Shiga toxigenic Escherichia coli incidence is related to small area variation in cattle density in a region in Ireland

Title	Shiga toxigenic Escherichia coli incidence is related to small area variation in cattle density in a region in Ireland
Author(s)	Brehony, Carina;Cullinan, John;Cormican, Martin;Morris, D.
Publication Date	2018-05-12
Publisher	Elsevier
Repository DOI	10.1016/j.scitotenv.2018.05.038

SHIGA TOXIGENIC *ESCHERICHIA COLI* INCIDENCE IS RELATED TO SMALL AREA VARIATION IN CATTLE DENSITY IN A REGION IN IRELAND.

C. Brehony[1], J. Cullinan[2], M. Cormican[1,3], and D. Morris[1]

1. Antimicrobial Resistance and Microbial Ecology Group, School of Medicine, National University of Ireland, Galway, Ireland
2. School of Business & Economics, National University of Ireland, Galway, Ireland
3. Department of Medical Microbiology, University Hospital Galway, Galway, Ireland

Abstract

Introduction:

Shiga toxigenic *Escherichia coli* (STEC) are pathogenic *E. coli* that cause infectious diarrhoea. In some cases infection may be complicated by renal failure and death. The incidence of human infection with STEC in Ireland is the highest in Europe. The objective of the study was to examine the spatial incidence of human STEC infection in a region of Ireland with significantly higher rates of STEC incidence than the national average and to identify possible risk factors of STEC incidence at area level.

Methods:

Anonymised laboratory records (n=379) from 2009-2015 were obtained from laboratories serving three counties in the West of Ireland. Data included location and sample date. Population and electoral division (ED) data were obtained from the Irish 2011 Census of Population. STEC incidence was calculated for each ED (n=498) and used to map hotspots/coldspots using the Getis-Ord Gi* spatial statistic and significant spatial clustering using the Anselin's Local Moran's I statistic. Multivariable regression analysis was used to consider the importance of a number of potential predictors of STEC incidence.

Results:

Incidence rates for the seven-year period ranged from 0 to 10.9 cases per 1000. A number of areas with significant local clustering of STEC incidence as well as variation in the spatial distribution of the two main serogroups associated with disease in the region *i.e.* O26 and O157 were identified. Cattle density was found to be a statistically significant predictor of STEC in the region.

Conclusions:

GIS analysis of routine data indicates that cattle density is associated STEC infection in this high incidence region. This finding points to the importance of agricultural practices for human health and the importance of a “one-health” approach to public policy in relation to agriculture, health and environment.

Keywords:

STEC, GIS, public-health, clustering, risk-factors, zoonosis

1. Introduction

Shiga toxigenic *Escherichia coli* (STEC) also known as Verocytotoxigenic *E. coli* (VTEC) are pathogenic *E. coli* that are associated with a spectrum of illness. After exposure some people remain asymptomatic whereas others develop self-limiting gastroenteritis or bloody diarrhoea (haemorrhagic colitis). Haemolytic uraemic syndrome (HUS), characterized by acute renal failure, occurs in 3-7% of cases, with those most at risk being those aged less than 5 years of age [1]. The natural host for STEC are ruminants. Some animals excrete the microorganism in very high numbers (termed 'super-shedders') in their faeces [2][3]. The risk to humans is posed from contact with animal faecal matter, through direct contact with the animals or indirectly through the contaminated environment *e.g.* recreational or drinking water, or contaminated foodstuffs *e.g.* raw milk, raw meat [4]. Managing the risk to humans from STEC is particularly challenging as it has a very low infectious dose (<100 bacterial cells) and it can survive in the environment for extended periods of time [5].

The incidence of human infection with STEC in Ireland has been the highest in Europe since 2008; it was ten times the EU average in 2016 with 15.6 cases per 100,000 inhabitants in Ireland versus an average EU incidence of 1.77/100,000 [6]. Within Ireland there is significant regional variation in the incidence of STEC infection. Among the high incidence areas is a region of three counties (Galway, Mayo and Roscommon) in the west of Ireland. In this region the crude incidence in 2016 was 26 per 100,000 inhabitants compared to the national average of 18 per 100,000 inhabitants [1]. By contrast the incidence in a region in the east of Ireland comprised of counties Dublin, Kildare and Wicklow was 7.7 per 100,000 inhabitants.

