Spatial and temporal epidemiology of sporadic human cases of *Escherichia coli* O157 in Scotland, 1996–1999

G. T. INNOCENT[1], D. J. MELLOR[1], S. A. McEWEN[2], W. J. REILLY[3],
J. SMALLWOOD[1], M. E. LOCKING[3], D. J. SHAW[4], P. MICHEL[2], D. J. TAYLOR[5],
W. B. STEELE[5], G. J. GUNN[6], H. E. TERNENT[1], M. E. J. WOOLHOUSE[4]
AND S. W. J. REID[1]* on behalf of the Wellcome Trust-funded IPRAVE Consortium

[1] *Comparative Epidemiology and Informatics, Institute of Comparative Medicine, Faculty of Veterinary Medicine, University of Glasgow, Glasgow, and Department of Statistics and Modelling Science, University of Strathclyde, Glasgow, Scotland, UK*
[2] *Department of Population Medicine, Ontario Veterinary College, University of Guelph, Ontario, Canada*
[3] *Scottish Centre for Infection and Environmental Health, Clifton House, Clifton Place, Glasgow, Scotland, UK*
[4] *Centre for Tropical Veterinary Medicine (CTVM), Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Easter Bush, Roslin, Midlothian, Scotland, UK*
[5] *Department of Veterinary Pathology, University of Glasgow Veterinary School, Glasgow, Scotland, UK*
[6] *SAC Veterinary Science Division, Drummondhill, Inverness, Scotland, UK*

SUMMARY

In Scotland, between 1995 and 2000 there were between 4 and 10 cases of illness per 100 000 population per year identified as being caused by *Escherichia coli* O157, whereas in England and Wales there were between 1 and 2 cases per 100 000 population per year. Within Scotland there is significant regional variation. A cluster of high rate areas was identified in the Northeast of Scotland and a cluster of low rate areas in central-west Scotland. Temporal trends follow a seasonal pattern whilst spatial effects appeared to be distant rather than local. The best-fit model identified a significant spatial trend with case rate increasing from West to East, and from South to North. No statistically significant spatial interaction term was found. In the models fitted, the cattle population density, the human population density, and the number of cattle per person were variously significant. The findings suggest that rural/urban exposures are important in sporadic infections.

INTRODUCTION

During the past two decades many high-profile outbreaks of human infectious intestinal disease have been caused by verocytotoxigenic *Escherichia coli* (VTEC) throughout the world. The serotypes commonly associated with such outbreaks are O157:H7 or a non-motile strain (O157:H−) [1]. *E. coli* O157 is the most frequently isolated VTEC serotype from human patients suffering gastrointestinal illness in Scotland, and rates of isolation are significantly higher in Scotland than in England and Wales [2, 3]. Sporadic cases of human disease caused by infection with *E. coli* O157, which cannot be linked to an outbreak, greatly outnumber cases that are considered to be part of a defined outbreak. In Scotland, between 1995 and 2000 there were between 4 and 10 cases

* Author for correspondence: Professor S. W. J. Reid, Comparative Epidemiology and Informatics, Institute of Comparative Medicine, Faculty of Veterinary Medicine, University of Glasgow Veterinary School, Bearsden Road, Glasgow G61 1QH, UK. (Email: stuart.reid@vet.gla.ac.uk)

of illness per 100 000 population per year identified as being caused by *E. coli* O157. This contrasts with the situation in England and Wales where there were between 1·27 and 2·08 cases per 100 000 population per year [3].

Outbreak investigations have shed important light on the foodborne aspects of the disease. Out of this work has emerged the significant role of cattle as reservoirs of *E. coli* O157, and beef consumption as a risk factor for human infection [4–6]. Considerable effort has been directed towards the elucidation of the biology of these organisms in cattle, and the identification of critical-control points in the beef-processing chain for reducing contamination of beef products with *E. coli* O157. Studies have focused on establishing the prevalence of *E. coli* O157 in cattle on the farm and in the abattoir [7–9] and identifying numbers of positive carcasses or numbers of positive sites on carcasses [10–12].

