx2* or both). There are more than 400 different serotypes of STEC and over 100 of these are known to cause symptoms of gastrointestinal (GI) disease in humans, including severe bloody diarrhoea and haemolytic uraemic syndrome (HUS) (Scheutz & Strockbine, 2005). Previous studies have indicated that the presence of *stx2*, specifically the *stx2a* subtype, is more frequently associated with severe disease (Ethelberg *et al.*, 2004; Persson *et al.*, 2007; Luna-Gierke *et al.*, 2014). Many STEC associated with human disease also have the intimin-encoding gene *eae* (*E. coli* attaching and effacing), located on a pathogenicity island called the locus of enterocyte effacement (LEE), and associated with the intimate attachment of the bacteria to the human gut mucosa (Frankel *et al.*, 2001). Recently, strains of STEC that do not have the *eae* gene but carry a plasmid encoding *aggR*, associated with the enteroaggregative *E. coli* group, have also been associated with causing HUS (Frank *et al.*, 2011; Dallman *et al.*, 2012).

In England, national protocols for the detection of STEC are specific for STEC serogroup O157 and focus on the isolation

Abbreviations: ESQ, enhanced surveillance questionnaire; GBRU, Gastrointestinal Bacteria Reference Unit; GEZI, Gastrointestinal, Emerging and Zoonotic Infections Department; GI, gastrointestinal; HUS, haemolytic uraemic syndrome; NESSS, National Enhanced STEC Surveillance System; OR, odds ratio; PHE, Public Health England; PHEC, Public Health England Centre; STEC, Shiga toxin-producing *Escherichia coli*; VTEC, verocytotoxin-producing *Escherichia coli*.

on selective media of non-sorbitol fermenting colonies of *E. coli* agglutinating with antisera to the O157 antigen (http://www.hpa.org.uk/ProductsServices/MicrobiologyPathology/UKStandardsForMicrobiologyInvestigations/TermsOfUseForSMIs/AccessToUKSMIs/SMIBacteriology/smiB30InvestigationofFaecalSpecimensforEnteric/). For cases of severe disease consistent with STEC infection where O157 is not isolated at the frontline hospital laboratory, it is recommended that faecal specimens are referred to the Gastrointestinal Bacteria Reference Unit (GBRU) for re-testing for the presence of both O157 and non-O157 STEC [Public Health England (PHE) STEC Operational Guidelines, http://www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1279889252950]. Over the last 10 years, approximately 15–20 isolates of non-O157 STEC (around 1.5 % of all STEC) were identified annually from faecal specimens referred to the GBRU using PCR and culture (Jenkins *et al.*, 2012). In other countries, non-O157 STEC have been identified as a significant cause of GI disease, including HUS, and their incidence may exceed that of STEC O157 (Tozzi *et al.*, 2003; Gould *et al.*, 2013; Preußel *et al.*, 2013). However, in England the bias towards detection of non-O157 STEC from patients with severe disease only, means that the true incidence and pathogenic potential of non-O157 STEC is unknown.

Between December 2012 and December 2013, three frontline hospital laboratories in England implemented commercial PCR assays targeting GI pathogens, including STEC, as a first-line diagnostic detection method (referred to as GI PCR). All faecal specimens submitted to the three frontline laboratories were tested regardless of the patient's age or symptoms. One laboratory used the assay throughout 2013 (12 months of the study period), while the other two laboratories implemented the assay in November 2013 (2 months of the study period). Two laboratories used the EntericBio assay and one used the BD Max assay. An increasing number of laboratories in England report plans to adopt a PCR approach in 2014 (GBRU in-house data). This move towards a more sensitive and comprehensive molecular approach to GI-pathogen diagnostics has had a significant impact on the detection of non-O157 STEC and led to an increase in the number of faecal specimens referred to the GBRU. Consequently, there has been an increase in the number of strains of non-O157 STEC detected. The aim of this report is to describe the characteristics of non-O157 strains detected between 2009 and 2013, and to examine their pathogenicity traits alongside the available epidemiological data.

METHODS

Microbiological characterization. Faecal specimens from patients suspected of having STEC infection (those with symptoms of HUS and severe bloody diarrhoea) that were culture negative for STEC O157, and more recently those specimens testing positive for the *stx* genes using commercial GI-pathogen PCR assays (GI PCR), were submitted to the GBRU for further testing (referred to as GBRU PCR) (Table 1). Briefly, DNA extracted from faecal culture broths was tested using a real-time PCR targeting *stx1*, *stx2*, *eae* (intimin) and O157 *rfbE* according to the method of the EU Reference Laboratory for *E. coli* (Jenkins *et al.*, 2012). For faecal specimens positive for *stx*, 20 colonies were picked from bacterial growth on MacConkey or sorbitol MacConkey agar plates and re-tested using the same PCR. Those colonies harbouring the *stx* genes were identified biochemically as *E. coli* using appropriate sugars and other metabolites in glass tubes. Positive reactions were observed by noting a change in the colour of the media, and profiles were analysed as described by Scheutz & Strockbine (2005). Strains were serotyped using antisera raised in rabbits to the *E. coli* somatic O antigens (Gross & Rowe, 1985). Each isolate was further characterized using additional PCRs targeting *aggR* (http://www.iss.it/binary/vtec/cont/Geraldine_Smith.pdf) and the *stx2* subtyping genes (Scheutz *et al.*, 2012).

Epidemiology. Local laboratories report presumptive isolates of STEC directly to PHE Centres (PHECs). Each PHEC arranges for a standard enhanced surveillance questionnaire (ESQ) to be administered to patients in a timely manner as part of routine public health follow-up. The ESQ collects data in the following categories: demographic details; risk status; clinical condition (including progression to HUS);

Table 1. Overview of the process of isolating non-O157 STEC in the GBRU from referred faecal specimens from frontline laboratories

	Frontline laboratories not using GI PCR	Frontline laboratories using GI PCR*
	Send faecal specimen to the GBRU if:	
At the frontline hospital laboratory	Symptoms indicate STEC infection (HUS and/or bloody diarrhoea) and stool culture negative for *E. coli* O157	Symptoms indicate STEC infection (HUS and/or bloody diarrhoea) and stool culture negative for *E. coli* O157 or faecal specimen PCR positive for *stx* genes
At the GBRU	GBRU PCR for *stx1*, *stx2*, *eae*, O157 *rfbE* genes on faecal enrichment cultures† and 10–20 colonies isolated on the agar plates; *stx* positive isolates are identified biochemically as *E. coli* and serotyped (O antigen with rabbit antisera), and virulence typing by PCR for *aggR* and *stx2* subtypes is performed	

*Two frontline laboratories used the EntericBio Realtime Gastro Panel 1 commercial assay and one used the BD MAX Enteric Bacterial Panel commercial assay. Both methods include a kit-based DNA extraction protocol, and DNA was extracted directly from the faecal specimen. All faecal specimens submitted to the three frontline laboratories were tested regardless of patient's age or symptoms.
†PCR performed as previously described by Jenkins *et al.* (2012).

household or other close contact details; exposures including travel, food and water consumption, contact with animals and environmental factors; case classification; and outbreak status. Completed questionnaires are forwarded for inclusion in the National Enhanced STEC Surveillance System (NESSS), which is managed by the PHE Gastrointestinal, Emerging and Zoonotic Infections Department (GEZI).

Epidemiological case definitions. A primary case was defined as follows: a symptomatic case with no history of close contact with a confirmed case in the 7 days prior to onset of illness. A secondary case was defined as follows: a case with a date of onset that is more than 4 days after the primary case or where transmission is believed to be through exposure to a primary case. An asymptomatic case was defined as follows: a person identified through contact screening procedures, with no symptoms consistent with STEC infection.

Data analyses. Comparisons were made between groups and Fisher's exact test was used to assess statistical significance. Mantel–Haenzel odds ratios (ORs) were calculated to assess the relative risk of developing HUS for different groups. A matched case–control analysis was undertaken to test the hypothesis that there were no differences in risk exposures between O157 and non-O157 STEC cases reported over the period 1 January 2009 to 31 December 2013. Cases were matched to controls on the basis of age, gender and PHE region. A matched design was chosen to control for both the different testing methods at different laboratories (the laboratories using PCR over the study period were in the south of England serving two PHE regions) and any differences in behaviour that may be associated with different age groups and genders. Univariable conditional logistic regression was performed to assess for statistically significant differences between the two groups. Exposure variables at a significance level of $\leqslant 0.2$ were included in a multivariable model. Multivariable conditional logistic regression was performed using a step-wise procedure and likelihood ratio tests undertaken to assess best fit for the model. For all tests a P value of $\leqslant 0.05$ was deemed statistically significant. All statistical analyses were performed using STATA version 12.0 software (Stata Corps, Texas).

RESULTS

Non-O157 serogroups

Between 1 January 2009 and 31 December 2013, a total of 97 non-O157 STEC isolates from 84 cases were reported in England, representing 34 different *E. coli* serotypes (Table 2). This compared with 4330 cases of STEC O157 infection reported during the same period. The most common non-O157 STEC serogroup detected was serogroup O26 with 21 isolates (21.6 %). There were six strains of STEC O104 from cases epidemiologically linked to the outbreak in Germany in 2011 (Frank *et al.*, 2011). Four isolates of STEC O145 (two cases reported recent travel to Ireland, one had travelled to Egypt and one had not travelled abroad in the exposure period) were reported and there were three isolates of STEC O117 associated with travellers returning from exotic locations, as described previously (Dallman *et al.*, 2013). Seventeen isolates did not agglutinate with any of the antisera in the serotyping scheme raised to the known *E. coli* serogroups and were designated 'O unidentifiable'; seven did not express the O antigen, and therefore could not be serotyped; these were designated 'rough'.

Three frontline hospital laboratories implemented the GI PCR approach during the study period: one in December 2012 and two in November 2013. This resulted in a significant increase in the detection of non-O157 STEC ($P<0.001$). In 2013, 42 (42.4 %) non-O157 STEC isolates were detected compared with a total of 57 in the preceding 4 years (2009–2012). In total, 22 cases were detected by the three frontline laboratories using GI PCR. One hospital laboratory in London, the region with the lowest STEC incidence in England, reported 19 confirmed non-O157 cases during the first 12 months of adopting the GI PCR approach, while only 5 STEC O157 isolates were reported over that same period. This laboratory had previously reported just three non-O157 cases between 2004 and 2012 (GBRU in-house data). The other two laboratories used GI PCR for only 2 months during the study period and reported no STEC O157 isolates during that time.

Virulence traits and clinical symptoms

Three-quarters of non-O157 isolates (73/97) carried *stx2*: 45 (46.4 %) had *stx2* only, 28 (28.3 %) carried *stx1* and *stx2*, and the remaining 24 (24.2 %) harboured *stx1* only (Table 2). The most frequently detected *stx2* subtype was *stx2a* (63.0 %, $n=46$ isolates). Forty isolates carried *stx2a* only, including the six STEC O104 isolates linked to the outbreak in Germany in 2011 and six STEC O26 isolates. *stx2b* only was detected in 14 isolates (Table 2). Other subtypes were rare, including one isolate that had *stx2d* only and two harbouring *stx2g* only (Table 2). The nine isolates that were originally positive by GBRU PCR for *stx2* tested negative with the *stx2* subtyping PCR in this study. On repeat testing using the GBRU PCR, these strains tested negative for *stx2* and it is likely that they had lost the *stx*-encoding phage during storage. The *eae* gene was detected in 50 (51.5 %) non-O157 STEC isolates (Table 2). Six strains of STEC O104 had *aggR*; however, *aggR* was not detected in any other strains in this study.

The 97 non-O157 STEC isolates were associated with 84 individual cases reported to the NESSS. Three cases had multiple serotypes of non-O157 STEC, including one case reporting travel to Egypt infected with both serogroups O145 and O26, one with serogroup O117 and a rough isolate, and one case with serogroups O26, O45 and O91. Nine isolates were from patients enrolled in a large study of infectious intestinal disease (Tam *et al.*, 2012) and were not reported through the NESSS.