In Ireland, the most commonly recognised modes of STEC transmission are direct animal contact, person-to-person and waterborne [1][7]. However, for a significant number (13%) of cases each year, the mode of transmission is reported as 'unknown' or 'unspecified' [1]. Small group water supplies, untreated private wells, livestock density and domestic wastewater treatment have all been implicated in STEC transmission and in STEC outbreaks [1][8][9].

The aim of this project was to examine the spatial incidence of STEC infection in a region of Ireland with higher incidence than the national average and to model the relationship between STEC incidence and a range of potential predictors at a geographical level. Diagnostic testing for STEC in the region is concentrated primarily in one laboratory which facilitated data access.

2. Methods

Ethical approval for this study was granted by the National University of Ireland, Galway Research Ethics Committee. Anonymised laboratory data for confirmed STEC cases from 2009 to 2015 inclusive were obtained from diagnostic laboratories serving the three counties Galway, Mayo and Roscommon. Duplicates and possible linked confirmed STEC cases were removed leaving 'primary cases' (n=379). Possible linked STEC cases were defined as those that were from the same electoral division (ED) with the same serogroup and within one month of a 'primary case'. Metadata recorded for each case included location geocoded at ED level and clinical laboratory sample receipt date. There were 498 EDs in the region studied. Population and ED data including number of households, land surface area (km^2), household water source and household wastewater treatment, were obtained from the 2011 Census of Population for Ireland from the Central Statistics Office [10]. ED population ranged from 83 to 14,384 individuals and total land surface area ranged from 0.56 km^2 to 162.3 km^2. Population data was used to calculate the cumulative incidence rate (per 1000 inhabitants) of confirmed STEC cases for each ED for the study period. Data on cattle and sheep numbers were obtained from the 2010 Census of Agriculture [11].

The calculated STEC incidence rate was used to investigate geographic clustering within the region using ‘hotspot’ analysis and ‘cluster’ analysis. The hotspot analysis involved calculating the Getis-Ord Gi* [12] statistic for STEC incidence in each ED with the resulting z-scores indicating where EDs with either high or low values cluster spatially. For statistically significant positive/negative z-scores, the larger/smaller the z-score is, the more intense the clustering of high/low values *i.e.* a hotspot/coldspot. In order to be classified as a statistically significant hotspot, for example, an ED will have a high z-score and be surrounded by other EDs with high z-scores. Cluster analysis, on the other hand, uses Anselin’s Local Moran’s I test statistic [13]. Again the approach involves the calculation of a test statistic (*i.e.* Local Moran's I value) along with a z-score and a pseudo p-value, indicating the cluster type (or outlier) for each statistically significant ED. The cluster/outlier type distinguishes between a statistically significant cluster of high values (HH), cluster of low values (LL), outlier in which a high value is surrounded primarily by low values (HL), and outlier in which a low value is surrounded primarily by high values (LH). Statistical significance was set at 95%.

Finally, to consider the relationship between STEC incidence by ED and a range of potential predictors, correlation and multivariable regression analyses were undertaken. First, Pearson correlations between ED-level STEC incidence and the various potential predictors were estimated to give an indication of the strength of linear association between variables. Associated significance levels of each correlation coefficient were at the 95% level. In the multivariable analysis, STEC incidence rate in an ED was the dependent variable and a range of potential predictors were considered (Table 1), based on data availability and previously published research [10, 11, 14, 15].

For the multivariable regression analysis, we estimated a stepwise (backward selection) model, applying a 5% significance level for removal. This meant that the variable with the greatest p-value was removed, one at a time, until all remaining variables had a p-value less than the 5% threshold. We also tested the robustness of our findings by applying a 10% significance level for removal, as well as by estimating a number of backward stepwise models with 5% and 10% significance levels for removal and addition. Overall, our main findings were robust across these different approaches. Moreover, we also considered models that accounted for spatial dependence in both the dependent and independent variables and found this did not alter our key conclusions.

