A spatial and temporal analysis of risk factors associated with sporadic Shiga toxin-producing *Escherichia coli* O157 infection in England between 2009 and 2015.

Richard Elson[1,2,6], Katherine Grace[4], Roberto Vivancos[1,2,3], Claire Jenkins[1], Goutam. K. Adak [1, 2], Sarah J. O'Brien [2,5] and Iain R. Lake [2,6].

1. National Infection Service, Public Health England, United Kingdom.

2. National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Gastrointestinal Infections, United Kingdom.

3. National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Emerging and Zoonotic Infections, United Kingdom.

4. Animal and Plant Health Agency, London, United Kingdom.

5. University of Liverpool, Liverpool, United Kingdom.

6. School of Environmental Sciences, University of East Anglia, United Kingdom.

Author for correspondence:

Richard Elson

National Infection Service, Public Health England, 61, Colindale Avenue, London, NW9 5EQ.

Email:richard.elson@phe.gov.uk

Article summary line: Rural residence, presence of private water supplies and proximity to high densities of farmed animals is associated with increased risk of STEC O157 infection in England.

Key words: Shiga-toxin producing E. coli, ruminant density, rurality, spatial, temporal.

Summary

Infection with STEC O157 is relatively rare but has potentially serious sequelae, particularly for children. Large outbreaks have prompted considerable efforts designed to reduce transmission primarily from food and direct animal contact. Despite these interventions, numbers of infections have remained constant for many years and the mechanisms leading to many sporadic infections remain unclear.

Here we show that two thirds of all cases reported in England between 2009 and 2015 were sporadic. Crude rates of infection differed geographically and were highest in rural areas during the summer months. Living in rural areas with high densities of cattle, sheep or pigs and those served by private water supplies were associated with increased risk. Living in an area of lower deprivation contributed to increased risk but this appeared to be associated with reported travel abroad. Fresh water coverage and residential proximity to the coast were not risk factors.

To reduce the overall burden of infection in England, interventions designed to reduce the number of sporadic infections with STEC should focus on the residents of rural areas with high densities of livestock and the effective management of non-municipal water supplies. The role of sheep as a reservoir and potential source of infection in humans should not be overlooked.

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; *stx*1 and *stx*2. Compared to other bacterial pathogens, it is a relatively rare infection in many parts

of the world but is of public health concern due its low infectious dose (<100 bacteria)[1]and potential to cause severe disease (3).

Worldwide, it is estimated that there are around 2.8 million cases annually, leading to 3,890 cases of haemolytic uraemic syndrome (HUS) and 230 deaths [2]. The Europe wide rate of infection is estimated to be 1.4 cases per 100,000 population but reported rates vary between countries (Range: <0.1 to 12.4 cases per 100,000 population) [3]. The O157 STEC serogroup is most commonly associated with human disease in the UK however, other serogroups are seen more frequently in other European countries[3]. Rates of infection in England have remained fairly constant for many years (around [4]. Europe has shown a similar pattern with an increase since 2011 attributed to wider use of molecular methods following a large outbreak linked to sprouted fenugreek seeds [5]. Rates of infection are highest in children and most cases occur in the late summer, at least in temperate areas, and this pattern is seen universally [6].

Healthy cattle are the main reservoir of STEC although they are also carried by sheep and other animals [7]. Animals shed a range of phage types (PTs) with the most prevalent in UK cattle being PT21/28, 8 and 34, PT 4 in sheep and PT2 in pigs [8, 9]. STEC O157 survives well in the environment, remaining viable for many months in temperate conditions [10, 11].

Transmission to humans occurs through multiple routes. Cases present themselves sporadically (occurring independently of other cases) or as part of small outbreaks due to person-to-person spread in closed settings, particularly childcare facilities. The low infectious dose of STEC means that once in the population, person-to-person spread is common [1]. Larger outbreaks tend to be associated with foodborne transmission, with an increasing trend towards salad vegetables and away from meat and dairy products [4]. Direct contact with the farming environment or ruminants, such as in open farms or petting zoos [1], are important risk factors for STEC infection [12]. Indirect contact with animals or environments

contaminated with their faeces is also of importance but the actual process leading to infection is less well understood. Heavy rainfall and flooding events can lead to contamination of fresh and marine water systems [13], beaches [14-18] and food crops. Poorly managed private water supplies present a particular risk in areas of high animal density [19, 20].

Phylogenetic analysis of strains circulating in humans and UK cattle during 2014 described three distinct lineages (I, II and I/II) descended from a common ancestor. Lineage I contains PT 21/28 and PT32; strains encoding *stx*2 only and associated with more severe disease. Lineage II contains PT8 and Lineage I/II PT2. Isolates from humans and UK cattle are closely related suggesting that PT8 and, in particular, PT21/28, have a domestic source and are domestically acquired [21]. With the advent of routine whole genome sequencing (WGS), it is now possible to identify links between cases that previously appeared sporadic in nature. These cases may exhibit spatial clustering, sometimes over long periods of time, suggesting that geographically restricted transmission of highly related strains can occur [22].

Ecological studies in the UK, Europe and further afield demonstrate a spatial association between rates of infection, cattle density and other factors and all describe a seasonally driven picture with rates highest in the summer [23-29]. There are limitations to these studies. Sheep were only considered in one study [29], despite being a known reservoir. The study populations were restricted to children [25] or focussed upon severe disease [25]. Some studies may also have included cases linked to outbreaks which is not ideal as the source of their infection may have differed from sporadic cases [28]. Cases reporting foreign travel were included in some studies [25, 26, 28], but not others [23-25]. Finally, all of these studies were performed at serogroup level only, even though different subgroups may have different sources and hence potentially different risk factors.

In this study we overcome these limitations using enhanced surveillance of STEC which has been performed in England since 2009. These data arguably represent the most comprehensive dataset for STEC infections in the world. This allows accurate identification of cases who have been part of an outbreak (and so are not representative of all cases) and those who report travel abroad or within the UK

This study had three aims. The first was to describe the spatial and temporal distribution of sporadic STEC O157 cases in England, the second was to test the relationship between the numbers of infections and hypothesised risk factors, the third was to test whether these risks differed by STEC subtype. Finally we explored how these risks varied between all sporadic cases and sporadic cases when those reporting national or foreign travel are excluded.

Methods

Isolates of E. coli O157 identified locally are sent for confirmation and typing at the Gastrointestinal Bacterial Reference Unit (GBRU). Detection and confirmation of STEC 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 [4].

The National Enhanced Surveillance System for STEC (NESSS) was introduced in England in 2009. The system collects clinical and epidemiological information for each laboratory confirmed case using a standardised questionnaire. This information is linked to reference microbiology information including PT, presence of virulence factors and whole genome sequence data.

The case definition for the purpose of this study was a sporadic case of STEC O157, confirmed by GBRU and reported to the national enhanced surveillance system for STEC

(NESSS) between January 1st 2009 and December the 31st 2015. An overview of the data selection process is shown in Figure 1.

The main aim of our analysis was to estimate the effect of hypothesised risk factors on the occurrence of sporadic cases (i.e. those occurring independently of each other). We therefore excluded cases linked to known outbreaks because their residential location rarely reflects exposure to the source of their infection, particularly for large outbreaks linked to widely distributed foodstuffs. Cases linked to household outbreaks were also excluded. Household outbreaks were identified as those where at least two cases had isolates of the same serotype and phage type that were collected within six months of each other, processed by a laboratory in the same Health Protection Team area and sharing the same surname and/or UK postcode.