Just over half ($n=46$, 54.8 %) of cases were female and almost a third ($n=30$, 35.7 %) were children aged under 15 years. The same gender difference was reported amongst STEC O157 cases (55.2 %), but a higher proportion (42.5 %) of children was observed with STEC O157 infection. For four cases, ESQs were lost to follow-up, and an additional six cases (screened because they were contacts of symptomatic cases) had asymptomatic infection. Among the 74 remaining cases (69 primary and 5 secondary), diarrhoea was reported in 66 (89.2 %) cases, including bloody diarrhoea in 35

Table 2. Microbiological characteristics of 97 non-O157 STEC isolates from 88 cases detected in England, 2009–2013

Serogroup	*stx* type	*stx1a*	*stx1c*	*stx1d*	*stx2a*	*stx2b*	*stx2c*	*stx2d*	*stx2e*	*stx2f*	*stx2g*	*eae*	*aggR*
O103 (2)	1	+	−	−	NA	NA	NA	NA	NA	NA	NA	+	−
O104 (6)	2	NA	NA	NA	+	−	−	−	−	−	−	−	+
O111 (2)	1, 2	+	−	−	+	−	−	−	−	−	−	+	−
O113 (1)	1, 2	−	+	−	−	+	−	−	−	−	−	−	−
O113 (1)	2	NA	NA	NA	+	−	−	−	−	−	−	−	−
O113 (1)	2	NA	NA	NA	+	−	+	+	−	−	−	+	−
O117 (3)	1	+	−	−	NA	NA	NA	NA	NA	NA	NA	−	−
O118 (1)	1	+	−	−	NA	NA	NA	NA	NA	NA	NA	+	−
O121 (1)	2	NA	NA	NA	+	−	−	−	−	−	−	+	−
O125 (1)	2	NA	NA	NA	−	−	−	−	−	−	−	+	−
O126 (1)	1, 2	−	+	−	−	+	−	−	−	−	−	−	−
O128 (1)	2	NA	NA	NA	−	+	−	−	−	−	−	−	−
O132 (1)	2	NA	NA	NA	−	−	−	−	−	−	−	+	−
O145 (4)	2	NA	NA	NA	+	−	−	−	−	−	−	+	−
O146 (2)	1	−	+	−	NA	NA	NA	NA	NA	NA	NA	−	−
O146 (1)	2	NA	NA	NA	−	+	−	−	−	−	−	−	−
O156 (2)	1	−	+	−	NA	NA	NA	NA	NA	NA	NA	+	−
O159 (1)	2	NA	NA	NA	+	−	−	+	−	−	−	−	−
O161 (1)	2	NA	NA	NA	−	−	−	−	−	−	−	−	−
O165 (1)	2	NA	NA	NA	+	−	+	−	−	−	−	+	−
O166 (1)	1, 2	−	+	−	−	+	−	−	−	−	−	−	−
O172 (1)	2	NA	NA	NA	+	−	−	−	−	−	−	+	−
O177 (1)	2	NA	NA	NA	+	−	+	−	−	−	−	+	−
O181 (1)	1, 2	+	−	−	+	−	+	−	−	−	−	−	−
O182 (1)	1	+	−	−	NA	NA	NA	NA	NA	NA	NA	−	−
O186 (1)	1, 2	+	−	−	−	+	−	−	−	−	−	+	−
O186 (1)	1, 2	+	−	−	+	−	−	−	−	−	−	−	−
O26 (2)	1	+	−	−	NA	NA	NA	NA	NA	NA	NA	−	−
O26 (11)	1, 2	+	−	−	+	−	−	−	−	−	−	+	−
O26 (8)	2	NA	NA	NA	+	−	−	−	−	−	−	+	−
O27 (1)	2	NA	NA	NA	−	−	−	−	−	−	−	+	−
O29 (1)	2	NA	NA	NA	−	−	−	−	−	−	−	−	−
O45 (1)	2	NA	NA	NA	−	−	−	−	−	−	−	−	−
O49 (1)	2	NA	NA	NA	−	−	−	−	−	−	−	−	−
O71 (1)	1, 2	+	−	−	+	−	+	−	−	−	−	+	−
O76 (1)	1	−	+	−	NA	NA	NA	NA	NA	NA	NA	−	−
O8 (1)	2	NA	NA	NA	+	−	−	+	−	−	−	−	−
O87 (1)	2	NA	NA	NA	−	+	−	−	−	−	−	−	−
O91 (1)	1, 2	+	−	−	−	+	−	−	−	−	−	−	−
O91 (1)	2	NA	NA	NA	+	−	−	−	−	−	−	+	−
O92 (1)	2	NA	NA	NA	−	−	−	−	−	−	−	+	−
O Rough (3)	1	+	−	−	NA	NA	NA	NA	NA	NA	NA	−	−
O Rough (1)	1, 2	+	−	−	−	−	−	−	−	−	+	−	−
O Rough (1)	1, 2	−	+	−	−	+	−	−	−	−	−	−	−
O Rough (1)	2	NA	NA	NA	−	−	−	−	−	−	−	−	−
O Rough (1)	2	NA	NA	NA	−	+	−	−	−	−	−	−	−
O* (3)	1	+	−	−	NA	NA	NA	NA	NA	NA	NA	−	−
O* (3)	1	−	+	−	NA	NA	NA	NA	NA	NA	NA	+	−
O* (1)	1, 2	+	−	−	−	−	−	−	−	−	+	+	−
O* (1)	1, 2	+	−	−	−	−	+	−	−	−	−	−	−
O* (2)	1, 2	+	−	−	−	+	−	−	−	−	−	−	−
O* (1)	1, 2	+	−	−	+	−	−	−	−	−	−	−	−
O* (2)	2	NA	NA	NA	−	−	−	+	−	−	−	+	−
O* (2)	2	NA	NA	NA	−	+	−	−	−	−	−	+	−

Table 2. cont.

Serogroup	*stx* type	*stx1a*	*stx1c*	*stx1d*	*stx2a*	*stx2b*	*stx2c*	*stx2d*	*stx2e*	*stx2f*	*stx2g*	*eae*	*aggR*
O* (1)	2	NA	NA	NA	+	−	−	−	−	−	−	−	−
O* (1)	2	NA	NA	NA	+	−	+	−	−	−	−	−	−

+, Present; −, not present; NA, not applicable.
*Unidentifiable.

(47.2 %) cases. Abdominal pain and vomiting were reported by 47 (63.5 %) and 29 (39.2 %) cases, respectively. These are comparable to symptoms reported by STEC O157 cases with the exception of bloody diarrhoea and abdominal pain, which are reported by a higher proportion of STEC O157 cases (61.0 and 79.2 %, respectively). A total of 28 (37.8 %) cases were hospitalized and a quarter developed HUS (n=18, 24.3 %). Hospitalization rates were similar (34.3 %) amongst STEC O157 cases, but reported development of HUS only occurs in approximately 5 % of infections (Lynn *et al.*, 2005; GEZI and GBRU in-house data). Two-thirds (12/18) of HUS cases were children aged under 15; children were at significantly greater odds of developing HUS compared with adults in this study population (OR 4.40, P=0.005). There was no significant difference (OR 1.14, P=0.800) in development of HUS by gender.

The development of HUS was significantly associated with non-O157 STEC strains possessing *eae* (OR 5.845, P=0.0235) and/or *stx2a* (OR 9.56, P=0.0034) subtypes, when compared with all other strains. All 18 HUS cases (24 isolates) were infected with at least one strain of non-O157 STEC carrying *stx2*, and most strains (n=19, 86.4 %) possessed the *stx2a* subtype (Table 3) and encoded intimin. Two cases that developed HUS had multiple non-O157 STEC serotypes detected. The first had STEC serotypes O26, O45 and O91. Although it is not possible to determine whether one strain was more significant in the progression to HUS for this case, both STEC O26 and O91 harboured *stx2a* and *eae* whereas the STEC O45 strain was *stx2a* and *eae* negative. The other case had STEC O26, harbouring *eae* and *stx2a*, and STEC O145 carrying *eae* and *stx2a*.

Unsurprisingly, disease was more severe among STEC cases identified by referral of the faecal specimen to the GBRU following the STEC guidelines (i.e. recommending referral of cases with HUS and bloody diarrhoea) than in those initially detected through the routine testing of all faecal specimens (regardless of severity of symptoms) at the frontline laboratory by PCR. A significantly smaller proportion of cases detected by PCR had strains encoding *stx2a* (P=0.001) and only one-third of strains detected by PCR had *eae* compared with two-thirds of those referred to the GBRU following the STEC guidelines (P=0.01). Among 22 cases detected by PCR, 3 reported bloody diarrhoea and 2 were hospitalized. None of the cases detected by PCR developed HUS.

Non-O157 STEC risk exposures

A matched case–control analysis was undertaken to compare reported risk exposures amongst 69 primary, symptomatic, non-O157 STEC cases and 2300 primary STEC O157 cases. The final multivariable model indicated significantly increased odds among non-O157 STEC cases of contact with farm animals, including cattle, sheep, pigs, goats, horses and poultry (Table 4). Contact with farm animals included visiting petting farms, living on or having access to a farm and incidental contact in the countryside. Cases of STEC O157 had greater odds than non-O157 STEC cases of freshwater swimming, handling or consuming raw beef and shopping at independent retailers, including butchers, grocers, delicatessens and markets.

DISCUSSION

The implementation of GI PCR at just one frontline laboratory, for the greater part of the year, significantly increased the detection of non-O157 STEC cases in England. Notably, almost four times as many non-O157 isolates than O157 isolates were detected at that laboratory during that time, indicating the extent to which non-O157 STEC is under-ascertained through traditional culture methods. Although the numbers in this study were low in comparison with STEC O157 cases, collection of clinical data through the NESSS coupled with strain characterization, including detection of genes associated with adherence to the gut mucosa and *stx2* subtyping, facilitated the assessment of virulence traits belonging to different strains of non-O157 STEC.

Elsewhere in Europe, STEC serogroups O26, O103, O111 and O145, along with STEC O157, are regarded as the most common STEC serogroups associated with severe disease and outbreaks (Gerber *et al.*, 2002; Beutin *et al.*, 2004; Brooks *et al.*, 2005; Carroll *et al.*, 2005). While all four of these non-O157 STEC serogroups were identified in this dataset, with STEC O26 being the most common (Jenkins *et al.*, 2008; Chase-Topping *et al.*, 2012; Bielaszewska *et al.*, 2013), a variety of other serogroups were also found. The wide range of serogroups and the fact that 24 % of the strains could not be serogrouped (strains designated rough and unidentifiable) using the current serotyping scheme suggests that reliance on serogroup alone to indicate pathogenicity would be misleading.

Table 3. Shiga-toxin types and development of HUS among non-O157 cases in England, 2009–2013

Shiga toxin type or subtype	*eae* +				*eae*−			
	No. of isolates*	No. of isolates associated with bloody diarrhoea (%)	No. of isolates of associated with hospitalization (%)	No of isolates associated with HUS (%)†	No. of isolates*	No. of isolates associated with bloody diarrhoea (%)	No. of isolates associated with hospitalization (%)	No of isolates associated with HUS (%)†
stx1 only	11	6 (54.5)	2 (18.2)	0	11	3 (27.3)	1 (9.1)	0
stx2a	32	17 (53.1)	21 (65.6)	17 (53.1)	7	6 (100.0)	4 (66.7)	3 (50.0)
stx2a, stx2c	4	2 (50.0)	2 (50.0)	1 (25.0)	1	0	0	0
stx2b	0	0	0	0	10	1 (10.0)	0	0
stx2c	0	0	0	0	1	1 (100.0)	0	0
stx2d	0	0	0	0	1	0	0	0
stx2g	0	0	0	0	1	0	0	0
stx2 no *stx2* subtype	3	1 (33.3 %)	0	0	6	2 (33.3)	1 (16.7)	1 (16.7)
Total	50	26 (52.0)	24 (48.0)	18 (36.0)	38	13 (34.2)	6 (15.8)	4 (10.5)

*Excludes nine isolates detected in the infectious intestinal disease II study (Tam *et al.*, 2012) as these clinical data are unavailable.
†Two HUS cases had multiple strains of non-O157 isolated.

Furthermore, non-O157 STEC are a heterogeneous group, and the ability of specific serogroups to cause severe illness in humans is likely to be determined by toxin subtype and additional attachment genes involved in pathogenesis. As described in other studies, intimin and *stx2a* subtypes were all significantly associated with progression to HUS (Gerber *et al.*, 2002; Beutin *et al.*, 2004; Brooks *et al.*, 2005; Ethelberg *et al.*, 2004; Persson *et al.*, 2007; Luna-Gierke *et al.*, 2014). It is important to note, however, that not all cases of HUS in this study were infected with strains exhibiting these properties. Other bacterial virulence traits, such as *aggR*, are associated with HUS, and human host factors, such as age, may also impact on the likelihood of disease progression. It is well documented that children are at greater risk of developing STEC–HUS, and our findings were consistent with this (Lynn *et al.*, 2005). There are also likely to be additional, as yet unproven, bacteria and host factors associated with disease progression.