3. Results

3.1 STEC cases and incidence

The overall seven-year incidence ranged from 0 to 10.9 cases per 1000 inhabitants across the region with a mean of 0.95/1000 – equivalent to 94.57/100,000 and a per year average of 13.51/100,000. There were no cases across the whole time period in 289 of the 498 EDs (58%). In EDs *with no* cases versus EDs *with* cases, population density was lower (108 versus 182 individuals per km^2), cattle density was lower (69 versus 81 cattle per km^2), and septic tank density was lower (5.2 versus 8.5 per per km^2). There was seasonality in number of cases with a majority of 66.2% (251/379) occurring in Spring and Summer (March to August). Within this there was a lag between the two main serogroups with most O157 cases occurring earlier in the year in Spring/Summer (March to August) and most O26 cases occurring slightly later in Summer/Autumn (June to November). For O157, 80.3% (102/127) of cases occurred in March to August while 71.7% of O26 cases occurred from June to November.

3.2 Cluster Analysis

Cluster analyses using the Anselin's Local Moran's I and Getis-Ord Gi* test statistics identified a number of areas with significant local clustering of STEC incidence (Figs. 1 and 2). For example, there were a number of both HH clusters and HL outliers, most obviously in county Roscommon and east county Mayo. Moreover, there were a number of HL outliers in the west and southeast of the region. In relation to the hotspot analysis, there were 37 EDs (7.4%) that were identified as significant hotspots (at least 95% confidence) - 21 EDs at the 95% level and 16 at the 99% level. There were no ED hotspots at the 99% level in county Galway while there were four in county Mayo (57.1% of hotspots) and 12 in county Roscommon (48% of hotspots). Roscommon county had the largest proportion of its EDs as significant hotspots at 22.7% versus 4.6% and 2.1% in counties Mayo and Galway respectively.

The majority of cases were of the two serogroups O26 (n=172) and O157 (n=126), accounting for 78.6% of all cases (O157-33.2% and O26-45.4%). Getis-Ord Gi* hotspot/coldspot analysis of each serogroup indicated spatial clustering (Supplementary Figs. 1 and 2) and apparent degree of non-overlapping hotspots for each serogroup. A total of 22/24 (91.7%) of the O26 significant hotspots (95% level) were in counties Galway and Mayo while all 33 significant O157 hotspots were in counties Galway and Roscommon.

3.3 Correlation Analysis

The Pearson correlation analysis suggested a statistically significant degree of linear association between STEC rates and cattle density (positive; p=0.000) and rurality (positive; p=0.0293) (Table 2). All other variables were found to have small pairwise correlations which were not statistically significant at 5%.

3.4 Multivariable Analysis

In the final iteration of the stepwise regression model (Table 3) we found that cattle density was the only statistically significant predictor of STEC incidence in our sample, a result that was robust across a range of alternative models. No other variable was found to be significantly and robustly associated with STEC incidence in any of our multivariable regression analyses. Moreover, in addition to the model presented, we also estimated similar models separately for the two main serogroups *i.e.* O26 and O157. In both cases we also found that cattle density was the only significant and robust predictor of incidence.

4. Discussion

GIS is a flexible and adaptable methodology with a wide range of uses such as identifying and ameliorating public health risks [16][17][18] including determining geographical variation in risk and transmission of infectious diseases [19][20][21]. In our study, we were able to apply GIS methodology and multivariable analysis to identify areas at increased risk of STEC incidence in a region of Ireland and also factors that may underlie this increased risk. Since STEC incidence varies significantly by region this type of analysis may indicate specific regional risk factors which may be addressed in order to reduce incidence. We found that cattle density was the primary risk factor for STEC incidence in this region. Rurality and population density were also positively and negatively correlated with incidence respectively.

Cattle density has repeatedly been found as significantly associated with STEC incidence [22][23][24]. One German study indicated increasing risk associated with increasing cattle density [22]. This is significant since Ireland plans to increase milk and beef production over the next 10 years as part of the Food Wise 2025 strategy [25] – the national dairy herd is set to increase from 1.3 million animals in 2016 to 1.7 million by 2025 [26]. Such an increase may impact on the incidence of STEC infection in areas with a high density of livestock farming. Previous work on risk factors for STEC infection in Ireland has indicated that as well as cattle density, domestic wastewater treatment and private well usage are significantly associated with STEC infection [9]. The study region in the west of Ireland is primarily rural with a low population density. However, factors that may be indicated in increased STEC incidence risk such as domestic wastewater treatment and private well usage did not appear as significant. This may indicate that contact with the zoonotic ruminant host and its waste is the primary risk factor in this region. The west of Ireland has almost a quarter of all farms in Ireland, a lower than average farm size but less than 10% of the country's population [10, 11]. Management of cattle to prevent onward transmission of this zoonosis may prevent a significant amount of human disease. This could include herd vaccination [27]; genome sequencing can identify lineages of STEC within cattle with higher zoonotic potential intelligently guiding prevention measures such as selective herd vaccination [28].