The postcode (an alphanumeric reference developed by the UK Post Office to facilitate the delivery of mail and each containing around 15 addresses) for each case was geocoded to provide a spatial reference, allow visual display of the location of cases and enable the details of each case to be spatially joined to other datasets at the Lower Super Output Area (LSOA) level defined by the Office for National Statistics [30]. LSOAs were chosen because they provide the most homogenous unit in terms of geographical size (Mean 3.9 Km^2, Range 0.02-684 Km^2) and population size (Mean 1,613 persons, Range 985-8,300 persons).

Crude incidence rates were calculated using the total population denominator data for each LSOA drawn from the last ONS Census performed in 2011.

For each LSOA, a dependent variable, indicating the number of cases that occurred during the study period was created. Because STEC is an uncommon infection, the majority (90.1%) of LSOAs had no cases, 9.3% had one case and less than 0.6% had more than one case.

Dependent variables were created for all PTs, PT21/28 and PT8, further divided into three classes (all reported cases, those not reporting foreign travel and those reporting no travel either within the UK or abroad) giving a total of nine dependent variables.

The following explanatory variables were constructed for each LSOA:

Livestock density variables for cattle, sheep and pigs were calculated using the agricultural census of 2010 [31]. This census is performed every ten years by the Department for the Environment, Food and Rural Affairs and collects detailed information on land usage and livestock populations. Farm level data are aggregated to a 5x5 km grid and individual farms are not identified.

Estimates of deprivation were obtained from the Office of National Statistics (ONS). The index of Multiple Deprivation (IMD) was obtained for England for 2011 [32] which provides a set of relative measures of deprivation for LSOAs. This is based on seven domains of deprivation (income, employment, education, health, crime, housing and living environment). These domains are combined and ranked to produce the overall IMD score for each LSOA. For our analysis these data were divided into quintiles where quintile 1 is most deprived.

The degree of rurality for each LSOA was derived from the ONS rural urban classification used to distinguish rural and urban areas in England and Wales in 2011[33]. The classification defines areas as rural if they are outside settlements with more than 10,000 resident population. For LSOAs, there are four urban classes (major conurbations, minor conurbations, cities and towns, cities and towns in sparse settings) and four rural classes (town and fringe, town and fringe in a sparse setting, village and dispersed settlements, village and dispersed settlements in sparse settings). Due to the small numbers of cases resident in areas considered sparse, we grouped these eight classes into five by merging the three sparse categories with the corresponding non-sparse categories.

Outbreaks have been linked to beaches [14, 18], hence the straight-line distance from the centroid of each LSOA to the GB coastline was calculated in kilometres.

Inland water coverage was identified as a risk factor in a Finnish study [27]. The shapes and areas of inland water features were extracted from the Ordnance Survey Master Map Topography Layer [34], summed and divided by the area of each LSOA to provide a proportional measure of fresh water coverage.

For each LSOA, the count of private water supplies was calculated using data submitted by local authorities to the Drinking Water Inspectorate during 2016. Local authorities are responsible for the enforcement and monitoring of the Private Water Supplies (PWS) Regulations 2005 which require PWS to meet certain standards and for the location of each supply to be recorded. Three classes were created (0, 1-20 and >20 supplies).

Because the datasets used for inland water coverage and animal density differed from LSOAs in terms of geographical area or shape, we used a geographical information system (GIS) overlay function to create proportional measure in Km^2 for each LSOA.

We used Jenks' natural breaks method to create four categorical variables for each animal species, distance from the coast and inland water coverage. This method is designed to determine the best arrangement of values into different classes by seeking to minimise the variance within classes and maximise the variance between classes [35].

Statistical analysis

We considered three methods of regression analysis: Poisson, negative binomial and zero inflated Poisson. The results of a likelihood ratio test of alpha and goodness of fit test following Poisson regression indicated that the data were overdispersed. The same analysis was repeated using negative binomial and zero inflated negative binomial regression

respectively. The Vuong test was not significant indicating that the standard negative binomial approach was best suited to the data. Proceeding with the negative binomial regression approach we conducted a multivariable analysis for each dependent and the independent variables. The first set of dependent variables were all sporadic cases, all sporadic cases minus those reporting foreign travel and all sporadic cases with no foreign or domestic travel. Two further models were then produced focussing upon PT21/28 and PT8 only. Person years (the total population of each LSOA multiplied by the years of observation) was included as an exposure variable. None of the multivariable analyses showed any associations with the distance from the coast and inland water coverage variable, hence these were removed from the analysis. The remaining independent variables were all included in the nine models to allow greater comparability between models. The dependent and independent variables were checked for correlation using Spearman's rank test. All coefficients showed low to moderate correlation with the exception of cattle and sheep density with a coefficient of 0.7. An analysis for collinearity indicated that the addition of each independent variable in turn did not lead to significant changes in the coefficients or significance of any other variables in the model. Presenting each livestock density variable to the model as continuous variables did not affect the results.

Results are presented in terms of Incidence Rate Ratio (IRR) estimates and the 95% Confidence Interval (CI). The overall significance for a variable was estimated using the Wald test. All statistical analyses were performed using Stata version 13 [36].

Results

Rates of infection

A total of 3,559 (34%) cases were eligible for inclusion in the statistical analysis (Figure 1).

The crude incidence of all sporadic confirmed STEC O157 cases (including those reporting foreign travel) reported during the study period was 9.1 per million person years. The rural rate (13.3 per million person years) was 1.6 times higher than that of the urban rate (8.1 per million person years). Rates varied across the country with the highest in the North of the country, the North West, Midlands and the South West Peninsula (Figure 2a) and this was seen each year during the study (Figure 3).

The crude incidence rate of PT21/28 was 2.5 per million person years and for PT8 it was 3.3 per million person years. There was a distinct seasonality both in rural and urban areas with rates comparable during the winter but higher in rural areas during the summer (Figure 3.). The rate of infection declined from 2012, particularly for PT21/28 infections in rural areas (Figure 3).

The spatial distribution of animals varied across the country (Figure 2 g-i). The mean cattle density ranged from 0-199 animals/Km^2 with the highest densities observed in the South West Peninsula, areas of the North West (Cheshire) and Midlands (Staffordshire) and in the North. Sheep density ranged from 0-572 animals/Km^2 with the highest densities observed along the Welsh Borders, Oxfordshire, the South West Peninsula and in the North. Pig density ranged from 0-499 animals/Km^2 with the highest densities observed in East Anglia and the North East.

Multivariable analysis

The results of the multivariable analysis for all sporadic cases are presented in Table 1.

These indicate that living in a rural village, in an area with high densities of farmed animals (cattle, sheep or pigs), the presence of PWS and those areas considered least deprived were positively associated with risk for all sporadic cases. Removing cases reporting foreign travel removed the effect seen in the IMD variable.

The dataset was then split into PT21/28 and PT8. The results for PT21/28 are presented in Table 2 and indicate that living in an area with high densities of farmed animals and being served by PWS were positively associated with risk. Living in a rural village was a risk factor only for cases who reported no travel. Areas regarded as the most deprived were associated with increased risk of PT 21/28 infection.

The results of the multivariable analysis for PT8 are presented in Table 3 and indicate that living in a rural village and areas with high densities of cattle and/or pigs and areas considered least deprived (Quintiles 4&5) were positively associated with risk. Removing cases reporting foreign travel from the dependent variable increased the risk effect for cattle and pig density but removed the effect of deprivation. PWS and sheep density were not significant predictors of infection for PT8.