The risk factors for sporadic disease are poorly understood but it is increasingly recognized that environmental exposure is important. A case-control study on sporadic cases conducted by the Scottish Centre for Infection and Environmental Health (SCIEH), identified contact with animal faeces, other than pet faeces, as a significant positive risk factor for disease [3]. Similar studies elsewhere have demonstrated similar associations [13, 14]. *E. coli* O157 has been isolated from the faeces of most domestic animal species, but cattle have been shown to carry, and shed, the organism in greater numbers than other species [7]. Studies carried out by Wang et al. [15] have shown that *E. coli* O157 can survive in cattle faeces for up to 11 weeks at 5 °C.

In Ontario, Canada, a study conducted by Michel et al. [16] demonstrated an association between cattle population density and the rate of human infection with VTEC. Although this study did not identify specific sources of infection and modes of transmission, the authors concluded that proximity to dense cattle populations was a risk factor for disease, and proposed that various mechanisms of infection should be the subject of further investigation. Further research identified a number of indicators of livestock density positively associated with the prevalence of human disease, most notably the spreading of manure or slurry on the ground, the ratio of the cattle to human population, and a protective effect of the porcine population density [17].

In view of the different frequencies of sporadic disease in humans caused by *E. coli* O157 within Great Britain, an investigation was undertaken to identify possible associations between the cumulative incidence of sporadic disease in Scotland and a number of explanatory variables. This paper describes the spatial epidemiology of sporadic cases of *E. coli* O157 infection in humans from 1996 to 1999 in Scotland. Measures of cattle density and human density are used to describe the variation in prevalence of the human disease in space.

METHODS

Case definition

A sporadic case was defined by the SCIEH as a symptomatic case of laboratory-confirmed *E. coli* O157 in which only members of a single household were affected. Cases for which an epidemiological link was established to a case outside the original household were excluded. Cases were also excluded if the patient had travelled abroad within 2 weeks of the appearance of clinical signs. A case was included only when spatial and temporal reference data were available: the postcode district in which the case resided, and the date that a positive stool sample was identified, respectively.

Data sources

Data on reported sporadic human cases of *E. coli* O157, during the period October 1996 to April 1999, a period of enhanced surveillance, were obtained from the SCIEH. The use of postcode district to geocode the data safeguarded patient confidentiality. Postcode districts in postal areas for part of the Highlands and Islands, and Orkney and Northeast Highlands were excluded because they did not meet the inclusion criteria for the study cases and did not participate. Within the United Kingdom addresses are provided with a postcode. Postcodes possess a hierarchical structure and are used for administrative purposes. Each premise has a unit postcode, which can be grouped into postcode sectors. These in turn can be grouped into postcode districts, which can be grouped into postcode areas. There are 17 postcode areas in Scotland, and over 400 postcode districts. Postcode districts vary greatly in both size and population. The smallest postcode district is 0·69 km², whereas the largest is 3400 km², and they have a median size of 69·5 km². In general, urban districts are smaller than rural ones to account, in part, for their higher population density. The human population in postcode districts varies

from 63 to 98 628, with a median population of 6187. Human population data were based on the 1991 census and retrieved, at postcode sector resolution, from the University of Manchester [18]. These data were subsequently aggregated at postcode district resolution. Digital boundary data for postcode districts from the 1991 human census were retrieved from EDINA, the University of Edinburgh [19]. Summarized data of cattle population, again at postcode district resolution, were obtained from the Department of Environment, Farming and Rural Affairs (DEFRA) 2000 agricultural census. For analyses, using area-based data, cases and human and cattle populations were assumed to be evenly distributed within postcode district. For analyses using point patterns, cases occurring within a postcode district were spatially referenced to the postcode district centroid, as calculated by ArcView GIS (Environmental Systems Research Institute Inc., Redlands, CA, USA).