Examination of reported risk exposures among non-O157 STEC cases revealed significantly increased reporting of farm animals as compared with O157 cases, while reporting of freshwater swimming (i.e. in rivers and lakes), shopping at independent retailers and handling or consuming raw beef was less likely. However, it is difficult to draw conclusions from these data as the numbers of non-O157 STEC in this study are low and because the non-O157 STEC are a heterogeneous group. Furthermore, the sensitivity of the ESQ is limited to the risk exposures included on the questionnaire, and is subject to both interviewer and recall biases. Isolation and confirmation of non-O157 STEC by the GBRU takes longer than confirmation of STEC O157, and notification to PHE health protection teams and subsequent public health follow-up, including administering the ESQ, is delayed in comparison. Thus, for non-O157 STEC, epidemiological data may be less reliable due to the time-lag affecting patient recall. The introduction of PCR at the frontline offers an opportunity to reduce this lag and subsequent recall bias if public health follow-up can be taken on the basis of PCR detection at the frontline.

Direct testing of all faecal specimens and the subsequent increase in the number of STEC cases raised questions regarding the clinical and public health impact of non-O157 STEC in England. The clinical symptoms, and therefore the public health significance, of non-O157 STEC cases in this study were severe, with a considerably higher proportion of cases developing HUS than observed for STEC O157. However, these data are biased by selective referral of specimens to the GBRU from cases with severe disease. Moving forward, cases identified at the frontline laboratories by direct testing of all faecal specimens using the GI PCR approach will readdress this bias. Widespread implementation of the GI PCR approach, and subsequent follow-up via the NESSS, will enhance our ability to determine the true incidence of non-O157 STEC infection, the burden in terms of morbidity and mortality, and whether there are niche risk exposures for particular strains.

Table 4. Multivariable conditional logistic regression analysis of risk exposures of non-O157 STEC cases compared to O157 STEC cases in England, 2009–2013

Risk exposure*	O157 STEC cases (n=2300)		Non-O157 STEC cases (n=67)		OR†	P-value	95 % confidence interval
	Exposed (%)	Unexposed (%)	Exposed (%)	Unexposed (%)			
Freshwater swimming	332 (14.4)	1968 (85.6)	3 (4.4)	66 (95.6)	0.217	0.012	0.066–0.715
Handling/consuming raw beef	317 (13.8)	1983 (86.2)	1 (1.5)	66 (98.5)	0.124	0.041	0.017–0.912
Shopping at independent retailers	720 (31.3)	1580 (68.7)	8 (11.6)	61 (88.4)	0.327	0.023	0.152–0.703
Contact with farm animals	293 (12.7)	2007 (87.3)	16 (23.2)	53 (76.8)	3.291	<0.001	1.69–6.41

*Only risk exposures included in the final conditional logistic model are presented.
†ORs adjusted for all other variables in the model.

ACKNOWLEDGEMENTS

We would like to thank all our colleagues from the National Health Service and PHE frontline laboratories who refer specimens to the GBRU and those who have taken part in the PCR debate so far. We are especially grateful to Raj Virdee and Debra Clark at Wexham Park Hospital, Clare Reynolds at Brighton Hospital, Kevin Carroll and Lisa Harvey-Vince at the Kent, Surrey and Sussex Health Protection Team, Claire Bayntun and Yvonne Young at the South West London Health Protection Team, and Rohini Manual and Derren Ready at the Public Health Laboratory, London. We would also like to acknowledge Marie Chattaway, Vivienne do Nascimento, Michael Wright and Yoshini Taylor for their hard work and microbiological expertise at the GBRU. Thanks to Kirsten Glen, Natalie Adams and Radha Patel for their contributions to the NESSS in England, and to all public health practitioners who report to the scheme.

REFERENCES

Beutin, L., Krause, G., Zimmermann, S., Kaulfuss, S. & Gleier, K. (2004). Characterization of Shiga toxin-producing *Escherichia coli* strains isolated from human patients in Germany over a 3-year period. *J Clin Microbiol* **42**, 1099–1108.

Bielaszewska, M., Mellmann, A., Bletz, S., Zhang, W., Köck, R., Kossow, A., Prager, R., Fruth, A., Orth-Höller, D. & other authors (2013). Enterohemorrhagic *Escherichia coli* O26:H11/H-: a new virulent clone emerges in Europe. *Clin Infect Dis* **56**, 1373–1381.

Brooks, J. T., Sowers, E. G., Wells, J. G., Greene, K. D., Griffin, P. M., Hoekstra, R. M. & Strockbine, N. A. (2005). Non-O157 Shiga toxin-producing *Escherichia coli* infections in the United States, 1983-2002. *J Infect Dis* **192**, 1422–1429.

Carroll, A. M., Gibson, A. & McNamara, E. B. (2005). Laboratory-based surveillance of human verocytotoxigenic *Escherichia coli* infection in the Republic of Ireland, 2002-2004. *J Med Microbiol* **54**, 1163–1169.

Chase-Topping, M. E., Rosser, T., Allison, L. J., Courcier, E., Evans, J., McKendrick, I. J., Pearce, M. C., Handel, I., Caprioli, A. & other authors (2012). Pathogenic potential to humans of bovine *Escherichia coli* O26, Scotland. *Emerg Infect Dis* **18**, 439–448.

Dallman, T., Smith, G. P., O'Brien, B., Chattaway, M. A., Finlay, D., Grant, K. A. & Jenkins, C. (2012). Characterization of a verocytotoxin-producing enteroaggregative *Escherichia coli* serogroup O111:H21 strain associated with a household outbreak in Northern Ireland. *J Clin Microbiol* **50**, 4116–4119.

Dallman, T., Cross, L., Bishop, C., Perry, N., Olesen, B., Grant, K. A. & Jenkins, C. (2013). Whole genome sequencing of an unusual serotype of Shiga toxin-producing *Escherichia coli*. *Emerg Infect Dis* **19**, 1302–1304.

Ethelberg, S., Olsen, K. E. P., Scheutz, F., Jensen, C., Schiellerup, P., Engberg, J., Petersen, A. M., Olesen, B., Gerner-Smidt, P. & Mølbak, K. (2004). Virulence factors for hemolytic uremic syndrome, Denmark. *Emerg Infect Dis* **10**, 842–847.

Frank, C., Werber, D., Cramer, J. P., Askar, M., Faber, M., an der Heiden, M., Bernard, H., Fruth, A., Prager, R. & other authors (2011). Epidemic profile of Shiga-toxin-producing *Escherichia coli* O104:H4 outbreak in Germany. *N Engl J Med* **365**, 1771–1780.

Frankel, G., Phillips, A. D., Trabulsi, L. R., Knutton, S., Dougan, G. & Matthews, S. (2001). Intimin and the host cell–is it bound to end in Tir(s)? *Trends Microbiol* **9**, 214–218.

Gerber, A., Karch, H., Allerberger, F., Verweyen, H. M. & Zimmerhackl, L. B. (2002). Clinical course and the role of Shiga toxin-producing *Escherichia coli* infection in the hemolytic-uremic syndrome in pediatric patients, 1997-2000, in Germany and Austria: a prospective study. *J Infect Dis* **186**, 493–500.

Gould, L. H., Mody, R. K., Ong, K. L., Clogher, P., Cronquist, A. B., Garman, K. N., Lathrop, S., Medus, C., Spina, N. L. & other authors (2013). Increased recognition of non-O157 Shiga toxin-producing *Escherichia coli* infections in the United States during 2000-2010: epidemiologic features and comparison with *E. coli* O157 infections. *Foodborne Pathog Dis* **10**, 453–460.

Gross, R. J. & Rowe, B. (1985). Serotyping of *Escherichia coli*. In *The Virulence of Escherichia coli: Reviews and Methods* (Special Publication of the Society for General Microbiology no. 13), pp 345–360. Edited by M. Sussman. London: Academic Press.

Jenkins, C., Evans, J., Chart, H., Willshaw, G. A. & Frankel, G. (2008). *Escherichia coli* serogroup O26–a new look at an old adversary. *J Appl Microbiol* **104**, 14–25.

Jenkins, C., Lawson, A. J., Cheasty, T. & Willshaw, G. A. (2012). Assessment of a real-time PCR for the detection and characterization of verocytotoxigenic *Escherichia coli*. *J Med Microbiol* **61**, 1082–1085.

Luna-Gierke, R. E., Griffin, P. M., Gould, L. H., Herman, K., Bopp, C. A., Strockbine, N. & Mody, R. K. (2014). Outbreaks of non-O157 Shiga toxin-producing *Escherichia coli* infection: USA. *Epidemiol Infect*, 1–11. doi:10.1017/S0950268813003233 [Epub ahead of print].

Lynn, R. M., O'Brien, S. J., Taylor, C. M., Adak, G. K., Chart, H., Cheasty, T., Coia, J. E., Gillespie, I. A., Locking, M. E. & other authors (2005). Childhood hemolytic uremic syndrome, United Kingdom and Ireland. *Emerg Infect Dis* **11**, 590–596.

Persson, S., Olsen, K. E., Ethelberg, S. & Scheutz, F. (2007). Subtyping method for *Escherichia coli* Shiga toxin (verocytotoxin) 2

variants and correlations to clinical manifestations. *J Clin Microbiol* **45**, 2020–2024.

Preußel, K., Höhle, M., Stark, K. & Werber, D. (2013). Shiga toxin-producing *Escherichia coli* O157 is more likely to lead to hospitalization and death than non-O157 serogroups–except O104. *PLoS ONE* **8**, e78180.

Scheutz, F., Teel, L. D., Beutin, L., Piérard, D., Buvens, G., Karch, H., Mellmann, A., Caprioli, A., Tozzoli, R. & other authors (2012). Multicenter evaluation of a sequence-based protocol for subtyping Shiga toxins and standardizing Stx nomenclature. *J Clin Microbiol* **50**, 2951–2963.

Scheutz, M. & Strockbine, N. A. (2005). Genus I. *Escherichia* Castellani and Chalmers 1919, 941T^{AL}. In *Bergey's Manual of Systematic Bacteriology*, vol. 2, pp. 607–624. Edited by G. M. Garrity, D. J. Brenner, N. R. Krieg & J. T. Staley. New York: Springer.

Tam, C. C., Rodrigues, L. C., Viviani, L., Dodds, J. P., Evans, M. R., Hunter, P. R., Gray, J. J., Letley, L. H., Rait, G. & other authors (2012). Longitudinal study of infectious intestinal disease in the UK (IID2 study): incidence in the community and presenting to general practice. *Gut* **61**, 69–77.

Tozzi, A. E., Caprioli, A., Minelli, F., Gianviti, A., De Petris, L., Edefonti, A., Montini, G., Ferretti, A., De Palo, T. & other authors (2003). Shiga toxin-producing *Escherichia coli* infections associated with hemolytic uremic syndrome, Italy, 1988-2000. *Emerg Infect Dis* **9**, 106–108.

Paper 6: The utility and public health implications of PCR and whole genome sequencing for the detection and investigation of an outbreak of Shiga toxin-producing *Escherichia coli* serogroup O26:H11.

Epidemiol. Infect. (2015), 143, 1672–1680. © Cambridge University Press 2014
doi:10.1017/S0950268814002696

The utility and public health implications of PCR and whole genome sequencing for the detection and investigation of an outbreak of Shiga toxin-producing Escherichia coli serogroup O26:H11

T. J. DALLMAN[1] †, L. BYRNE[2] †, N. LAUNDERS[2], K. GLEN[2], K. A. GRANT[1] AND C. JENKINS[1] *

[1] Gastrointestinal and Emerging Zoonotic Infections Department, Public Health England, London, UK
[2] Gastrointestinal Bacteria Reference Unit, Public Health England, London, UK

Received 9 May 2014; Final revision 30 August 2014; Accepted 19 September 2014;
first published online 15 October 2014

SUMMARY

Many serogroups of Shiga toxin-producing Escherichia coli (STEC) other than serogroup O157 (non-O157 STEC), for example STEC O26:H11, are highly pathogenic and capable of causing haemolytic uraemic syndrome. A recent increase in non-O157 STEC cases identified in England, resulting from a change in the testing paradigm, prompted a review of the current methods available for detection and typing of non-O157 STEC for surveillance and outbreak investigations. Nineteen STEC O26:H11 strains, including four from a nursery outbreak were selected to assess typing methods. Serotyping and multilocus sequence typing were not able to discriminate between the stx-producing strains in the dataset. However, genome sequencing provided rapid and robust confirmation that isolates of STEC O26:H11 associated with a nursery outbreak were linked at the molecular level, had a common source and were distinct from the other strains analysed. Virulence gene profiling of DNA extracted from a polymerase chain reaction (PCR)-positive/culture-negative faecal specimen from a case that was epidemiologically linked to the STEC O26:H11 nursery outbreak, provided evidence at the molecular level to support that link. During this study, we describe the utility of PCR and the genome sequencing approach in facilitating surveillance and enhancing the response to outbreaks of non-O157 STEC.