Discussion

Risk was positively associated with cattle density across all models. The risk of a case occurring in areas with 87 animals/Km^2 or more was more than twice that of area with fewer than 18 animals/Km^2. This finding was somewhat expected as cattle are regarded as the main reservoir of STEC O157 [7] and the most common subtypes shed in cattle faeces are PT21/28 followed by PT8 [9].

Sheep density was positively associated with risk for all STEC O157 cases and PT21/28 cases but not for PT8. The greatest effect of sheep density was seen in the PT21/28 model where the risk was increased almost threefold in areas with high densities of sheep. There is increasing evidence that sheep and other small ruminants are an important reservoir of STEC and that sheep density is associated with non O157 STEC infections [29]. The association with PT21/28 is interesting because the carriage of PT21/28 in sheep is low (14%) compared to cattle (37%) [37] yet they appear strongly significant in our model. This exposure to sheep

and lambs has been linked to at least two PT21/28 outbreaks at petting farms/lambing events in England and an extended outbreak of closely related strains was linked to an ovine source. [38, 39]. A recent study in the Republic of Ireland demonstrated a geographical association between sheep density and STEC O26 but not STEC O157 [29]. Disentangling the relative contribution of ruminant species to the overall burden of infection is difficult due to scant contemporary information on shedding by sheep compared to cattle, a lack of genetic difference between cattle and sheep strains [40] and the fact that sheep and cattle are farmed in the same geographical areas in the UK. However, our results suggest that the role of sheep as a reservoir and potential source of infection in humans should not be overlooked.

Pig density was positively associated with risk across all models. However, for the PT8 and PT21/28 models, the observed effect was not linear. Pigs can shed STEC [41], and pork products have been implicated as the source or vehicle of infection in outbreaks worldwide, but they are not generally considered to be an important reservoir for STEC O157 [42] or source for human infection [43]. Studies of intestinal carriage in England showed a low carriage rate in pigs and that the characteristics of pig strains differed from those seen in humans during the same time period [9, 37, 44]. Pig density was not identified as a risk factor for STEC infection in the Netherlands [24]. In spite of this, associations with pig density appear in this study. This dichotomy may be partly explained by the few areas of the country with high pig densities or may be due to differences in pig husbandry practices between countries.

We found an increased risk associated with the presence of PWS for all STEC O157 cases, PT21/28 cases but not PT8. Private water supplies that do not meet the requirements of the EC directive present a high risk of infection with STEC [20], however, the reason for the difference between PTs is unclear and may relate to factors not considered by this study.

Living in a rural village was a risk factor for all STEC O157 cases and for PT8. For the PT21/28 model, living outside a major urban conurbation was a significant risk factor but only for cases reporting no travel. Residents of rural areas are more likely to come into contact with contaminated environments either through work or leisure [28] and residential proximity to the ruminant reservoir also increases the possibility of exposure from wildlife and insect vectors [45, 46].

Living in less deprived areas was strongly associated with all STEC O157 cases and PT8. What is intriguing was that when foreign travel cases were removed, this effect was removed for all STEC O157 cases and PT8. This indicates that the deprivation effect is strongly driven by foreign travel and that risk factors for these groups differ from indigenously acquired infection. The strong association with foreign travel in the PT8 model was anticipated as a greater proportion of cases of PT8 report travel abroad compared to other PTs [1]. Lower deprivation was protective for cases of PT21/28. reporting no travel but the reasons for this are unclear. One explanation could be related to deprivation and/or likelihood of exposure to PT21/28 compared to other strains. PT21/28 is a strain indigenous to the UK and rates of infection are highest in the north of England where there are also more areas considered to be the most deprived. However, crude rates of infection with STEC are lower in the most deprived areas, cases are less likely to travel abroad or within the UK [47] and levels of social interaction differ from residents of less deprived areas [48]. Socioeconomic status has been shown to be associated with risk for other gastrointestinal infections [49-51] introducing the possibility that whilst risk factors may differ broadly across the country, within regions, socio-economic status has a greater influence on risk factors and transmission dynamics. This is an area that would benefit from further research.

In developed countries, disadvantaged children, but not adults, appear to be at greater risk of gastrointestinal infection and that living in deprived areas is protective for infectious

intestinal disease (IID) overall but is associated with more severe symptoms in those who do become infected [52] . Our findings indicate that once the effect of foreign travel is removed, deprivation has little effect on sporadic infection with STEC O157. This suggests that infection is a result of a localised and stochastic process driven by exposure to the local environment and that exposures related to affluence, such as diet, occupation or leisure pursuits are likely to be less important.

Residential distance from the coast and living in an area with a high proportion of fresh water coverage were not significant and removed at an early stage of the modelling process. These environments may act as reservoirs for STEC [53] and recreational exposure to fresh water has been suggested as a risk factor for STEC infection in epidemiological studies from other countries. The lack of an association in the UK may relate to the general unpopularity of freshwater swimming in the UK in comparison to other countries [54-57]. These variables were proxies for exposure and so do not capture details of individual interactions with these environments.

There are several potential limitations to our study. First, molecular typing methods, used routinely in England since 2015, are superior to the phenotypic methods we used to discriminate between cases and have been shown to reduce the number of cases considered to be sporadic [58]. It's therefore possible that a small number of cases included in our study were microbiologically linked and therefore may not be considered truly sporadic. . Secondly, we excluded cases linked to household outbreaks. We made this decision based on the difficulty in identifying the primary case amongst co-primary and asymptomatic (microbiologically confirmed) cases generated by microbiological screening of household contacts. In addition, we noted epidemiological links between foodborne outbreaks and secondary transmission within households which may have introduced a bias away from the exposures of interest. Thirdly, for every STEC O157 infection reported to national

surveillance systems in England, there are an estimated 7.4 in the community [59]. The reasons for this are likely to be related to severity of disease, health seeking behaviours and whether or not a clinician takes a sample and requests a microbiological examination from a laboratory. Notwithstanding that the ratio of STEC O157 reports is considerably smaller than other pathogens, the cases reported to national surveillance represent a biased sample of the true community burden of STEC O157 in England. Transmission routes are varied and infection is a result of complex interactions between people and their local environment. Our approach meant that individually reported exposures could not be considered in our analysis. Finally, this is an ecological study and association does not equal causation which could only be inferred from other study designs involving an intervention. We are confident that the association with animal density is most likely due to environmental exposure, however, other factors not included in our study (for example, locally sourced food) cannot be ruled out as a potential route of infection.

In conclusion, using arguably one of most comprehensive enhanced surveillance of STEC datasets in the world, we found that two thirds of infections were sporadic and that the spatial and temporal distribution of these cases showed distinct variation within England. We provide evidence that living in a rural area with high densities of farmed animals and served by private water supplies partly explain this variation. Our results indicate that travel abroad may expose individuals to risks not present in their local residential environment and that this risk is influenced by socio-economic status. Further analysis is required to elucidate the relative importance of exposures reported by individual cases including travel, contact with animals and the agricultural environment and consumption of food and water.

To reduce the overall burden of infection in England, interventions designed to reduce the number of sporadic infections with STEC O157 should focus on the residents of rural areas with high densities of livestock and the effective management of non-municipal water supplies.

Acknowledgements

We thank Marie Anne Chattaway, Vivienne DoNascimento, and Neil Perry for their microbiology expertise. We also acknowledge the roles of Amy Mikhail, Lisa Byrne, Kirsten Glen and Nalini Purohit for the maintenance of the enhanced surveillance system for STEC in England and extend thanks to Sue Pennison of the Drinking Water Inspectorate for sharing their data on PWS locations. 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 the National Infections Service of Public Health England.