This work has a number of limitations. The cases' geocode ED level related to residential address but it is possible the exposure may have occurred elsewhere. However, it is likely that the majority of cases will have acquired the infection in their own locale and the use of all confirmed case data over a seven-year period should reduce the effects. Also, not using data from cases within an ED within a month from a 'primary case' will reduce the effects of outbreaks and person-to-person transmission on the data which may include outbreaks originating from outside a primary case's ED.

This work has provided further evidence of the key role cattle play in the transmission of STEC to humans. Since the agricultural sector plays a particularly important role in the Irish economy and will grow in the coming years it will be important to have in place means of managing the sector from a public health standpoint also. ‘One Health’ is a concept that recognises that “*human health and animal health are interdependent and bound to the health of the ecosystems*” [29]. Our findings demonstrate this interdependence with respect to STEC and emphasizes the need to link agricultural policy and practice with impact on human health at a local area level. As incidence can vary significantly within a country/county, mapping can be used to target interventions and monitor outcomes.

5. Conclusions

STEC incidence in this region shows striking local variation and this variation is related to cattle density. GIS should be applied more generally to map local variation in zoonotic disease incidence, to target interventions and to monitor outcomes.

Acknowledgements

Thanks to the Division of Clinical Microbiology for their help particularly Belinda Hanahoe. This work was funded through an Irish Research Council Government of Ireland Postdoctoral Fellowship. Icons in graphical abstract made by Freepik (www.freepik.com) from Flaticon (www.flaticon.com). Flaticon is licensed by Creative Commons BY 3.0.

1. HPSC. Annual Epidemiological Report 2016. 2018.

2. Murphy BP, McCabe E, Murphy M, Buckley JF, Crowley D, Fanning S, et al. Longitudinal study of two Irish dairy herds: low numbers of shiga toxin-producing Escherichia coli O157 and O26 super-shedders identified. Front Microbiol. 2016;7.

3. Chase-Topping M, Gally D, Low C, Matthews L, Woolhouse M. Super-shedding and the link between human infection and livestock carriage of Escherichia coli O157. Nat Rev Microbiol. 2008;6:904–12.

4. Kintz E, Brainard J, Hooper L, Hunter P. Transmission pathways for sporadic Shiga-toxin producing E. coli infections: A systematic review and meta-analysis. Int J Hyg Environ Health. 2017;220:57–67.

5. Money P, Kelly AF, Gould SWJ, Denholm-Price J, Threlfall EJ, Fielder MD. Cattle, weather and water: Mapping Escherichia coli O157:H7 infections in humans in England and Scotland. Environ Microbiol. 2010;12:2633–44.

6. ECDC. ECDC Surveillance Atlas. ECDC. 2017. https://atlas.ecdc.europa.eu/public/. Accessed 17 April 2018.

7. Garvey P, Carroll A, McNamara E, McKeown PJ. Verotoxigenic Escherichia coli transmission in Ireland: a review of notified outbreaks, 2004–2012. Epidemiol Infect. 2016;144:917–26.

8. O'Sullivan M, Garvey P, O'Riordan M, Coughlan H, McKeown P, Brennan A, et al. Increase in VTEC cases in the south of Ireland: link to private wells? Eurosurveillance. 2008;13.

9. ÓhAiseadha C, Hynds PD, Fallon UB, O'Dwyer J. A geostatistical investigation of agricultural and infrastructural risk factors associated with primary verotoxigenic E. coli (VTEC) infection in the Republic of Ireland, 2008–2013. Epidemiol Infect. 2017;145:95–105.

10. CSO. Census 2011. 2011.

11. CSO. Census of Agriculture 2010 – Final Results. Dublin; 2012.

12. Getis A, Ord JK. The analysis of spatial association by use of distance statistics. Geogr Anal. 1992;24:189–206.

13. Anselin L. Local Indicators of Spatial Association-LISA. Geogr Anal. 1995;27:93–115.

14. Haase T, Pratschke J, Gleeson J. The 2011 All-Island HP Deprivation Index. 2014.

15. Teljeur C, Kelly A. An urban–rural classification for health services research in Ireland. Irish Geogr. 2008;41:295–311.