Key words: Bacterial typing, molecular epidemiology, outbreaks, public health microbiology, Shiga-like toxin-producing E. coli.

INTR ODUCTION

Shiga toxin-producing Escherichia coli (STEC), also known as Verocytotoxin-producing E. coli (VTEC), cause a range of symptoms from mild gastroenteritis to severe bloody diarrhoea, and about 6% of cases develop haemolytic uraemic syndrome (HUS) [1]. HUS is the most common cause of acute kidney failure in children and can be associated with cardiac and neurological complications. Strains of E. coli belonging to the STEC pathotype are defined by the presence of the phage-encoded Shiga toxin genes, stx1 and/or stx2. In England, national protocols for the detection of STEC are specific for serogroup O157 and focus on the isolation on selective media of non-sorbitol-fermenting colonies of E. coli agglutinating with antisera to the O157 antigen

* Author for correspondence: Dr C. Jenkins, Gastrointestinal Bacteria Reference Unit, Public Health England, 61 Colindale Avenue, London NW9 5EQ, UK.
(Email: claire.jenkins@phe.gov.uk)
† These authors contributed equally to this work.

(http://www.hpa-standardmethods.org.uk/). However, there are over 400 different serotypes of STEC and over 100 of these are known to be associated with severe disease in humans [2]. Non-O157 STEC are generally sorbitol-fermenting strains and, therefore, are not detected by the current national protocols.

STEC O26:H11 is the second most common STEC serogroup detected in England, after serogroup O157 [3] and is commonly isolated from patient with symptoms of gastrointestinal infection elsewhere [4]. Like STEC O157, STEC O26:H11 can harbour stx1 or stx2 or both. Prior to 1994, STEC O26:H11 in Europe almost exclusively exhibited the stx1 genotype; however, since that time there has been a steady increase in the number of strains of STEC O26:H11 isolated harbouring stx2, either alone or with stx1 [5]. This shift in genotype has been associated with strains of STEC O26:H11 that cause more severe disease [6].

In England, the guidance in the VTEC operational manual (http://www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1279889252950) recommends that faecal specimens from cases of bloody diarrhoea or HUS should be referred to Gastrointestinal Bacteria Reference Unit (GBRU) by local hospital laboratories for testing for STEC other than serogroup O157 (non-O157) [7]. About 15–20 non-O157 STEC strains were isolated each year following this testing paradigm. The selective referral of specimens from cases of the most severe disease only results in an unquantifiable under-ascertainment of non-O157 STEC in England. Recently, a number of local hospital laboratories have implemented the polymerase chain reaction (PCR) approach for the detection of gastrointestinal pathogens directly from faecal specimens using commercially available PCR assays, specifically EntericBio (Ireland) and BD Max (UK). This change in the testing paradigm has resulted in a significant increase in the number of non-O157 STEC cases identified as non-O157 STEC, previously not detected using the current national protocols described above, being detected using this approach [3]. This increase in non-O157 cases prompted a review of the current methods available at GBRU for typing isolates of non-O157 STEC to inform routine surveillance and outbreak investigations.

Multilocus variable number tandem repeat (VNTR) analysis (MLVA) is routinely used at GBRU for typing strains of STEC O157 to inform epidemiology of cases, and facilitate outbreak detection and investigation [8]. This MLVA scheme is serotype-specific and, therefore, not applicable to strains of non-O157 STEC. Currently, multilocus sequence typing (MLST) is not discriminatory enough for outbreak investigations of non-O157 STEC and provides approximately the same resolution as serotyping. Pulsed-field gel electrophoresis is the method of choice in many national reference laboratories for the molecular typing of non-O157 STEC but it is laborious and technically demanding [9] and international laboratory comparisons of fingerprint patterns are difficult (GBRU in-house data).

Recently whole genome sequencing (WGS) has been employed at Public Health England (PHE) for typing bacterial strains associated with gastrointestinal outbreaks [10, 11]. Other groups have investigated the use of metagenomics, the direct sequencing of DNA extracted from microbiologically complex samples such as faecal specimens, to identify and characterize bacterial strains without laboratory culture [12].

In this study, we describe a complex nursery outbreak of STEC O26:H11 which highlights both the pathogenic potential of this strain, and the public health issues arising from currently employed testing strategies. Strains from the outbreak and other STEC O26:H11 from the GBRU archive were used to retrospectively evaluate the use of genome sequencing, including metagenomic analysis of a PCR-positive but culture-negative faecal specimen, as a suitable molecular typing approach for non-O157 STEC outbreak investigations.

MATERIAL AND METHODS

Direct detection of STEC from faecal specimens

DNA was extracted from the faecal specimens using the QiaSymphony Automated DNA extraction platform [13]. DNA from faecal extracts was tested using real-time PCR primers and probes detecting stx1, stx2, eae (intimin) and O157rfbE [7]. For all faecal specimens positive for stx and/or eae (intimin), 10 colonies were picked from either the MacConkey or SMAC plate and retested by the same PCR. Those colonies harbouring stx genes were identified biochemically and serotyped using antisera raised in rabbits.

Epidemiological investigations

Local laboratories report presumptive isolates of STEC directly to PHE centres (PHEC) who undertake public health follow-up and risk assessment.

Table 1. Molecular and epidemiological data associated with strains of E. coli O26:H11 isolated at GBRU between 2009 and 2013

Reference no.	Stx profile	MLST	Date culture isolated	Sex/ age	Travel	Additional information*
181/09	1&2	21	Feb. 2009	F/2	No travel	Outbreak 2009
461/09	1&2	21	Feb. 2009	M/14	No travel	Outbreak 2009
460/09	1&2	21	Feb. 2009	F/0	No travel	Outbreak 2009
259/10	1&2	21	Sept. 2010	M/3	No travel	HUS
605/10	1&2	21	Sept. 2010	F/3	France	HUS
467/10	2	21	Nov. 2010	M/13	Turkey	HUS
519/11	1&2	21	July 2011	M/4	No travel	
624/12	2	21	Apr. 2012	F/13	Egypt	HUS
165/12	1&2	21	May 2012	F/42	No travel	Fatal case
483/12	2	21	July 2012	M/35	Ireland and Switzerland	Outbreak 2012
482/12	2	21	July 2012	M/4	No travel	Outbreak 2012
626/12	2	21	July 2012	M/4	No travel	Outbreak 2012
627/12	2	21	July 2012	M/2	No travel	Outbreak 2012
2270–502/12	2	21	July 2012	F/1	Ireland	
2290–502/12	—	29	July 2012	M/40	No data	
670/13	—	29	Aug. 2013	F/0	Egypt	HUS
680/13	—	29	Sept. 2013	F/3	Italy	Hospitalized with severe bloody diarrhoea
075/13	2	21	Sept. 2013	M/0	Albania	
637/13	2	21	Sept. 2013	F/7	No data	

MLST, Multilocus sequence typing; HUS, haemolytic uraemic syndrome.
* Additional information includes whether or not cases were associated with an outbreak and clinical symptoms.

Household contacts of confirmed cases are sampled and those found to be positive and belonging to risks groups are excluded from school or work until they have two negative clearance specimens taken 24–48 h apart. Each PHEC arranges for the standard enhanced surveillance questionnaire to be administered to cases in order to collect demographic details; risk status; clinical condition (including progression to HUS); household or other close contact details; laboratory results; exposures including travel, food and water consumption, contact with animals and environmental factors; case classification; outbreak/ cluster status. Completed questionnaires are forwarded for inclusion in the National Enhanced Surveillance System for STEC in England (NESSS), Gastrointestinal and Emerging Zoonotic Infections (GEZI).

WGS and analysis of strains of STEC O26:H11

DNA was extracted from 19 strains of E. coli O26 using the Wizard kit (Promega, UK). These strains were isolated from faecal specimens submitted to GBRU by colleagues at local hospital laboratories from cases with severe bloody diarrhoea or HUS between 2009 and 2013 (Table 1). Genomic DNA was fragmented and tagged for multiplexing with Nextera XT DNA Sample Preparation kits (Illumina, UK) and sequenced at the Animal Health Veterinary Laboratory Agency, Surrey, using the Illumina GAII platform with paired-end 150 bp reads. Multiplexing allowed 96 samples to be sequenced per run. Short reads were mapped to the reference STEC O26:H11 strain 11 368 (NC_011361·1) using BWA-SW [14]. The Sequence Alignment Map output from BWA was sorted and indexed to produce a Binary Alignment Map (BAM) using Samtools [15]. GATK2 [16] was used to create a variant call format (VCF) file from each of the BAMs, which were further parsed to extract only single nucleotide polymorphism (SNP) positions which were of high quality in all genomes (MQ > 30, DP > 10, GQ > 30, variant ratio >0·9) [17]. Pseudosequences of polymorphic positions were used to create maximum-likelihood trees using RaxML [18]. Spades version 2.5·1 [19] was used to produce de novo assemblies of the sequenced paired-end FASTQ files.

FASTQ sequences for these 19 isolates have been deposited in the NCBI Short Read Archive under BioProject PRJNA 259 827.

MLSTs were identified by mapping the reads against all E. coli allele variants held in the MLST database (www.mlst.ucc.ie/mlst/dbs/Ecoli) using a modification of SRST software [20].

Identification of the flagella antigen and intimin subtype were performed by BLAST [21] comparisons of the assembled genomes to a database of fliC genes for the determination of the flagella antigen and a database of eae genes for the determination of the intimin subtype.

Metagenomic analysis of a faecal specimen from a culture-negative case epidemiologically linked to the outbreak

Faecal DNA from a PCR-positive/culture-negative faecal specimen, from a case epidemiologically linked to a nursery outbreak, was extracted on the QiaSymphony Automated DNA extraction platform [13] in triplicate to achieve the required yield for DNA sequencing. Genomic DNA was fragmented and tagged for multiplexing with Nextera XT DNA Sample Preparation kits (Illumina, UK) and sequenced on an Illumina MiSeq platform. About 18·5 m high-quality paired-end 150 bp reads were generated after trimming using Trimmomatic [22]. Reads were mapped against a diarrhoegenic E. coli virulence gene panel including stx1 and stx2 for STEC, eae (encoding intimin) for STEC and enteropathogenic E. coli (EPEC) [23], aggR for enteroaggregative E. coli (EAEC) [24], ipaH for Shigella, heat-stable enterotoxin (ST) and heat-labile enterotoxin (LT) genes for enterotoxigenic E. coli (ETEC) and the following serotype-specific targets: rfbE (O157), wzx (O26), wzx (O103), wbdI (O111) and ihp1 (O145) (http://www.iss.it/vtec/index.php?lang=2&anno=2014&tipo=3), using BWA-SW. Coverage of each gene target was calculated using Samtools depth command.

FASTQ sequences of this metagenomic sample have been deposited in the NCBI Short Read Archive under BioProject PRJNA 259 829.

Metagenomic reads were also mapped to virulence factors of pathogenic bacteria database (VFDB) [25] which contains 28 966 virulence genes from 26 pathogens including common enteric genera, such as Salmonella, Listeria, Campylobacter, Shigella, Yersinia and Vibrio species.

RESULTS

Microbiology and epidemiology of the STEC O26:H11 outbreak

On 8 July 2012 a faecal specimen from a 1-year-old infant (case A) with symptoms of HUS was tested for bacterial gastrointestinal pathogens at a local hospital laboratory. The culture was reported as negative for Campylobacter, Salmonella, Shigella species and E. coli O157 and was submitted to GBRU for further testing for other non-O157 STEC, as recommended by the guidelines in the PHE VTEC operational manual. PCR tests at GBRU showed the faecal specimen was positive for the stx2 and eae (intimin) genes and subsequently a strain of STEC O26:H11 stx2a was cultured from the specimen (Fig. 1 and Table 2).

Following the positive culture from case A, faecal specimens were collected from the household contacts on 10 July, including case B, a 4-year-old sibling who attended nursery and was therefore considered in a risk group, and both parents (cases C and D). All household contacts were asymptomatic at this time. The first specimen from case B was negative for STEC although subsequent specimens taken over 2 weeks later were positive by PCR and culture (Fig. 1, Table 2). Both parents (cases C and D) were positive by PCR, although STEC O26:H11 was cultured from case D only (Table 2). Cases A and B continued to excrete STEC O26:H11 until 3 August. Subsequent specimens were culture-negative although specimens from case A continued to be PCR-positive up to the point when sampling ceased on 16 August.