The research was funded by the National Institute for Health Research Health Protection Research Unit (NIHR HPRU) in Gastrointestinal Infections at University of Liverpool in partnership with Public Health England (PHE), in collaboration with University of East Anglia, University of Oxford and the Institute of Food Research.

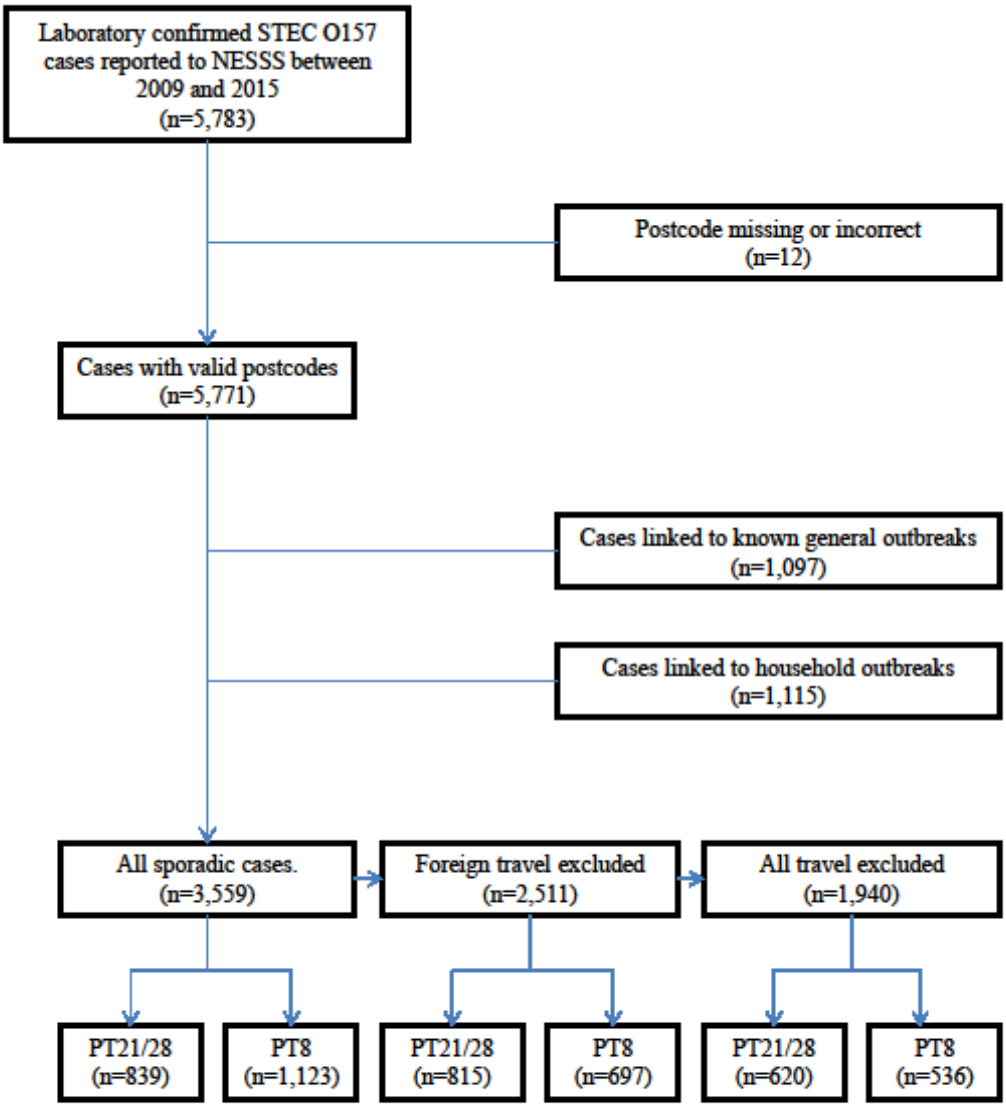
Richard Elson is based at Public Health England. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health or Public Health England.

Biographical sketch

Richard Elson is an epidemiologist employed by Public Health England with a particular interest in the spatial and spatio-temporal distribution of gastrointestinal infections.

Word count: 4342 including summary.

Figure 1. Flow chart for case selection.