16. McGrory ER, Brown C, Bargary N, Williams NH, Mannix A, Zhang C, et al. Arsenic contamination of drinking water in Ireland: A spatial analysis of occurrence and potential risk. Sci Total Environ. 2017;579:1863–75.

17. Zhang C, Luo L, Xu W, Ledwith V. Use of local Moran's I and GIS to identify pollution hotspots of Pb in urban soils of Galway, Ireland. Sci Total Environ. 2008;398:212–21.

18. Stopka TJ, Krawczyk C, Gradziel P, Geraghty EM. Use of spatial epidemiology and hot spot analysis to target women eligible for prenatal women, infants, and children services. Am J Public Health. 2014;104:183–90.

19. Meyers DJ, Hood ME, Stopka TJ. HIV and hepatitis C mortality in Massachusetts, 2002-2011: Spatial cluster and trend analysis of HIV and HCV using multiple cause of death. PLoS One. 2014;9:2002–11.

20. Khormi HM, Kumar L. The importance of appropriate temporal and spatial scales for dengue fever control and management. Sci Total Environ. 2012;430:144–9.

21. Patterson B, Morrow CD, Kohls D, Deignan C, Ginsburg S, Wood R. Mapping sites of high TB transmission risk: Integrating the shared air and social behaviour of TB cases and adolescents in a South African township. Sci Total Environ. 2017;583:97–103.

22. Frank C, Kapfhammer S, Werber D, Stark K, Held L. Cattle density and shiga toxin-

producing Escherichia coli infection in Germany: Increased risk for most but not all serogroups. Vector-Borne Zoonotic Dis. 2008;8:635–44.

23. Friesema IHM, Van de Kassteele J, De Jager CM, Heuvelink AE, van Pelt W. Geographical association between livestock density and human Shiga toxin-producing Escherichia coli O157 infections. Epidemiol Infect. 2011;139:1081–7.

24. Kistemann T, Zimmer S, Vgsholm I, Andersson Y. GIS-supported investigation of human EHEC and cattle VTEC O157 infections in Sweden: Geographical distribution, spatial variation and possible risk factors. Epidemiol Infect. 2004;132:495–505.

25. DAFM. Foodwise 2025: A 10-year vision for the Irish agri-food industry. 2015.

26. Teagasc. Dairy Sector Road Map. 2016. https://www.teagasc.ie/media/website/publications/2016/Road-map-2025-Dairy.pdf.

27. Matthews L, Reeve R, Gally DL, Low JC, Woolhouse MEJ, McAteer SP, et al. Predicting the public health benefit of vaccinating cattle against Escherichia coli O157. Proc Natl Acad Sci. 2013;110:16265–70.

28. Lupolova N, Dallman TJ, Matthews L, Bono JL, Gally DL. Support vector machine applied to predict the zoonotic potential of E. coli O157 cattle isolates. Proc Natl Acad Sci. 2016;113:11312–7.