There was a complex exposure history within the household. In the 7 days prior to case A's onset of illness the family had visited two petting farms and consumed burgers at a family barbeque. Case D had travelled to mainland Europe on two occasions (Table 1) during the incubation period and had also experienced mild gastrointestinal illness at the same time as case A became ill on 25 June.

A 3-year-old (case E) who attended the same nursery as cases A and B experienced diarrhoea, vomiting and abdominal pain commencing on 5 July 2012. A faecal specimen collected on 10 July was positive for stx2 and eae and a strain of STEC O26:H11 stx2a was cultured (Table 2). Those cases (A, B, E) who attended the nursery were excluded until two culture-negative results were reported. However, case F, a 6-month-old infant with symptoms of HUS, became ill on 19 July 2012. This case had direct contact with case E outside of a nursery setting during the time

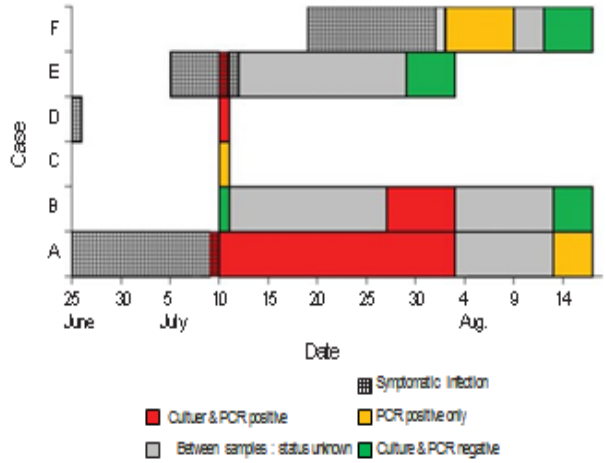


Fig. 1. Epidemic curve for the six outbreak cases showing the timeline for symptomatic infection and STEC excretion monitored by culture and PCR. Diagnostic results for cases C and D are presented for sampling date only and not duration as only one specimen was taken from these two cases (see Table 1 for details).

case E was still excreting STEC. Faecal specimens collected from case F were PCR-positive for stx2 and eae but STEC O26:H11 was never cultured from either PCR-positive specimens (Table 2).

WGS of strains of STEC O26:H11

During the outbreak, the serotyping and stx profiles of the isolates along with the epidemiological data were regarded as strong evidence that these cases were linked. Retrospectively, we sequenced the whole genomes of 19 strains of E. coli O26:H11, including the outbreak strains, isolated between 2009 and 2013 (Table 1). The phylogenetic relationship of these strains, mapped to the reference STEC O26:H11 strain 11 368 (NC_ 011 361·1), is shown in Figure 2 and comprises of 3543 polymorphic positions. Case E (482/12) was 0 SNPs different at the core genome level to case D (483/12), two SNPs different from case A (627/12) and three SNPs different from case B (626/12). The genome sequencing data indicated that the strains linked to this nursery outbreak are closely related and clearly differentiated from the sequences of the other 15 strains of E. coli O26:H11 from the GBRU archive. Strains from a household outbreak in 2009 were also identified as being closely related to each other with <3 SNPs between each sequence and 272 SNPs from the closest strain 637/13.

Two different MLSTs were identified, ST21 and ST29. In this collection, ST21 and ST29 were associated with the stx-positive and stx-negative strains, respectively. Two of the three stx-negative strains were associated with cases with severe disease including one case of HUS and this phenomenon has been described previously [26].

All of the sequenced strains had 100% identity over the complete length to the fliC gene of O26:H11 strain 11 368 (NC_ 011 361·1) and >99% identity to the complete length of the β-intimin gene.

We performed a metagenomic analysis on the PCR-positive/culture-negative faecal specimen from case F. Sequenced reads from the extracted faecal DNA were mapped against a panel of E. coli virulence genes (Table 3). Reads were successfully mapped to the STEC virulence factors eae and stx as well as wzx, the E. coli O26:H11 lipopolysaccharide gene and fliC encoded the flagella antigen H11. There were no reads mapped to virulence factors indicative of the presence of Shigella spp., ETEC or EAEC. When all metagenomic reads from the faecal specimen from case F were mapped to the assembled contigs of the isolate from case B, 75% of the genome was covered with depth greater than 10 reads comprising 52% of metagenomic reads.

Sequenced reads mapped to VFDB resulted in significant hits or matches to 1663 virulence factors. Of the 1312 that could be attributed to a species 994 (76%) were associated with E. coli, 24 (2%) with Salmonella spp., 208 (16%) with Shigella spp., 62 (5%) with Yersinia spp. and 1% with other species. Genes matching those found in other enteric pathogens could be accounted for by homology of certain genes to similar genes found in E. coli.

Table 2. Real-time PCR results and corresponding culture results for each faecal specimen received from each outbreak case

	Sample date (2012)	PCR	Culture
Case A	8 July	+	+
	10 July	+	+
	27 July	+	+
	1 Aug.	+	+
	3 Aug.	+	+
	13 Aug.	+	–
	16 Aug.	+	–
Case B	10 July		–
	27 July	+	+
	1 Aug.	+	+
	3 Aug.	+	+
	13 Aug.	–	–
	16 Aug.	–	–
Case C	10 July	+	–
Case D	10 July	+	+
Case E	10 July	+	+
	27 July	–	–
	31 July	–	–
Case F	2 Aug.	+	–
	2 Aug.	+	–
	8 Aug.	+	–
	12 Aug.	–	–
	16 Aug.	–	–

PCR+/culture+, stx DNA detected and STEC isolated by culture; PCR+/culture-, stx DNA detected but STEC not isolated by culture; PCR-/culture-, stx DNA not detected and STEC not isolated by culture.

DISCUSSION

PCR for the direct detection of STEC is rapid, sensitive and facilitates the detection of all STEC serogroups. In this outbreak, direct PCR led to the rapid identification of cases, including one PCR-positive/culture-negative case with HUS, for which STEC infection would not have been confirmed through traditional culture-dependent methods. However, this approach raises a number of public health issues. First, although the initial diagnosis is rapid, the follow-up culture and isolation of the pathogen can be labour-intensive and time-consuming and may delay public health action. Second, the increased sensitivity of the PCR can lead to PCR-positive but culture-negative results that can be difficult to interpret with respect to the risk of person-to-person transmission. Providing guidelines for public health action in such cases, especially in those belonging to risk groups including young children and food handlers, may be challenging. Following the outbreak described here, case A returned to nursery once he was asymptomatic and had two culture-negative stools, despite the positive PCR results. No new cases were detected in the nursery cohort during this investigation following case A's return. Third, the clinical and public health significance of all serotypes of non-O157 STEC is unclear [27, 28]. STEC O26:H11, specifically those strains harbouring stx2 only, are highly pathogenic and capable of causing HUS. This outbreak strain had stx2 and eae, both pathogenicity factors associated with severe disease [29–31] and the pathogenic potential (two cases developed symptoms of HUS) and transmissibility between close contacts of this strain, was clearly demonstrated.

While two cases developed HUS, four cases had less severe or asymptomatic infection. The two cases that developed HUS were aged <5 years, the recognized risk group for progression to HUS following STEC infection [32]. While it is unclear who the index case was, case D reported mild gastrointestinal symptoms at the same time as case A was ill and it is possible that case D acquired STEC infection abroad and then transmitted it within the household. The mild/asymptomatic nature of the illness in adults in this outbreak (cases C and D) illustrate the limitations of only testing those cases with more severe disease for non-O157 STEC infection, as recommended by the current guidelines in the VTEC operational manual, and the advantages of testing all specimens by PCR at a local level. Given the high transmissibility of STEC within households, an index case in an adult with mild symptoms may go on to infect more vulnerable members of the household leading to more severe disease. The use of PCR for the direct detection of STEC at local hospital laboratories will facilitate the detection of non-O157 STEC, particularly in cases with less severe disease which may otherwise have gone undetected.

Long-term shedding of STEC is well recognized, particularly in children [33]. Given the cases of mild and asymptomatic infection described in this study and the long-term shedding of infection, it was not possible to determine the order of infection or the source of the outbreak. Transmission to humans occurs through direct or indirect contact with animals or their environment, consumption of contaminated food or water and through person-to-person contact [34, 35]. The main reservoir of STEC in England is cattle although it is carried by other animals, mainly ruminants. The family of the index case had recent contact with the farming environment on two

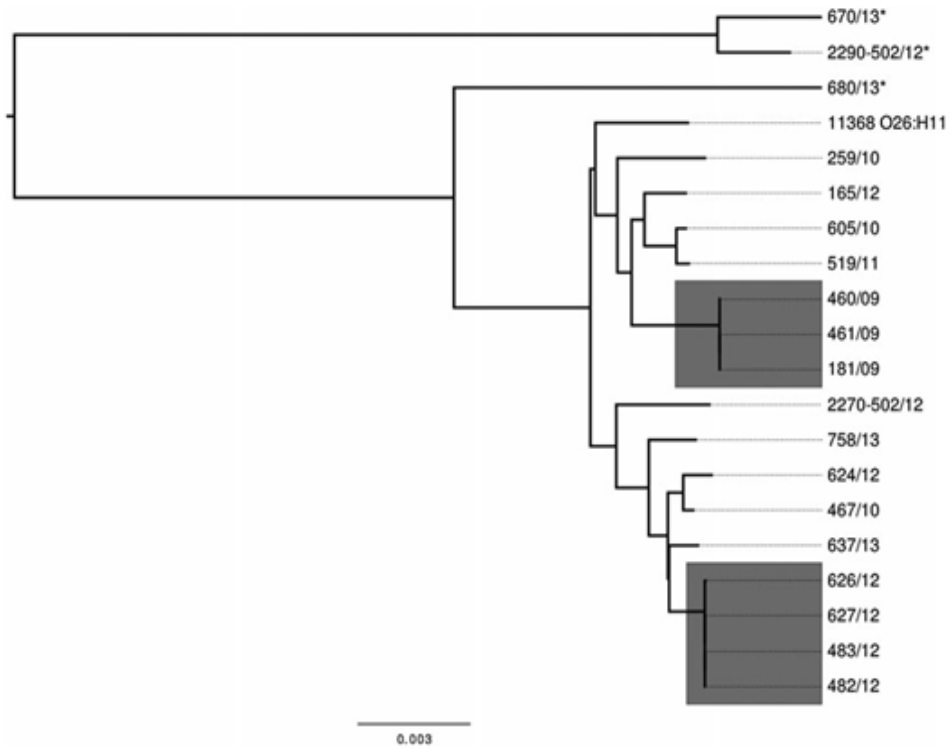


Fig. 2. Phylogenetic relationship of 19 strains of E. coli O26 isolated at GBRU between 2009 and 2013. Outbreak strains of STEC O26 from 2009 and 2012 are highlighted in grey. The scale represents number of nucleotide substitutions per site. * Strains negative for the stx genes.

Table 3. Number of reads of the metagenomic sequencing data taken from the faecal specimen from case F to map to virulence genes associated with diarrhoeagenic E. coli

Gene	Stx1	Stx2	eae	ipaH	aggR	bfpA
No. of reads	–	25	8	–	–	–
Gene	ST	rfbE O157	Wzx O26	Wzx O103	Wbdl O111	Ihp O145
No. of reads	–	–	14	–	–	–

occasions prior to onset of symptoms and these farm visits were initially regarded as the likely source of the infection. However, the stx profile of the outbreak strain (stx2a only) is rarely seen in the UK cattle population where strains characteristically harbour stx1 only or stx1 and stx2 [36]. Taking into account the travel history of case D, the possibility exists that the outbreak strain may have been acquired abroad, although there is no direct evidence for this.

As well as a cluster of cases in the same household (cases A–D), two cases (E and F) were identified at the nursery attended by cases A and B. STEC nursery outbreaks are not uncommon, and STEC O26:H11 has been previously linked to a number of such outbreaks [37–39]. In this outbreak, the epidemiological links (same household or attendance at the same nursery) between the cases were strong. However, the serotyping results did not provide conclusive evidence that the strains were microbiologically linked as STEC O26:H11 is the second most commonly isolated STEC serotype in England [3]. In contrast, the genome sequencing performed during this study, provided rapid, robust and highly discriminatory confirmation at the molecular level that the strains of STEC O26:H11 associated with the outbreak were linked (<5 SNPs) and had a common source (Fig. 2). In this example, we demonstrated that WGS was a suitable method for typing STEC O26:H11 in order to facilitate outbreak investigations.