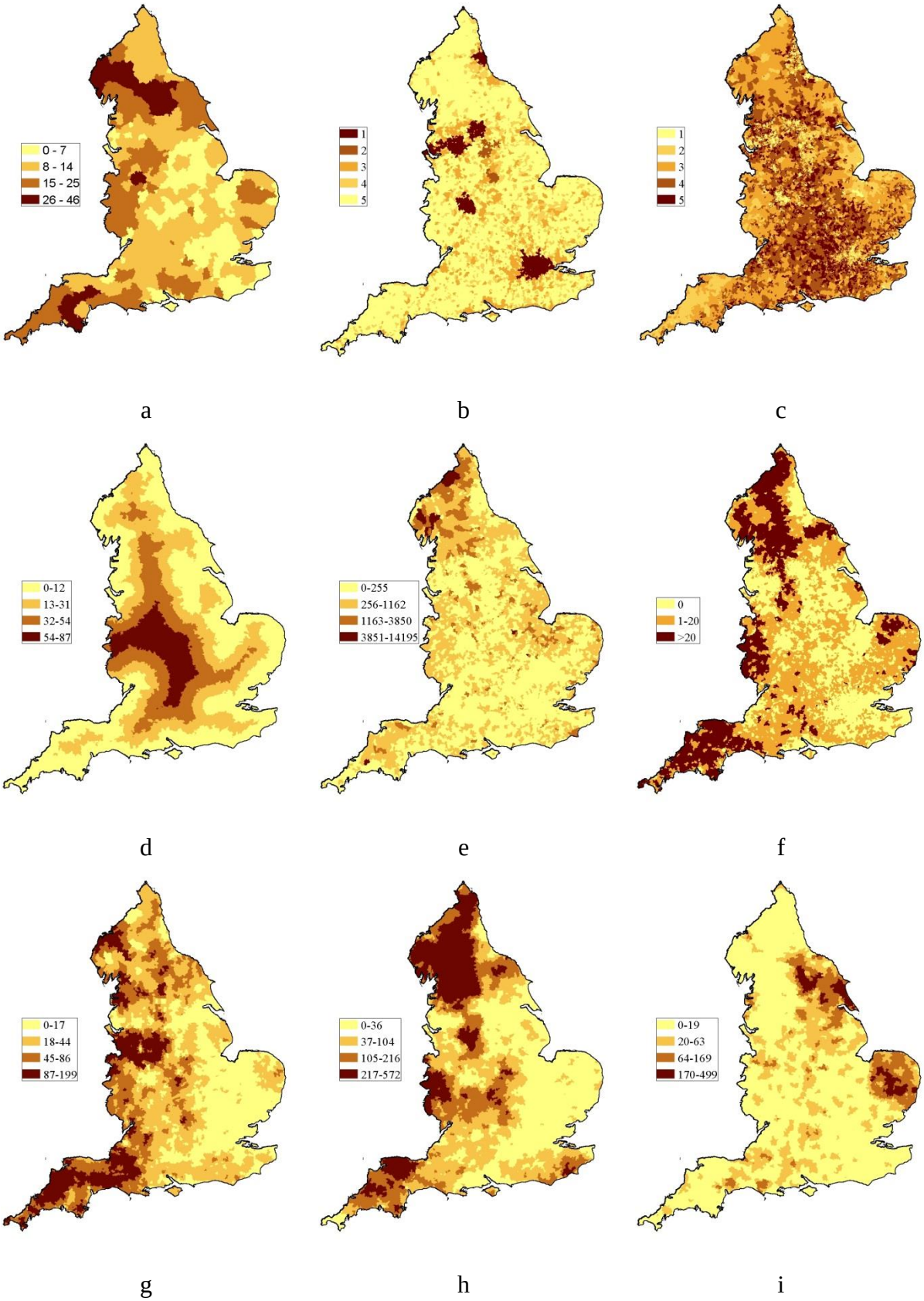


Figure 2. a. Cumulative incidence rate (Sporadic cases/million person years) and spatial distribution of the eight independent variables used in the analysis b.) Residence (1: Urban-Major conurbation, 2: Urban-Minor conurbation, 3: Urban-City and town, 4: Rural-Town and fringe, 5:Rural-Village. c.) IMD (deciles). d.) Distance from LSOA centroid to GB coast in Km. e.) Proportion of inland fresh water coverage of each LSOA (Km^2). f.) Numbers of private water supplies in each LSOA. g.) Cattle density Animals/Km^2). h.) Sheep density (Animals/Km^2). i.) Pig density (Animals/Km^2).

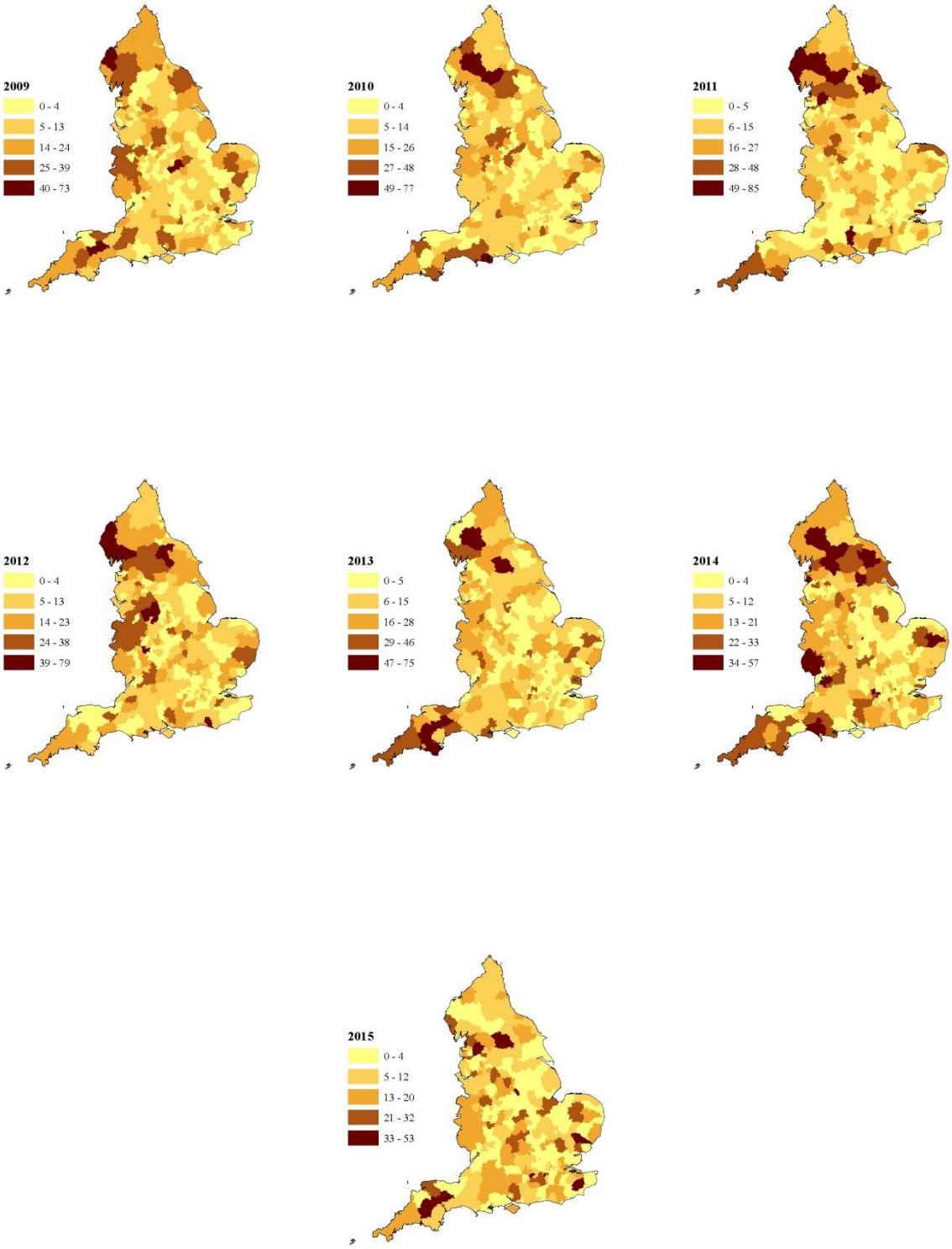


Figure 3. Annual incidence rate of STEC O157 per million population including cases reporting travel outside the UK in England between 2009-2015. (Unit of analysis is a local authority area)

Figure 4. Monthly rate of sporadic STEC O157 infection per million population in urban and rural settings in England between 2009 and 2015 (Travel included).