29. OIE. One Health. http://www.oie.int/en/for-the-media/onehealth/.

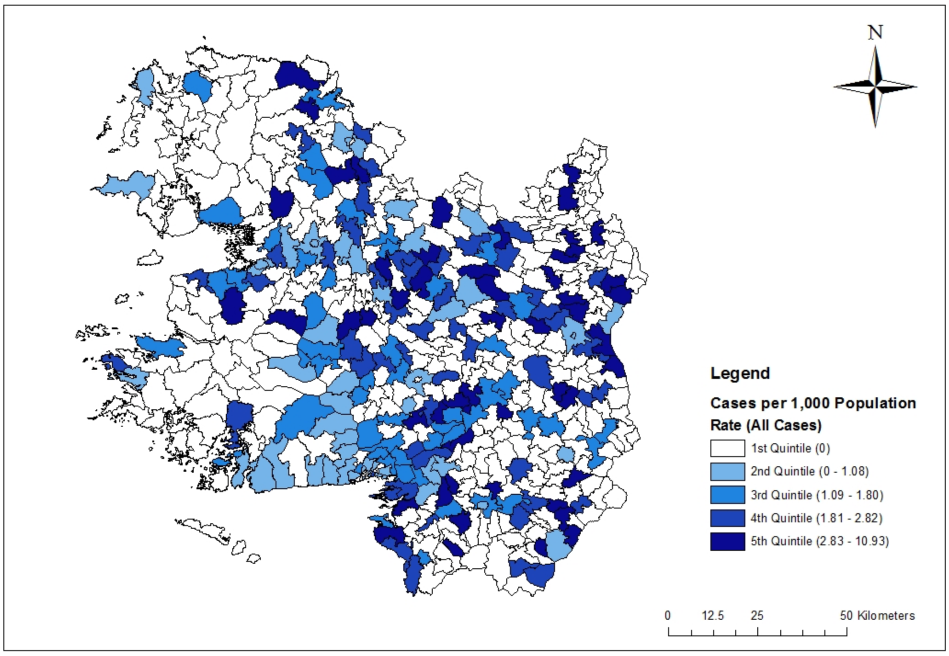

Fig. 1: Incidence per 1000 of confirmed Shiga-toxigenic *E. coli* cases in study region 2009-2015 inclusive (n=379).

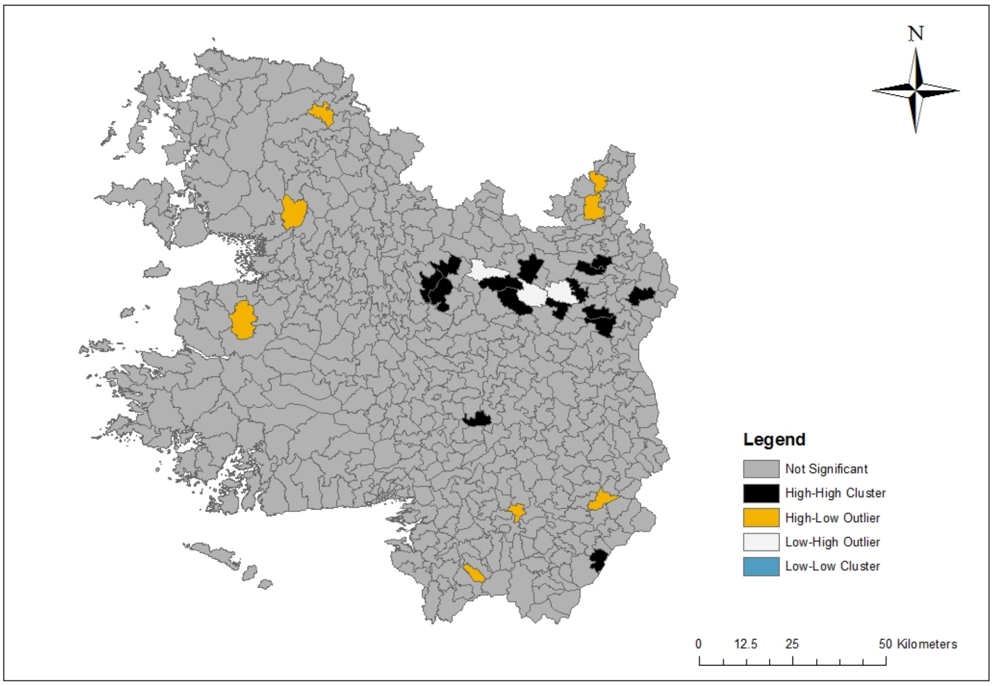

Fig. 2: Anselin's Local Moran's I cluster analysis (n=379).