Although the epidemiological data indicated that case F was part of the outbreak, the metagenomic analysis provided further evidence of that link in the absence of positive culture results. This analysis

confirmed the likely presence of wzx O26, stx2 and eae (intimin) genes in the specimen. We were also able to show that there was no evidence from the metadata sequence set that other gastrointestinal pathogens were present. The number of reads mapping to each virulence gene was close to the threshold of a detectable limit suggesting that this approach is not yet sensitive enough to be used routinely as a diagnostic tool. Furthermore, PCR assays for the detection of the E. coli O26 target gene (wzx O26) from faecal specimens have been well validated and are currently used in many diagnostic laboratories [40]. However, the value of using of using metagenomic sequencing analyses as an 'open-ended, culture-independent' approach during outbreaks of unknown aetiology has been previously documented [12].

Non-O157 outbreaks are thought to be rare in England but as the current frontline diagnostic laboratory protocols are not designed to detect non-O157 STEC, there is no surveillance data to support this. Elsewhere, outbreaks of non-O157 STEC can involve large numbers of cases and associated symptoms can be severe and include HUS [41, 42]. The cost of WGS, including DNA extraction is currently about €50 per isolate which compares favourably with the costs of the multiple biochemical, serological and alternative molecular tests required to type non-O157 STEC to facilitate outbreak investigations currently. Furthermore, unlike the STEC O157 MLVA scheme used at GBRU, WGS is not serotype-specific. WGS is more discriminatory than MLST, as demonstrated here, and sequencing data is easily transferable. At PHE, we have established a robust and reliable pipeline for routine genome sequencing and analysis of isolates from non-O157 STEC outbreaks within 48–72 h. WGS also provided a full virulence profile of the pathogen from DNA extracted directly from a PCR-positive but culture-negative faecal specimen from a case epidemiologically linked to a nursery outbreak. This study demonstrates the potential of WGS to determine the pathogenicity profile and evolutionary origin of certain strains as well as facilitating outbreak investigations of non-O157 STEC in England.

ACKNOWLEDGEMENTS

The authors thank all colleagues from the National Health Service and Public Health England frontline laboratories who refer specimens to GBRU. We also acknowledge Marie Chattaway, Neil Perry, Vivienne do Nascimento, Michael Wright and Yoshini Taylor for their hard work and microbiological expertise at GBRU. Thanks are also due to Natalie Adams and Radha Patel for their contributions to the National Enhanced Surveillance System for STEC in England and to all public health practitioners who report to the scheme.

DECLARATION OF INTEREST

None.

REFERENCES

1. Lynn RM, et al. Childhood hemolytic uremic syndrome, United Kingdom and Ireland. Emerging Infectious Diseases 2005; 11: 590–596.
2. Garrity GM, Brenner DJ, Krieg NR, Staley JT (eds). In: Bergey's Manual of Systematic Bacteriology, 2nd edn, Vol. 2: The Proteobacteria. New York: Springer, 2005.
3. Byrne et al. The epidemiology and microbiology of Shiga toxin-producing Escherichia coli other than serogroup O157 in England 2009–2013. Journal of Medical Microbiology 2014; 639: 1181–1188
4. Mellman et al. Analysis of a collection of haemolytic uremic syndrome associated enterohaemorrhagic Escherichia coli. Emerging Infectious Diseases 2008; 14: 1287–1290
5. Zhang WL, et al. Molecular characteristics and epidemiological significance of Shiga toxin-producing Escherichia coli O26 strains. Journal of Clinical Microbiology 2000; 38: 2134–2140.
6. Bielaszewska M, et al. Enterohemorrhagic Escherichia coli O26:H11/H-: a new virulent clone emerges in Europe. Clinical Infectious Diseases 2013; 56: 1373–1381.
7. Jenkins C, et al. Assessment of a real-time PCR for the detection and characterization of verocytotoxigenic Escherichia coli. Journal of Medical Microbiology 2012; 61: 1082–1085.
8. Byrne L, et al. Evaluating the use of multilocus variable number tandem repeat analysis of Shiga toxin-producing Escherichia coli O157 as a routine public health tool in England. PLoS ONE 2014; 9: e85901.
9. Karama M & Gyles CL. Methods for genotyping Verotoxin-producing Escherichia coli. Zoonoses Public Health 2010; 57: 447–462.
10. McDonnell J, et al. Retrospective analysis of whole genome sequencing compared to prospective typing data in further informing the epidemiological investigation of an outbreak of Shigella sonnei in the UK. Epidemiology & Infection 2013; 141: 2568–2575.
11. Byrne, et al. A multi-country outbreak of Salmonella Newport gastroenteritis in Europe associated with watermelon from Brazil confirmed by whole genome sequencing: October 2011 to January 2012. Eurosurveillance 2014; 19: 6–13.
12. Loman NJ, et al. A culture-independent sequence-based metagenomics approach to the investigation of an

outbreak of Shiga-toxigenic Escherichia coli O104:H4. Journal of the American Medical Association 2013; 309: 1502–1510.
13. Halstead FD, et al. Olympics GI Group. Universal extraction method for gastrointestinal pathogens. Journal of Medical Microbiology 2013; 62: 1535–1539.
14. Li H and Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 2010; 26: 589–595.
15. Li H, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009; 25: 2078–2079.
16. DePristo MA et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics 2011; 43: 491–498.
17. McKenna A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010; 20: 1297–1303.
18. Price MN, et al. FastTree 2 – approximately maximum-likelihood trees for large alignments. PLoS ONE 2010; 5: e9490.
19. Bankevich A, et al. SPAdes: a new genome assembly al- gorithm and its applications to single-cell sequencing. Journal of Computational Biology 2012; 19: 455–477
20. Inouye M, et al. Short read sequence typing (SRST): multi-locus sequence types from short reads. BMC Genomics 2012; 13: 338.
21. Altschul SF, et al. Basic local alignment search tool. Journal of Molecular Biology 1990; 215: 403–410.
22. Bolger AM, et al. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014; 30: 2114–2120.
23. Frankel G, et al. Intimin and the host cell – is it bound to end in Tir(s)? Trends in Microbiology 2001; 9: 214–218.
24. Nataro JP. Enteroaggregative Escherichia coli pathogenesis. Current Opinion in Gastroenterology 2005; 21: 4–8.
25. Chen LH, et al. VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors. Nucleic Acids Research 2012; 40 (database issue): D641–645.
26. Bielaszewska M, et al. Shiga toxin gene loss and transfer in vitro and in vivo during enterohemorrhagic Escherichia coli O26 infection in humans. Applied Environmental Microbiology 2007; 73: 3144–3150.
27. Feng PC & Reddy S. Prevalences of Shiga toxin subtypes and selected other virulence factors among Shiga-toxigenic Escherichia coli strains isolated from fresh produce. Applied Environmental Microbiology 2013; 79: 6917–6923.
28. Preußel K, et al. shiga toxin-producing Escherichia coli o157 is more likely to lead to hospitalization and death than non-O157 serogroups – except O104. PLoS ONE 2013; 8: 78180.
29. Persson S, et al. Subtyping method for Escherichia coli Shiga toxin (verocytotoxin) 2 variants and correlations to clinical manifestations. Journal of Clinical Microbiology 2007; 45: 2020–2024.
30. Melton-Celsa A, et al. Pathogenesis of Shiga-toxin producing Escherichia coli. Current Topics in Microbiology and Immunology 2012; 357: 67–103
31. Pollock KG, et al. Highly virulent Escherichia coli O26, Scotland. Emerging Infectious Diseases 2011; 17: 1777–1779.
32. Karmali MA. Host and pathogen determinants of verocytotoxin-producing Escherichia coli-associated hemolytic uremic syndrome. Kidney International Supplement 2009; 112: S4–7.
33. Dabke G, et al. Duration of shedding of Verocytotoxin-producing Escherichia coli in children and risk of transmission in childcare facilities in England. Epidemiology & Infection 2014; 142: 327–34
34. Locking ME, et al. Risk factors for sporadic cases of Escherichia coli O157 infection: the importance of contact with animal excreta. Epidemiology & Infection 2001; 127: 215–220.
35. Gillespie IA, et al. Foodborne general outbreaks of Shiga toxin-producing Escherichia coli O157 in England and Wales 1992–2002: where are the risks? Epidemiology & Infection 2005; 133: 803–808
36. Chase-Topping ME, et al. Pathogenic potential to humans of bovine Escherichia coli O26, Scotland. Emerging Infectious Diseases 2012; 18: 439–448.
37. Brown JA, et al. Outbreak of shiga toxin-producing Escherichia coli serotype O26: H11 infection at a child care center in Colorado. Pediatric Infectious Diseasse Journal 2012; 31: 379–383.
38. Sonoda C, et al. An enterohemorrhagic Escherichia coli O26 outbreak at a nursery school in Miyazaki, Japan. Japanese Journal of Infectious Diseases 2008; 61: 92–93.
39. Werber D, et al. A multistate outbreak of Shiga toxin-producing Escherichia coli O26:H11 infections in Germany, detected by molecular subtyping surveillance. Journal of Infectious Diseases 2002; 186: 419–422.
40. Kagkli DM, et al. Application of the modular approach to an in-house validation study of real-time PCR methods for the detection and serogroup determination of verocytotoxigenic Escherichia coli. Applied Environmental Microbiology 2011; 77: 6954–6963.
41. Frank C, et al. HUS Investigation Team. Epidemic profile of Shiga-toxin-producing Escherichia coli O104: H4 outbreak in Germany. New England Journal of Medicine 2011; 365: 1771–1780.
42. Luna-Gierke RE, et al. Outbreaks of non-O157 Shiga toxin-producing Escherichia coli infection: USA. Epidemiology & Infection 2014; 142: 2270–2280.

APPENDIX II: Co-authors statements of candidate's contributions

PhD by published work at University of Warwick Medical School

Statement of contribution by Lisa Anne Byrne

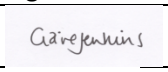
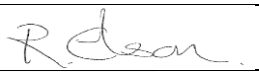
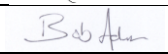
Paper to be used towards the PhD:

Byrne L, Jenkins C, Launders N, R. Elson, Adak GK. The Epidemiology, microbiology and clinical impact of Shiga toxin-producing *Escherichia coli* in England, 2009-2012. *Epidemiol* Infect. 2015 Apr 29:1-13

Study circumstances: This study was conducted to provide a recent and comprehensive update on the epidemiological distribution, microbiological characterisation and clinical impact of STEC infections in England. The study utilised data collected via an enhanced surveillance for STEC, newly introduced in 2009 and was the first paper describing these data. The study was performed at Public Health England. Lisa Byrne was the epidemiological scientist responsible for the day to day management of national surveillance of STEC at the time of the study.

Contribution: Lisa Byrne designed the study, managed and cleaned the data, performed descriptive and statistical analyses, led writing the manuscript in liaison with co-authors and responded to reviewers as corresponding author.

I agree that Lisa Anne Byrne made the contribution to the paper as stated above:

Name	Signature	Date
Claire Jenkins		07-09-17
Naomi Launders		07-09-17
Richard Elson		08-09-17
Goutam Adak		08-09-17

PhD by published work at University of Warwick Medical School

Statement of contribution by Lisa Anne Byrne

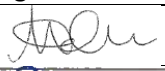
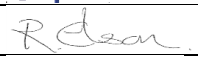
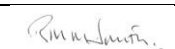
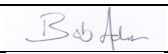
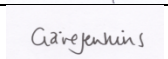
Paper to be used towards the PhD:

Natalie L. Adams, **Lisa Byrne**, Geraldine A. Smith, Richard Elson, John Harris, Roland Salmon, Robert Smith, Sarah O'Brien, Goutam K. Adak, and Claire Jenkins. Shiga toxin-producing Escherichia coli (STEC) O157 in England and Wales. Three decades on, what has changed? *Emerg Infect* Dis. 2016 Apr;22(4):590-7

Study circumstances: The study was carried out to provide a longitudinal review of STEC O157 infections across a thirty-year period since its emergence in the 1980's. In the study, changes in microbiological testing over the period are described alongside the description and interpretation of trends in the number and microbiological subtype of infections. Outbreaks were reviewed across the time period and changes in vehicles and settings over time considered. The study was performed at Public Health England. Lisa Byrne was the epidemiological scientist responsible for the day to day management of national surveillance of STEC at the time of the study.

Contribution: Lisa Byrne contributed to the study design and interpretation of data, extracted and cleaned outbreak data, wrote the outbreak section, drafted and critically revised the manuscript alongside co-authors.