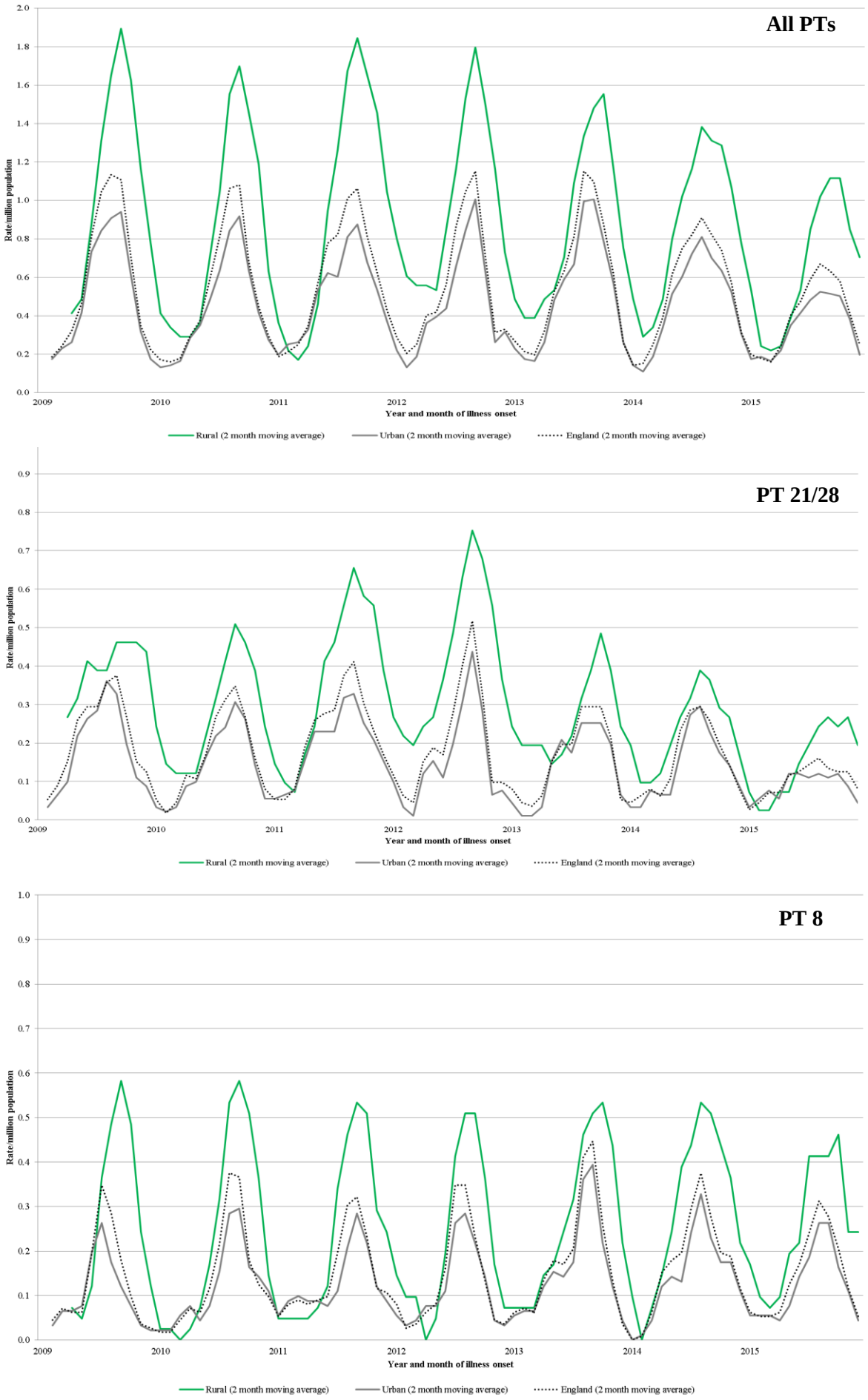


Table 1. Results of multivariable analysis. All STEC O157 PTs

Variable	Detail	All (n=3,559)			Excluding foreign travel (n=2,511)			Excluding all travel (n=1,940)		
		IRR	95% CI	P	IRR	95% CI	P	IRR	95% CI	P
Cattle density (Animals/Km^2)	0-17	1	-	<0.001			<0.001			<0.001
	18-44	1.25	1.14 - 1.37		1.34	1.19 - 1.50		1.28	1.12 - 1.45	
	45-86	1.33	1.16 - 1.51		1.52	1.30 - 1.77		1.49	1.25 - 1.78	
	87-199	1.84	1.58 - 2.14		2.24	1.88 - 2.67		2.24	1.83 - 2.74	
Sheep density (Animals/Km^2)	0-36	1	-	<0.001			<0.001			<0.001
	37-104	1.09	0.98 - 1.20		1.08	0.96 - 1.21		1.13	0.98 - 1.29	
	105-216	1.20	1.05 - 1.38		1.25	1.07 - 1.46		1.26	1.05 - 1.51	
	217-572	1.56	1.30 - 1.88		1.63	1.32 - 2.01		1.71	1.34 - 2.17	
Pig density (Animals/Km^2)	0-19	1	-	0.005			<0.001			0.015
	20-63	1.03	0.93 - 1.14		1.10	0.98 - 1.23		1.08	0.94 - 1.23	
	64-169	1.29	1.11 - 1.51		1.43	1.20 - 1.71		1.40	1.15 - 1.71	
	170-499	1.35	0.94 - 1.93		1.59	1.07 - 2.35		1.46	0.91 - 2.32	
Number of private water supplies	0	1	-	0.003			0.002			0.019
	1-20	1.19	1.06 - 1.34		1.22	1.07 - 1.41		1.18	1.01 - 1.39	
	>20	1.35	1.09 - 1.68		1.48	1.16 - 1.88		1.44	1.09 - 1.89	
Residence	Urban: Major conurbation	1	-	<0.001			<0.001			<0.001
	Urban: Minor conurbation	1.23	1.02 - 1.49		1.18	0.93 - 1.49		1.33	1.04 - 1.71	
	Urban: City and town	1.06	0.97 - 1.16		1.04	0.94 - 1.17		1.01	0.89 - 1.14	
	Rural: Town and Fringe	1.12	0.98 - 1.28		1.16	0.99 - 1.37		1.17	0.98 - 1.41	
	Rural: Village	1.43	1.23 - 1.66		1.54	1.29 - 1.84		1.61	1.31 - 1.97	
IMD	1 (Most deprived)	1	-	<0.001			0.965			0.082
	2	1.04	0.92 - 1.17		1.01	0.88 - 1.15		0.90	0.77 - 1.04	
	3	1.12	1.00 - 1.26		1.03	0.90 - 1.18		0.88	0.75 - 1.02	
	4	1.13	1.01 - 1.27		0.99	0.86 - 1.13		0.80	0.69 - 0.94	
	5(Least **deprived**)	1.27	1.14 - 1.43		1.02	0.89 - 1.17		0.85	0.73 - 1.00	

. **Table 2. Results of multivariable analysis – STEC O157 PT21/28.**

Variable	Detail	All (n=839)			Excluding foreign travel (n=815)			Excluding all travel (n=620)		
		IRR	95% CI	P	IRR	95% CI	P	IRR	95% CI	P
Cattle density (Animals/Km^2)	0-17	1	-	<0.001	1	-	<0.001	1	-	0.002
	18-44	1.42	1.17 - 1.73		1.43	1.17 - 1.74		1.34	1.07 - 1.69	
	45-86	1.36	1.04 - 1.77		1.35	1.03 - 1.77		1.25	0.91 - 1.71	
	87-199	1.98	1.46 - 2.68		2.01	1.48 - 2.73		1.93	1.36 - 2.76	
Sheep density (Animals/Km^2)	0-36	1	-	<0.001	1	-	<0.001	1	-	<0.001
	37-104	1.19	0.97 - 1.46		1.21	0.98 - 1.50		1.27	1.00 - 1.62	
	105-216	1.83	1.42 - 2.38		1.85	1.43 - 2.41		2.04	1.50 - 2.77	
	217-572	2.48	1.77 - 3.46		2.49	1.78 - 3.50		2.75	1.85 - 4.09	
Pig density (Animals/Km^2)	0-19	1	-	0.029	1	-	0.021	1	-	0.034
	20-63	1.10	0.90 - 1.34		1.12	0.92 - 1.37		1.02	0.80 - 1.30	
	64-169	1.38	1.01 - 1.88		1.39	1.02 - 1.90		1.60	1.13 - 2.25	
	170-499	2.02	1.11 - 3.69		2.09	1.15 - 3.83		1.71	0.79 - 3.73	
Number of private water supplies	0	1	-	0.002	1	-	0.002	1	-	0.056
	1-20	1.35	1.07 - 1.70		1.34	1.06 - 1.69		1.10	0.83 - 1.46	
	>20	1.94	1.33 - 2.82		1.94	1.32 - 2.84		1.70	1.09 - 2.64	
Residence	Urban: Major conurbation	1	-	0.212	1	-	0.193	1	-	0.018
	Urban: Minor conurbation	1.23	0.83 - 1.83		1.23	0.82 - 1.84		1.34	0.87 - 2.05	
	Urban: City and town	1.00	0.83 - 1.21		1.00	0.83 - 1.22		0.92	0.74 - 1.15	
	Rural: Town and Fringe	1.09	0.83 - 1.44		1.11	0.84 - 1.47		1.09	0.79 - 1.51	
	Rural: Village	1.35	0.99 - 1.83		1.37	1.01 - 1.87		1.52	1.07 - 2.18	
IMD	1 (Most deprived)	1	-	0.256	1	-	0.167	1	-	0.002
	2	0.91	0.73 - 1.14		0.89	0.70 - 1.11		0.75	0.58 - 0.96	
	3	0.79	0.62 - 0.99		0.76	0.60 - 0.97		0.65	0.50 - 0.84	
	4	0.80	0.63 - 1.01		0.77	0.61 - 0.98		0.60	0.46 - 0.78	
	5(Least deprived)	0.85	0.68 - 1.08		0.83	0.66 - 1.06		0.67	0.51 - 0.87	

Table 3. Results of multivariable analysis – STEC O157 PT8.