Statistical methods

Three methods were employed. First, a Poisson model was fitted to the data using a generalized linear model (GLM) with the number of cases for each postcode district as the dependent variable, the log of the human population of the district set as offset, and a number of putative independent variables were offered to the model, namely the cattle population density, the human population density, and the number of cattle per person. Trends in two orthogonal axes were considered (West–East and South–North), followed by a spatial lag model based upon the mean case cumulative incidence in neighbouring districts. The residuals from the models were then examined using Moran's I [20] in order to determine if there was any overall spatial autocorrelation. Models were developed using the full dataset and also with data from Aberdeen postcode districts excluded. This approach was adopted to investigate relationships in the study region aside from the Aberdeen area which was known to have a high prevalence and, therefore, likely to have a large influence in the models.

Second, the spatial scan statistic and the space–time scan statistic [21, 22] were used to identify possible clusters in space and time. Using the population for each postcode district and the coordinates of the centroid of each postcode district, as calculated by ArcView, a Poisson model was fitted exploring purely spatial, purely temporal and space–time clusters. SaTScan software [23] was used for the analysis.
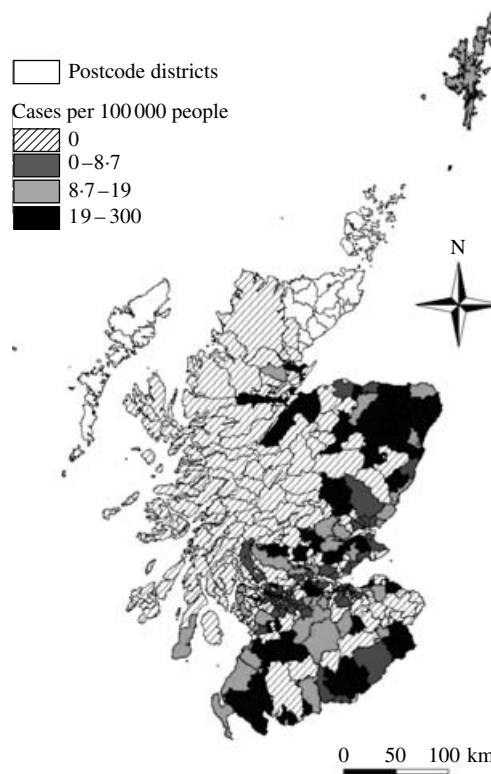

Fig. 1. Cumulative incidences of sporadic cases of *E. coli* O157 in Scotland during the period October 1996 to April 1999. Missing data are from areas that did not participate. Each different band is associated with an equal number of postcode districts, with the exception of those postcode districts with zero cases.

Third, in order to examine further the space–time relationship of sporadic cases, a second-order analysis, as described by Diggle et al. [24] extending the K function of Ripley [25], was applied. This can be interpreted as purely a spatial component, a purely temporal component and a space–time interaction component. The output presented can be considered as the attributable risk of living at specified distances and at specified times from an existing case.

The spatial models and K-function analysis were performed using SPlus (Insightful Corporation, Seattle, WA, USA) with Spatial statistics and ArcView modules. Data manipulation and visualization were carried out using ArcView GIS. For all analyses statistical significance was set at the 5% level.

RESULTS

Spatial models

A map depicting the spatial distribution of cumulative incidences in postcode districts is given in Figure 1.

Table 1. *Estimated coefficients and S.E. for the best-fit model of the spatial distribution of cases of sporadic human* E. coli *O157 infection in Scotland 1996–1999*