I agree that Lisa Anne Byrne made the contribution to the paper as stated above:

Name	Signature	Date
Natalie Adams		08-09-17
Geraldine Smith		08-09-17
Richard Elson		08-09-17
John Harris		07-09-17
Roland Salmon		08-09-17
Robert Smith		07-09-17
Sarah O'Brien*		
Goutam Adak		08-09-17
Claire Jenkins		07-09-17

PhD by published work at University of Warwick Medical School

Statement of contribution by Lisa Anne Byrne

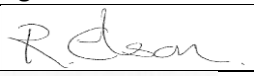
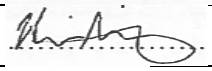
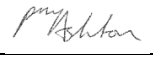
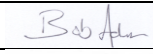
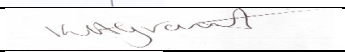
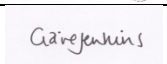
Paper to be used towards the PhD:

Byrne L, Elson R**,** Dallman TJ, Perry N, Ashton P, Wain J, Adak GK, Grant KA, Jenkins C. Evaluating the use of multilocus variable number tandem repeat analysis of Shiga toxin-producing *Escherichia coli* O157 as a routine public health tool in England**.** *PLoS ONE* 9(1): e85901.

Study circumstances: The study was carried out following an initial six-month period of routinely conducting multilocus variable number tandem repeat analysis (MLVA) on all STEC O157 strains isolated by the Gastrointestinal Bacteria Reference Laboratory in England. The study aimed to evaluate the impact of this molecular method on the enhanced detection of clusters of infection and the defining of sporadic and linked cases, alongside consideration of other epidemiological factors linking cases. The study was performed at Public Health England. Lisa Byrne was the epidemiological scientist responsible for the day to day management of national surveillance of STEC at the time of the study.

Contribution: Lisa Byrne designed the study, managed and cleaned the data, performed all descriptive and statistical analyses, led writing the manuscript in liaison with co-authors and responded to reviewers as corresponding author.

I agree that Lisa Anne Byrne made the contribution to the paper as stated above:

Name	Signature	Date
Richard Elson		08-09-17
Timothy Dallman		14-09-17
Neil Perry		07-09-17
Philip Ashton		12-09-17
John Wain*		
Goutam Adak		08-09-17
Kathie Grant		11-09-17
Claire Jenkins		07-09-17

PhD by published work at University of Warwick Medical School

Statement of contribution by Lisa Anne Byrne

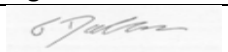
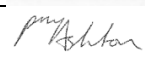
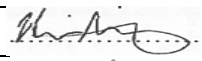
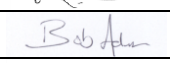
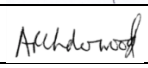
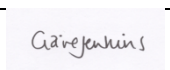
Paper to be used towards the PhD:

Dallman TJ,' **Byrne L**, Ashton P,' Cowley LA, Perry NT, Elson R, Adak GK, Underwood A, Green J, Jenkins C, Grant KA Wain J. Whole Genome Sequencing for National Surveillance of Shiga Toxin Producing *Escherichia coli* O157. *Clin Infect Dis*. 2015 Apr 17.

Study circumstances: This study was conducted in response to the advent, and planned implementation of WGS for analysing strains of enteric pathogens. The study aimed to assess the impact of WGS on detecting clusters of disease, determine a threshold for relatedness and compare it against existing typing methods. The study was performed at Public Health England. Lisa Byrne was the epidemiological scientist responsible for the day to day management of national surveillance of STEC at the time of the study.

Contribution: Lisa Byrne (joint first author) led on the epidemiological aspects of study design, provided epidemiological context to clusters described; Conducted the evaluation of MLVA versus WGS including survival analysis and contributed to the drafting and critical revision of the manuscript.

I agree that Lisa Anne Byrne made the contribution to the paper as stated above:

Name	Signature	Date
Timothy Dallman		14-09-17
Philip Ashton		12-09-2017
Lauren Cowley		08-09-17
Neil Perry		07-09-17
Richard Elson		08-09-17
Goutam Adak		08-09-17
Anthony Underwood		11-09-17
John Green*		
Claire Jenkins		07-09-17
Kathie Grant		11-09-17
John Wain*		

PhD by published work at University of Warwick Medical School

Statement of contribution by Lisa Anne Byrne

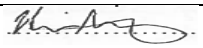
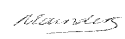
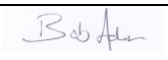
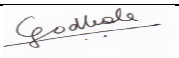
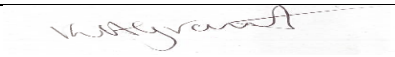
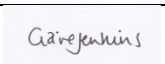
Paper to be used towards the PhD:

Byrne L, Vanstone G. Perry N, Launders N, Adak GK, Godbole G, Grant K. Smith R., Jenkins C. The epidemiology and microbiology of Shiga-toxin producing *Escherichia coli* other than serogroup O157 in England 2009-2013. J *Med Microbiol*. 2014 Sep;63(Pt 9):1181-8

Study circumstances: From December 2012, a small number of frontline laboratories implemented a commercial PCR assay which directly detects the *stx* genes and can detect all STEC serogroups not just serogroup O157. This study was conducted in response to the resultant increase in detection of Non-O157 STEC and described the characteristics of strains and compared disease severity amongst cases based on strain characteristics .The study was performed at Public Health England. Lisa Byrne was the epidemiological scientist responsible for the day to day management of national surveillance of STEC at the time of the study.

Contribution: Lisa Byrne designed the study, managed and cleaned the data, performed all descriptive and statistical analyses, and wrote the manuscript in liaison with co-authors.

I agree that Lisa Anne Byrne made the contribution to the paper as stated above:

Name	Signature	Date
Gemma Vanstone*		
Neil Perry		07-09-17
Naomi Launders		07-09-17
Goutam Adak		08-09-17
Gauri Godbole		11-09-17
Katie Grant		11-09-17
Robin Smith*		
Claire Jenkins		07-09-17

PhD by published work at University of Warwick Medical School

Statement of contribution by Lisa Anne Byrne

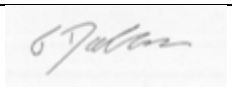
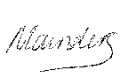
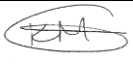
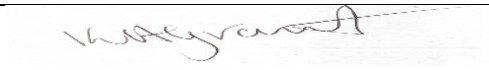
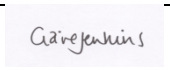
Paper to be used towards the PhD:

Dallman TJ, **Byrne L**, Launders N, Glen K, Grant KA, Jenkins C. The utility and public health implications of PCR and whole genome sequencing for the detection and investigation of an outbreak of Shiga toxin-producing Escherichia coli serogroup O26:H11. *Epidemiol Infect*. 2014 Oct 15:1-9.

Study circumstances: An outbreak of non-O157 STEC in a nursery school was described in respect to the application of both PCR and WGS in investigations. The utility of these methods in real-time was demonstrated. The study was performed at Public Health England. Lisa Byrne was the epidemiological scientist responsible for the day to day management of national surveillance of STEC at the time of the study.

Contribution: Lisa Byrne (joint first author) provided epidemiological input into the study design, provided the epidemiological context to the study and contributed to the drafting and revising of the manuscript.

I agree that Lisa Anne Byrne made the contribution to the paper as stated above:

Name	Signature	Date
Timothy Dallman		14-09-17
Naomi Launders		07-09-17
Kirsten Glen		07-09-17
Kathie Grant		11-09-17
Claire Jenkins		07-09-17

*These co-authors have been approached for their attested statements but due to them being on leave were unable to provide them in the time needed for timely submission of this thesis. I will continue to communicate with them and expect to have all signatures for a subsequent, final version of the thesis

APPENDIX III: Full list of the candidate's publications

Byrne L, Campbell H, Andrews N, Ribeiro S, Amirthalingam G. Hospitalisation of preterm infants with pertussis in the context of a maternal vaccination programme in England. *Arch Dis Child*. 2017 Aug 16. pii: archdischild-2016-311802. doi: 10.1136/archdischild-2016-311802.

Ward C, **Byrne L**, White JM, Amirthalingam G, Tiley K, Edelstein M. Sociodemographic predictors of variation in coverage of the national shingles vaccination programme in England, 2014/15. Vaccine. 2017;35(18):2372-8

Byrne L, Adams N, Glen K, Dallman TJ, Kar-Purkayastha I, Beasley G, et al. Epidemiological and Microbiological Investigation of an Outbreak of Severe Disease from Shiga Toxin-Producing Escherichia coli O157 Infection Associated with Consumption of a Slaw Garnish. J Food Prot. 2016;79(7):1161-8

Launders N, **Byrne L**, Jenkins C, Harker K, Charlett A, Adak GK. Disease severity of Shiga toxin-producing E. coli O157 and factors influencing the development of typical haemolytic uraemic syndrome: a retrospective cohort study, 2009-2012. BMJ Open. 2016;6(1):e009933.

Natalie L. Adams, **Lisa Byrne**, Geraldine A. Smith, Richard Elson, John Harris, Roland Salmon, Robert Smith, Sarah O'Brien, Goutam K. Adak, and Claire Jenkins. Shiga toxin-producing Escherichia coli (STEC) O157 in England and Wales. Three decades on, what has changed? *Emerg Infect* Dis. 2016 Apr; 22(4):590-7. doi: 10.3201/eid2204.151485.

Timothy J. Dallman, Philip M. Ashton, **Lisa Byrne**, Neil T. Perry, Liljana Petrovska, Richard Ellis,Lesley Allison, Mary Hanson, Anne Holmes, George J. Gunn, Margo E. Chase-Topping, Mark E. J. Woolhouse, Kathie A. Grant, David L. Gally, John Wain and Claire Jenkins. Applying phylogenomics to understand the emergence of Shiga-toxin-producing Escherichia coli O157:H7 strains causing severe human disease in the UK. *Microbial Genomics*, Sep 2015.

Byrne L, Jenkins C, Launders, R. Elson Adak GK. The Epidemiology, microbiology and clinical impact of Shiga toxin-producing *Escherichia coli* in England, 2009-2012. *Epidemiol Infect*. 2015 Apr 29:1-13

Dallman TJ,′ **Byrne L**, Ashton P,′ Cowley LA, Perry NT, , Elson R, Adak GK, Underwood A, Green J, Jenkins C, Grant KA Wain J. Whole Genome Sequencing for National Surveillance of Shiga Toxin Producing *Escherichia coli* O157. *Clin Infect Dis*. 2015 Aug 1;61(3):305-12. doi: 10.1093/cid/civ318. Epub 2015 Apr 17

Jenkins C, Dallman TJ, Launders N, Willis C, **Byrne** L, Jorgensen F, et al. Public Health Investigation of Two Outbreaks of Shiga Toxin-Producing Escherichia coli O157 Associated with Consumption of Watercress. Appl Environ Microbiol. 2015;81(12):3946-52.

Simms I, Gilbart VL, **Byrne L**, Jenkins C, Adak GK, Hughes G, Crook PD. Identification of verocytotoxin-producing Escherichia coli O117:H7 in men who have sex with men, England, November 2013 to August 2014. *Euro Surveill*. 2014 Oct 30;19(43).

Dallman TJ, **Byrne L**, Launders N, Glen K, Grant KA, Jenkins C. The utility and public health implications of PCR and whole genome sequencing for the detection and investigation of an outbreak of Shiga toxin-producing Escherichia coli serogroup O26:H11. *Epidemiol Infect*. 2014 Oct 15:1-9.

Byrne L, Fisher I, Peters T, Mather A, Thomson N, Rosner B, Bernard H, McKeown P, Cormican M, Cowden J, Aiyedun V, Coia J, Brown D, Lane C. **A** Multi-Country Outbreak of *Salmonella* Newport gastroenteritis in Europe associated with watermelon from Brazil, confirmed by whole genome sequencing: October 2011-January 2012. *Euro Surveill.* 2014 Aug 7;19(31):6-13

Byrne L, Vanstone G. Perry N, Launders N, Adak GK, Godbole G, Grant K. Smith R., Jenkins C. The epidemiology and microbiology of Shiga-toxin producing *Escherichia coli* other than serogroup O157 in England 2009-2013. J Med Microbiol. 2014 Sep;63(Pt 9):1181-8

Byrne L, Elson R**,** Dallman TJ, Perry N, Ashton P, Wain J, Adak GK, Grant KA, Jenkins C. Evaluating the use of multilocus variable number tandem repeat analysis of Shiga toxin-producing *Escherichia coli* O157 as a routine public health tool in England**.** *PLoS ONE* 9(1): e85901. doi:10.1371/journal.pone.008590. January 2014.