Variable	Detail	All (n=1,123)			Excluding foreign travel (n=697)			Excluding all travel (n=536)		
		IRR	95% CI	P	IRR	95% CI	P	IRR	95% CI	P
Cattle density (Animals/Km^2)	0-17	1	-	<0.001	1	-	<0.001	1	-	<0.001
	18-44	1.35	1.15 - 1.59		1.44	1.17 - 1.77		1.39	1.10 - 1.77	
	45-86	1.55	1.24 - 1.94		1.85	1.40 - 2.45		1.95	1.42 - 2.68	
	87-199	1.90	1.45 - 2.49		2.30	1.66 - 3.19		2.24	1.54 - 3.27	
Sheep density (Animals/Km^2)	0-36	1	-	0.790	1	-	0.420	1	-	0.076
	37-104	1.06	0.89 - 1.26		1.05	0.85 - 1.30		1.11	0.87 - 1.42	
	105-216	1.01	0.79 - 1.29		0.83	0.61 - 1.13		0.73	0.51 - 1.06	
	217-572	1.16	0.82 - 1.64		1.05	0.70 - 1.60		1.19	0.76 - 1.88	
Pig density (Animals/Km^2)	0-19	1	-	0.021	1	-	0.013	1	-	0.038
	20-63	1.12	0.95 - 1.33		1.20	0.98 - 1.48		1.25	0.99 - 1.58	
	64-169	1.46	1.13 - 1.88		1.58	1.16 - 2.14		1.54	1.09 - 2.19	
	170-499	1.42	0.77 - 2.62		1.54	0.75 - 3.16		1.54	0.67 - 3.53	
Number of private water supplies	0	1	-	0.932	1	-	0.467	1	-	0.372
	1-20	1.00	0.81 - 1.24		1.06	0.82 - 1.37		1.07	0.79 - 1.44	
	>20	1.08	0.72 - 1.61		1.32	0.85 - 2.07		1.42	0.87 - 2.34	
Residence	Urban: Major conurbation	1	-	0.011	1	-	0.001	1	-	0.003
	Urban: Minor conurbation	1.20	0.85 - 1.68		1.17	0.75 - 1.82		1.51	0.96 - 2.37	
	Urban: City and town	1.06	0.90 - 1.24		1.10	0.89 - 1.36		1.07	0.84 - 1.36	
	Rural: Town and Fringe	1.16	0.92 - 1.47		1.20	0.89 - 1.63		1.28	0.91 - 1.80	
	Rural: Village	1.58	1.21 - 2.05		1.92	1.39 - 2.65		1.94	1.34 - 2.81	
IMD	1 (Most deprived)	1	-	0.001	1	-	0.246	1	-	0.764
	2	1.17	0.94 - 1.45		1.09	0.83 - 1.43		1.04	0.77 - 1.41	
	3	1.32	1.07 - 1.64		1.31	1.01 - 1.71		1.19	0.89 - 1.60	
	4	1.38	1.12 - 1.70		1.23	0.94 - 1.61		1.07	0.79 - 1.44	
	5(Least deprived)	1.52	1.24 - 1.88		1.26	0.96 - 1.64		1.05	0.77 - 1.42	

References

(1) **Byrne L, et al.** The epidemiology, microbiology and clinical impact of Shiga toxin-producing Escherichia coli in England, 2009-2012. *Epidemiology and Infection* 2015; **143**(16): 3475-3487.

(2) **Majowicz SE, et al.** Global incidence of human Shiga toxin-producing Escherichia coli infections and deaths: a systematic review and knowledge synthesis. *Foodborne Pathogens and Disease* 2014; **11**(6): 447-455.

(3) **European Centre for Disease Prevention and Control.** Annual Epidemiological Report 2016 – Shigatoxin/verocytotoxin-producing Escherichia coli infection. Stockholm; 2016.

(4) **Adams NL, et al.** Shiga Toxin-Producing Escherichia coli O157, England and Wales, 1983-2012. *Emerging Infectious Diseases* 2016; **22**(4): 590-597.

(5) **Buchholz U, et al.** German outbreak of Escherichia coli O104:H4 associated with sprouts. *New England Journal of Medicine* 2011; **365**(19): 1763-1770.

(6) **Lal A, et al.** Seasonality in human zoonotic enteric diseases: a systematic review. *PloS One* 2012; **7**(4): e31883.

(7) **Persad AK, LeJeune JT.** Animal Reservoirs of Shiga Toxin-Producing Escherichia coli. *Microbiology Spectrum* 2014; **2**(4): EHEC-0027-2014.

(8) **Chapman PA, et al.** Escherichia coli O157 in cattle and sheep at slaughter, on beef and lamb carcasses and in raw beef and lamb products in South Yorkshire, UK. *International Journal of Food Microbiology* 2001; **64**(1-2): 139-150.

(9) **Milnes AS, et al.** Intestinal carriage of verocytotoxigenic Escherichia coli O157, Salmonella, thermophilic Campylobacter and Yersinia enterocolitica, in cattle, sheep and pigs at slaughter in Great Britain during 2003. *Epidemiology and Infection* 2008; **136**(6): 739-751.

(10) **Avery SM, Moore A, Hutchison ML.** Fate of Escherichia coli originating from livestock faeces deposited directly onto pasture. *Letters in Applied Microbiology* 2004; **38**(5): 355-359.

(11) **Williams AP, et al.** Persistence of Escherichia coli O157 on farm surfaces under different environmental conditions. *Journal of Applied Microbiology* 2005; **98**(5): 1075-1083.

(12) **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. *Emerging Infectious Diseases* 2001; **7**(6): 1049-1051.

(13) **Williams AP, et al.** Leaching of bioluminescent Escherichia coli O157:H7 from sheep and cattle faeces during simulated rainstorm events. *Journal of Applied Microbiology* 2008; **105**(5): 1452-1460.

(14) **Anon.** VTEC O157 outbreak linked to beach holidays. *Communicable Disease Report CDR Weekly* 1999; **9**(37): 327, 330.

(15) **Paunio M, et al.** Swimming-associated outbreak of Escherichia coli O157:H7. *Epidemiology and Infection* 1999; **122**(1): 1-5.

(16) **Bruneau A, et al.** Outbreak of E. coli O157:H7 associated with bathing at a public beach in the Montreal-Centre region. *Canada Communicable Disease Report* 2004; **30**(15): 133-136.

(17) **Harrison S, Kinra S.** Outbreak of Escherichia coli O157 associated with a busy bathing beach. *Communicable Disease and Public Health* 2004; **7**(1): 47-50.

(18) **Ihekweazu C, et al.** Outbreak of E. coli O157 infection in the south west of the UK: risks from streams crossing seaside beaches. *Euro Surveillance* 2006; **11**(4): 128-130.