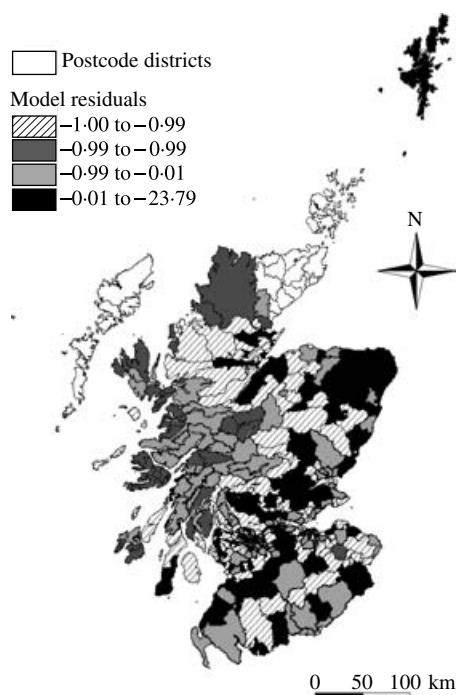|	Coefficient	S.E.	z value	P (>|z|)
(Intercept)	-13.58	0.3587	-37.846	<0.001
West to East (km)	7.212×10^{-3}	1.150×10^{-3}	6.274	<0.001
Cattle per head of human population	2.885×10^{-1}	3.730×10^{-2}	7.735	<0.001
South to North (km)	2.643×10^{-3}	5.739×10^{-4}	4.605	<0.001
Cattle population density (km²)	3.332×10^{3}	1.426×10^{3}	2.336	0.020

Fig. 2. Residuals from the best model to fit the data for the whole dataset. Note the large residuals clustered in the Northeast (Aberdeen area). Each different band is associated with an equal number of postcode districts.

The spatial models fitting cattle and human population identified a small, but significant geographic trend. There was an approximate doubling in the mean case rate per 100 km from West to East: the natural logarithm of the incidence increased by 7.2×10^{-3} per km ($P < 0.001$). Similarly there was a 30% increase in incidence per 100 km from South to North ($P < 0.001$). No interaction term was found to be statistically significant.

The lack of cases in the Northwest of the country that can be seen in Figure 1 is due to the lack of population in this area. The population in Scotland is mainly clustered around four cities: Edinburgh, Glasgow, Aberdeen and Dundee. The Northwest of the country is largely mountainous and has a very low population density. As a result, the lack of cases does not affect the model fit to any great extent.

The best-fit model included the position South–North, the position West–East of the postcode district centroid, the cattle density and the ratio of cattle to the human population (Table 1). Figure 2 shows a plot of the residuals from this model. Moran's *I* calculated on the residuals was 0.075 demonstrating negligible spatial autocorrelation. However, from Figure 2 it would appear that this model did not completely account for the number of cases present in the Aberdeen area, although there was no autocorrelation overall. The model was fitted to the reduced dataset, with Aberdeen postcode districts removed, in order to explore the effect of the Aberdeen area (Table 2). The model was similar to the model based on the full dataset, except that the coefficients for the position South–North and the position West–East approximately halved. This was as expected because a large number of cases in the Northeast of the country were removed. However, the coefficient associated with cattle density also decreased markedly, and was not significant. The Northeast of the country is characterized by having a higher ratio of cattle to humans, a higher cattle density and similar human population density when compared to the whole of Scotland (Table 3).

The subsequent best model fitted to the reduced dataset is given in Table 4. This model was forced to include a coefficient for both distance South–North and West–East, even though the latter was not significant. This model was then fitted to the whole dataset for comparison (Table 5). In this case human population density was significant, with an increase in population density being associated with a decrease in the cumulative incidence rate. The coefficients for position South–North and East–West doubled with the inclusion of Aberdeen postcodes.

Table 2. *Estimated coefficients and S.E. for the model with the same independent variables as the best model for the whole dataset applied to the reduced dataset, i.e. with Northeastern postcode districts removed, of the spatial distribution of cases of sporadic human* E. coli *O157 infection in Scotland 1996–1999*