Launders N, **Byrne L**, Adams N, Glen K, Jenkins C, Tubin-Delic D, Locking M, Williams C, Morgan D, Outbreak Control Team. Outbreak of Shiga toxin-producing E.coli O157

associated with consumption of watercress, United Kingdom, August to September 2013. *Euro Surveill* 2013 Oct 31; 18(44)

Sakkejha H, **Byrne L**, Lawson AJ, Jenkins C. An update on the microbiology and epidemiology of enteropathogenic Escherichia coli in England 2010-2012. *J Med Microbiol*. 2013 Oct; 62(pt10): 1531-4

Byrne L, Brant L, Reynolds C, Ramsay M. Seroprevalence of low rubella IgG antibody levels among antenatal women in England tested by NHS Blood and Transplant: 2004-2009. Is rubella susceptibility increasing? *Vaccine*. 2012 Jan 5;30(2):161-7.

Lisa J Brant, Claire Reynolds, **Lisa Byrne**, Katy L Davison. Hepatitis B and residual risk of infection in English and Welsh blood donors, 1996 through 2008. *Transfusion* 04/2011; 51(7):1493-502

Lisa Byrne, Lisa J Brant, Katy Davison, Patricia Hewitt .Transfusion-transmitted human immunodeficiency virus (HIV) from seroconverting donors is rare in England and Wales: results from HIV lookback, October 1995 through December 2008. *Transfusion* 12/2010; 51(6):1339-45

Chapter: Blood Donation Testing and the Safety of the Blood Supply David Wenham FIBMS PTI Manager, **Lisa Byrne** BSc Surveillance Officer (Information, Simon J. Stanworth MA, MRCP (Paeds), D. Phil, FRCPath Consultant Haematologist. Practical Transfusion Medicine, Third Edition, 05/2009: pages 200 - 208; , ISBN: 9781444311761

APPENDIX IV: Enhanced Surveillance Questionnaire

Vero cytotoxin-producing *Escherichia coli* Enhanced surveillance questionnaire

SECTION A:QUESTIONNAIRE DETAILS

Interviewer name: Interview date : / /

Interviewer office: Interviewer Telephone:

Person interviewed name:

Details from: ☐ Case ☐ Case's Parent ☐ Other (specify):

SECTION B: CASE CLASSIFICATION – SEE GUIDANCE NOTES FOR DEFINITIONS

Is this case: ☐ Primary ☐ Co-primary ☐ Secondary ☐ Not sure ☐ Asymptomatic

If secondary, name of primary case:

Investigation is: ☐ Ongoing ☐ Complete

Outbreak keyword or number:

Outcome:

Select all that apply ☐ Recovered ☐ Still ill ☐ HUS/TTP ☐ Died

if died, then → Date of death: / /

SECTION C:PERSONAL DETAILS

First name: Family name:

Address:

Postcode: Tel (h): Tel (m):

Email:

Sex: ☐ M ☐ F Date of birth (dd/mm/yyyy): / / Age: yrs

NHS No: GP name:

GP address: GP Tel:

Are there any children living in the household? (other than the case) ☐ Y ☐ N ☐ U

Occupation:

Tick if any of the below risk groups apply

☐ Foodhandler (e.g. handle food professionally)
☐ Work in/attend healthcare setting
☐ Work in/attend childcare setting
☐ Work in contact with faeces (e.g. lab, farm etc.)
☐ Have difficulty maintaining personal hygiene
☐ Other risk category

If yes to any of the above, details:

Work/School: Tel:

Address:

Postcode: Date of last attendance:

Ethnicity: ☐ White ☐ Mixed ☐ Asian/Asian British ☐ Black/Black British
☐ Chinese Other:

April 2013 v3

SECTION D: SYMPTOMS OF ILLNESS

Onset date: ☐☐☐ Still ill: ☐ Y ☐ N Duration of illness(days): ☐☐

Symptoms experienced:	Yes	No	Not Sure	Ongoing	Duration (d)	Date of onset
Diarrhoea (3 or more loose stools in 24hrs)	☐	☐	☐	☐	☐	☐☐/☐☐/☐☐
Bloody stools	☐	☐	☐	☐	☐	☐☐/☐☐/☐☐
Nausea	☐	☐	☐	☐	☐	☐☐/☐☐/☐☐
Vomiting	☐	☐	☐	☐	☐	☐☐/☐☐/☐☐
Abdominal pain	☐	☐	☐	☐	☐	☐☐/☐☐/☐☐
Fever	☐	☐	☐	☐	☐	☐☐/☐☐/☐☐

Other (specify): ☐

Sought healthcare: ☐ NHS Direct ☐ GP visit ☐ A&E ☐ Other (specify): ☐

Submitted stool sample: ☐ Y ☐ N

Admitted to hospital for this illness: ☐ Y ☐ N Admission date: ☐☐/☐☐/☐☐

Hospital name: ☐ Duration of stay (d): ☐☐

Self/Medicated with antibiotics: ☐ Y ☐ N ☐ Not sure If Y, specify: ☐

Self/Medicated with antidiarrhoeals: ☐ Y ☐ N ☐ Not sure If Y, specify: ☐

Detail any other health concerns: ☐

SECTION E: TRAVEL IN THE WEEK PRIOR TO ILLNESS

Travelled OUTSIDE of the UK: ☐ Y ☐ N ☐ Not sure

Specify countries visited (from most recent to least recent)

Country/Region	Date arrived	Date departed	Details
☐	☐☐/☐☐/☐☐	☐☐/☐☐/☐☐	☐
☐	☐☐/☐☐/☐☐	☐☐/☐☐/☐☐	☐
☐	☐☐/☐☐/☐☐	☐☐/☐☐/☐☐	☐

Type of accommodation used:

☐ Hotel ☐ Bed & Breakfast ☐ Guesthouse ☐ Friends/Family

☐ Tourist Campsite ☐ Holiday dwelling ☐ Other (specify): ☐

Name of accommodation: ☐

Travelled WITHIN the UK: ☐ Y ☐ N ☐ Not sure

Specify town/resort visited (from most recent to least recent)

Town/Resort	Date arrived	Date departed	Details
☐	☐☐/☐☐/☐☐	☐☐/☐☐/☐☐	☐
☐	☐☐/☐☐/☐☐	☐☐/☐☐/☐☐	☐
☐	☐☐/☐☐/☐☐	☐☐/☐☐/☐☐	☐

Type of accommodation used:

☐ Hotel ☐ Bed & Breakfast ☐ Guesthouse ☐ Friends/Family

☐ Tourist Campsite ☐ Holiday dwelling ☐ Other (specify): ☐

Name of accommodation: ☐ Postcode: ☐

April 2013 v3

SECTION F: FOOD HISTORY IN THE WEEK PRIOR TO ILLNESS

Ate out: ☐ Y ☐ N

Venues	Yes	No	Name/Location of place	Dates	Eaten
Restaurant 1	☐	☐			
Restaurant 2	☐	☐			
Takeaway1	☐	☐			
Takeaway2	☐	☐			
Café/Canteen	☐	☐			
Party/BBQ/function	☐	☐			
Other1	☐	☐			
Other2	☐	☐			

Do you handle/prepare most of the food within the household: ☐ Always ☐ Mostly ☐ Occasionally ☐ Never

Ate or handled any of the following:

		Yes	No	Product (e.g. boneless, skinless chicken breast)	Where purchased-name & location (e.g. Asda Acton, London)
HANDLED	Raw beef	☐	☐		
	Raw poultry	☐	☐		
	Raw lamb	☐	☐		
	Raw pork/gammon	☐	☐		
	Other raw meat (e.g. game, goat, ostrich etc.)	☐	☐		
	Raw vegetables	☐	☐		
	Pet/animal feed	☐	☐		
CONSUMED	Any meat	☐	☐	*If yes, indicate below:*	
	Beef	☐	☐		
	Cooked poultry	☐	☐		
	Cooked lamb	☐	☐		

April 2013 v3

CONSUMED				
Cooked pork/gammon	☐	☐		
Other cooked meat e.g. game, goat, ostrich etc.)	☐	☐		
Cured meats	☐	☐		
Other processed meat	☐	☐		
Fish	☐	☐		
Shellfish	☐	☐		
Pasteurised milk	☐	☐		
Unpasteurised milk	☐	☐		
Hard cheese	☐	☐		
Soft cheese	☐	☐		
Yoghurt/fromage frais	☐	☐		
Cream	☐	☐		
Ice Cream	☐	☐		
Unpasteurised dairy products	☐	☐		
Pre-packaged salad	☐	☐		
Other salad	☐	☐		
Raw vegetables	☐	☐		
Soft fruit/berries	☐	☐		
Pre cut fruits	☐	☐		
Other Raw fruit	☐	☐		
Sprouted seeds/ beansprouts	☐	☐		
Fresh herbs	☐	☐		
Fruit juices	☐	☐		
Pre-packaged sandwiches etc.	☐	☐		
Other foods (e.g.nuts, confectionery, etc.)	☐	☐		

SECTION G: WATER EXPOSURE IN THE WEEK PRIOR TO ILLNESS

Drank unboiled water from any of the following:

Water supply	Yes	No	Details
Mains (municipal) water	☐	☐	
Private water supply (spring/well/borehole)	☐	☐	
Bottled water	☐	☐	
Unboiled river/stream/lake water	☐	☐	

Exposed to floodwater: ☐ Y ☐ N ☐ Not sure

If yes, details:

Experienced any household drainage/plumbing problems: ☐ Y ☐ N ☐ Not sure

If yes, details:

Participated in any of the following activities – either recreationally or for occupation:

Activity	Fresh water	Sea Water	No	Details
Swimming/paddling	☐	☐	☐	
Other (e.g. canoeing, fishing, sailing, sufing).	☐	☐	☐	

Is it possible that water was accidentally swallowed during any of the above: ☐ Y ☐ N ☐ Not sure

If yes, details:

SECTION H: ANIMAL CONTACT IN THE WEEK PRIOR TO ILLNESS

Contact with domestic animals/pets: ☐ Y ☐ N

If yes, indicate which animals below:

☐ Dogs ☐ Cats ☐ Rabbits ☐ Rodents ☐ Reptiles ☐ Birds ☐ Fish Other:

Did any of the above animals have diarrhoea: ☐ Y ☐ N ☐ Not sure

If yes, specify:

Was there a veterinary investigation? ☐ Y ☐ N ☐ Not sure

Contact with non domestic animals: ☐ Y ☐ N ☐ Not sure

Indicate which animals below:

Cattle ☐ Calves ☐ Sheep ☐ Lambs ☐ Goats ☐ Horses ☐ Pigs ☐ Reptiles ☐

Poultry ☐ Rabbits ☐ Deer ☐ Birds ☐ Fish ☐ Rodents ☐ Other:

Lives on, works at or access to a private farm: ☐ Y ☐ N ☐ Not sure

Attended an agricultural event e.g horse show ☐ Y ☐ N ☐ Not sure Attended date: / /

Visited a farm/petting zoo/bird reserve or such: ☐ Y ☐ N ☐ Not sure Visit date: / /

If yes, specify:

If Yes, Handled the animals: ☐ Y ☐ N ☐ Not sure

If yes, specify:

Bottle fed any of the animals: ☐ Y ☐ N ☐ Not sure

If yes, washed hands before eating food or before leaving: ☐ Y ☐ N ☐ Not sure

If yes, specify:

Consumed any food whilst there: ☐ Y ☐ N ☐ Not sure

If yes, specify: _____

Was the food (tick all that apply):

☐ Purchased on premises
☐ Brought from home/elsewhere
☐ Eaten in a separate area
☐ Eaten whilst in contact with the animals
☐ Eaten at a picnic table
☐ Eaten while sat on the grass/soil

SECTION I: ENVIRONMENTAL EXPOSURE IN THE WEEK PRIOR TO ILLNESS

Walked in a paddock/field where farm animals graze: ☐ Y ☐ N ☐ Not sure

If yes, specify: _____

Taken any day trips (beach, countryside etc.): ☐ Y ☐ N ☐ Not sure

If yes, specify: _____

In contact with wildlife species or their droppings: ☐ Y ☐ N ☐ Not sure

If yes, specify: _____

Had contact with soil, manure or sewage: ☐ Y ☐ N ☐ Not sure

If yes, specify: _____

SECTION J: ANY OTHER COMMENTS RELEVANT TO THIS CASE

_____

_____

_____

_____

_____

_____

_____

Can the case be contacted again if further details are required: ☐