(19) **Richardson HY, et al.** Microbiological surveillance of private water supplies in England: the impact of environmental and climate factors on water quality. *Water Research* 2009; **43**(8): 2159-2168.

(20) **Riseboro H lI, Pitchers R, Hunter PR.** . Risk Assessment of VTEC Infections in English and Welsh Drinking Water report for contract DWI/70/2/256.

(21) **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. *Microbial Genomics* 2015; **1**(3).

(22) **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. *Epidemiology and Infection* 2016; **144**(13): 2812-2823.

(23) **Innocent GT, et al.** Spatial and temporal epidemiology of sporadic human cases of Escherichia coli O157 in Scotland, 1996-1999. *Epidemiology and Infection* 2005; **133**(6): 1033-1041.

(24) **Friesema IH, et al.** Geographical association between livestock density and human Shiga toxin-producing Escherichia coli O157 infections. *Epidemiology and Infection* 2011; **139**(7): 1081-1087.

(25) **Haus-Cheymol R, et al.** Association between indicators of cattle density and incidence of paediatric haemolytic-uraemic syndrome (HUS) in children under 15 years of age in France between 1996 and 2001: an ecological study. *Epidemiology and Infection* 2006; **134**(4): 712-718.

(26) **Frank C, et al.** Cattle density and Shiga toxin-producing Escherichia coli infection in Germany: increased risk for most but not all serogroups. *Vector Borne and Zoonotic Diseases* 2008; **8**(5): 635-643.

(27) **Jalava K, et al.** Agricultural, socioeconomic and environmental variables as risks for human verotoxigenic Escherichia coli (VTEC) infection in Finland. *BMC Infectious Diseases* 2011; **11**: 275.

(28) **Pearl DL, et al.** A multi-level approach for investigating socio-economic and agricultural risk factors associated with rates of reported cases of Escherichia coli O157 in humans in Alberta, Canada. *Zoonoses and Public Health* 2009; **56**(8): 455-464.

(29) **Ohaiseadha C, et al.** A geostatistical investigation of agricultural and infrastructural risk factors associated with primary verotoxigenic E. coli (VTEC) infection in the Republic of Ireland, 2008-2013. *Epidemiology and Infection* 2017; **145**(1): 95-105.

(30) **Office for National Statistics.** Census Geography. April 2017.

(31) **Department for the Environment, Food and Rural Affairs.** June Survey of Agriculture and Horticulture: Methodology. 2013.

(32) **Office for National Statistics.** English indices of deprivation 2015. April 2017.

(33) **Bibby P., Brindley P.** School of Computer Science, University of Nottingham. Urban and Rural Area Definitions for Policy Purposes in England and Wales: Methodology (v1.0). 2013.

(34) **Ordnance Survey.** OS MasterMap® Topography Layer

(35) **Jenks GF, Caspall FC.** Error on choroplethic maps:Definition, measurement, reduction. *Annals of the Association of American Geographers* 1971; **61**(2): 217-244.

(36) **StataCorp. 2011. Stata Statistical Software: Release 13. College Station TSL.**

(37) **Milnes AS, et al.** Factors related to the carriage of Verocytotoxigenic E. coli, Salmonella, thermophilic Campylobacter and Yersinia enterocolitica in cattle, sheep and pigs at slaughter. *Epidemiology and Infection* 2009; **137**(8): 1135-1148.

(38) **Mikhail A, Jenkins, C.,Dallman, T.,Douglas,A., Wight, A. Lighthill, J., Elviss,N., Elson,R.** Evolution of outbreak clusters detected by whole genome sequencing; investigation of phylogenetically linked sporadic cases reveals an ovine source. *PHE Research & Applied Epidemiology Scientific Conference*. Warwick, United Kingdom, 2017.

(39) **Mikhail AFW, et al.** An outbreak of Shiga toxin-producing Escherichia coli O157:H7 associated with contaminated salad leaves: epidemiological, genomic and food trace back investigations. *Epidemiology and Infection* 2018; **146**(2): 187-196.

(40) **Strachan NJ, et al.** Escherichia coli O157: burger bug or environmental pathogen? *International Journal of Food Microbiology* 2006; **112**(2): 129-137.

(41) **Cha W, et al.** Prevalence and characteristics of Shiga toxin-producing Escherichia coli in finishing pigs: Implications on public health. *International Journal of Food Microbiology* 2018; **264**: 8-15.

(42) **Tseng M, et al.** Shiga toxin-producing Escherichia coli in swine: the public health perspective. *Animal Health Research Reviews* 2014; **15**(1): 63-75.

(43) **Mughini-Gras L, et al.** Attribution of human infections with Shiga toxin-producing Escherichia coli (STEC) to livestock sources and identification of source-specific risk factors, The Netherlands (2010-2014). *Zoonoses and Public Health* 2018; **65**(1): e8-e22.

(44) **Chapman PA, et al.** A 1-year study of Escherichia coli O157 in cattle, sheep, pigs and poultry. *Epidemiology and Infection* 1997; **119**(2): 245-250.

(45) **Nielsen EM, et al.** Verocytotoxin-producing Escherichia coli in wild birds and rodents in close proximity to farms. *Applied and Environmental Microbiology* 2004; **70**(11): 6944-6947.

(46) **Nichols GL.** Fly transmission of Campylobacter. *Emerging Infectious Diseases* 2005; **11**(3): 361-364.

(47) **Adams NL, et al.** Socioeconomic status and Shiga toxin-producing *Eschericia coli* infecion in England: risk factors and clinical features. In: Public Health England.

(48) **Rotheram S, et al**. An ethnographic approach to examining health inequalities in gastrointestinal infections. In: Public Health Research and Science Conference. University of Warwick, UK. 2018.

(49) **Lake IR, et al.** Case-control study of environmental and social factors influencing cryptosporidiosis. *European Journal of Epidemiology* 2007; **22**(11): 805-811.

(50) **Simonsen J, Frisch M, Ethelberg S.** Socioeconomic risk factors for bacterial gastrointestinal infections. *Epidemiology* 2008; **19**(2): 282-290.

(51) **Newman KL, et al.** The impact of socioeconomic status on foodborne illness in high-income countries: a systematic review. *Epidemiology and Infection* 2015; **143**(12): 2473-2485.

(52) **Adams NL, et al.** Socioeconomic status and infectious intestinal disease in the community: a longitudinal study (IID2 study). *European Journal of Public Health* 2018; **28**(1): 134-138.

(53) **Williams AP, et al.** Persistence, dissipation, and activity of Escherichia coli O157:H7 within sand and seawater environments. *FEMS Microbiology Ecology* 2007; **60**(1): 24-32.

(54) **Joosten R, et al.** Risk factors for gastroenteritis associated with canal swimming in two cities in the Netherlands during the summer of 2015: A prospective study. *PloS One* 2017; **12**(4): e0174732.

(55) **Schets FM, De Roda Husman AM, Havelaar AH.** Disease outbreaks associated with untreated recreational water use. *Epidemiology and Infection* 2011; **139**(7): 1114-1125.

(56) **Sartorius B, et al.** Outbreak of norovirus in Vastra Gotaland associated with recreational activities at two lakes during August 2004. *Scandinavian Journal of Infectious Diseases* 2007; **39**(4): 323-331.

(57) **Marion JW, et al.** Association of gastrointestinal illness and recreational water exposure at an inland U.S. beach. *Water Research* 2010; **44**(16): 4796-4804.

(58) **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**(1): e85901.

(59) **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.