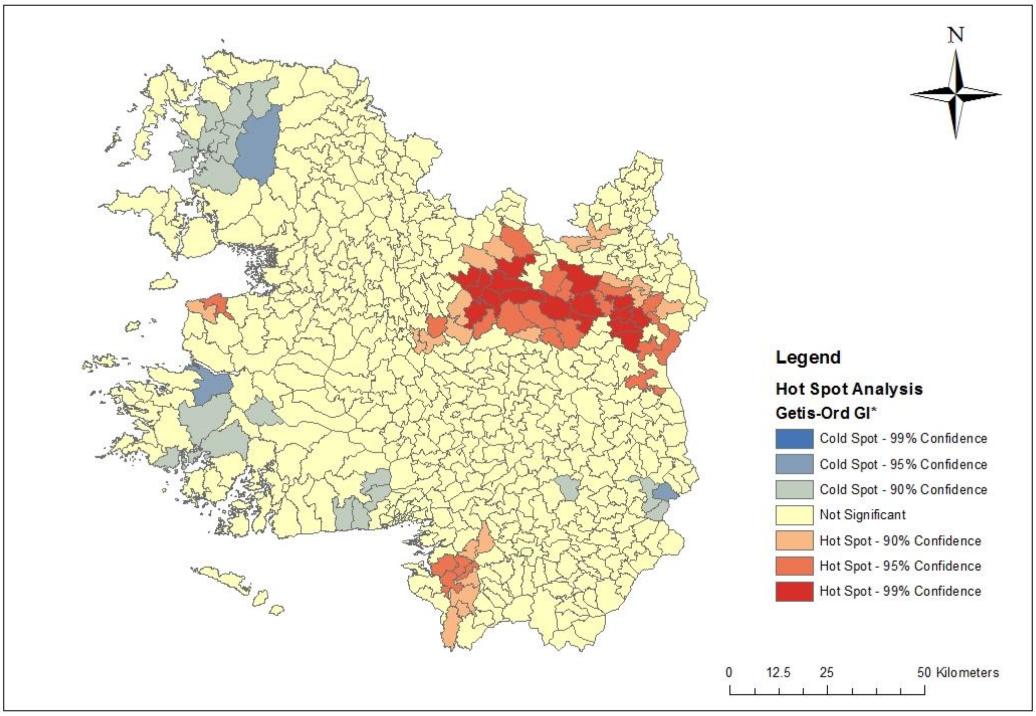

Fig. 3: Getis-Ord Gi* 'Hotspot'/'Coldspot' analysis for all cases (n=379).

Tables

Table I: Variable Definitions and Sample Descriptive Statistics

Variable Name	Variable Description	Mean (SD) or %
Dependent Variable		
STEC Rate	STEC incidence rate (cases per 1,000 population)	0.946 (1.568)
Potential Predictors		
Cattle Density	Number of cattle per km^2	74.313 (46.031)
Sheep Density	Number of sheep per km^2	90.328 (91.853)
Private Water Source	Percentage of households with a private water source	0.320 (0.261)
Septic Tank Density	Percentage of households with a septic tank	0.741 (0.262)

Deprivation	Relative deprivation (Trutz Haase Index)	-6.730 (3.450)
Population Density	Number of persons per km^2	138.829 (516.467)
Rurality	= Rural ED	90.56%
	= Urban ED	9.44%

Source: Analysis of hospital records, Census of Population 2011 data (CSO, 2011), Census of Agriculture 2010 data (CSO, 2012), data from Haase et al. (2014) (Haase et al., 2014) and Teljeur and Kelly (2008) (Teljeur and Kelly, 2008).
Note: All variables are calculated at ED level.

Table 2: Pairwise (Pearson) Correlations between STEC Incidence Rate and Potential Predictors

Variable Name	Correlation (Significance Level)
Cattle Density	0.1816 (0.0000)
Sheep Density	0.0314 (0.4849)
Private Water Source	0.0478 (0.2875)
Septic Tank Density	0.0017 (0.9697)
Deprivation Index	-0.0482 (0.2834)
Population Density	-0.0831 (0.0638)
Rurality	0.0977 (0.0293)
Observations	498

Source: Analysis of hospital records, Census of Population 2011 data (CSO, 2011), Census of Agriculture 2010 data (CSO, 2012), data from Haase et al. (2014) (Haase et al., 2014) and Teljeur and Kelly (2008) (Teljeur and Kelly, 2008).
Note: All variables are calculated at ED level.

Table 3: Regression Model of STEC Incidence Rates at ED Level

Variable	***Coefficient***	***Standard Error***	***p-Value***
Cattle Density	0.0062	0.0015	0.0000
Constant	0.4859	0.1314	0.0000
Observations	498		

Source: Analysis of hospital records, Census of Population 2011 data (CSO, 2011), Census of Agriculture 2010 data (CSO, 2012), data from Haase et al. (2014) (Haase et al., 2014) and Teljeur et al (2008) (Teljeur and Kelly, 2008).

Note: The model is a stepwise (backward selection) multivariable linear regression model, applying a 5% significance level for removal. This means that only variables found to be statistically significant at the 5% level are included in the final model. All analysis conducted at ED level.