Coefficients	Estimate	S.E.	z value	$P (>|z|)$
(Intercept)	$-1 \cdot 186 \times 10^1$	$5 \cdot 373 \times 10^{-1}$	$-22 \cdot 064$	$<0 \cdot 001$
West to East (km)	$3 \cdot 689 \times 10^{-3}$	$1 \cdot 356 \times 10^{-3}$	$2 \cdot 720$	$0 \cdot 007$
Cattle per head of human population	$2 \cdot 504 \times 10^{-1}$	$4 \cdot 486 \times 10^{-2}$	$5 \cdot 582$	$<0 \cdot 001$
South to North (km)	$1 \cdot 641 \times 10^{-3}$	$7 \cdot 562 \times 10^{-4}$	$2 \cdot 169$	$0 \cdot 030$
Cattle population density (km²)	$1 \cdot 689 \times 10^3$	$1 \cdot 544 \times 10^3$	$1 \cdot 094$	$0 \cdot 274$

Table 3. *Comparison of Northeast area of Scotland with the whole of Scotland with respect to parameters of interest for spatial models*

Area	Minimum	Q1	Median	Q3	Maximum
Cattle per head of human population					
Northeast	$0 \cdot 020$	$0 \cdot 36$	$0 \cdot 86$	$1 \cdot 9$	$4 \cdot 1$
Scotland	$0 \cdot 0$	$0 \cdot 034$	$0 \cdot 34$	$1 \cdot 1$	13
Cattle population density					
Northeast	$0 \cdot 66$	21	34	47	7100
Scotland	$0 \cdot 0$	$2 \cdot 0$	19	42	220
Human population density (km²)					
Northeast	$1 \cdot 5$	17	40	95	2100
Scotland	$0 \cdot 000034$	$9 \cdot 6$	56	490	6500

Q1, first quartile; Q3, the third quartile. All data given to two significant figures.

Table 4. *Estimated coefficients and S.E. for the best-fit model of the spatial distribution of cases of sporadic human* E. coli *O157 infection in Scotland 1996–1999, using the reduced dataset, i.e. with Northeastern postcode districts removed*

Coefficients	Estimate	S.E.	z value	$P (>|z|)$
(Intercept)	$-1 \cdot 127 \times 10^1$	$5 \cdot 556 \times 10^{-1}$	$-20 \cdot 290$	$<0 \cdot 001$
Cattle per head of human population	$2 \cdot 260 \times 10^{-1}$	$4 \cdot 787 \times 10^{-2}$	$4 \cdot 720$	$<0 \cdot 001$
West to East (km)	$3 \cdot 251 \times 10^{-3}$	$1 \cdot 327 \times 10^{-3}$	$2 \cdot 451$	$0 \cdot 014$
Human population density (km²)	$-1 \cdot 025 \times 10^2$	$4 \cdot 859 \times 10^1$	$-2 \cdot 109$	$0 \cdot 035$
South to North (km)	$1 \cdot 234 \times 10^{-3}$	$7 \cdot 714 \times 10^{-4}$	$1 \cdot 600$	$0 \cdot 110$

Space-scan statistic

The most likely spatial cluster identified was located in the Aberdeen area, with 79 cases, rather than the 12 expected ($P<0 \cdot 001$). The second most likely spatial cluster identified was a cluster of low-prevalence areas, located in the Glasgow area, with 102 cases rather than the 184 expected ($P<0 \cdot 001$). A single temporal cluster, from April to September 1997 was identified, with 166 cases rather than the 86 expected ($P<0 \cdot 001$).

Two space–time clusters were identified. The first was geographically identical to the purely spatial cluster identified in the Aberdeen area, but was

Table 5. *Estimated coefficients and s.e. for the model with the same independent variables as the best model for the reduced dataset applied to the whole dataset of the spatial distribution of cases of sporadic human E. coli O157 infection in Scotland 1996–1999*

Coefficients	Estimate	S.E.	z value	P (>|z|)
(Intercept)	-1.295×10^{1}	3.715×10^{-1}	-34.861	<0.001
Cattle per head of human population	2.852×10^{-1}	3.833×10^{-2}	7.440	<0.001
West to East (m)	6.559×10^{-3}	1.119×10^{-3}	5.864	<0.001
Human population density	-9.276×10^{1}	4.685×10^{1}	-1.980	0.048
South to North (m)	2.328×10^{-3}	5.794×10^{-4}	4.018	<0.001
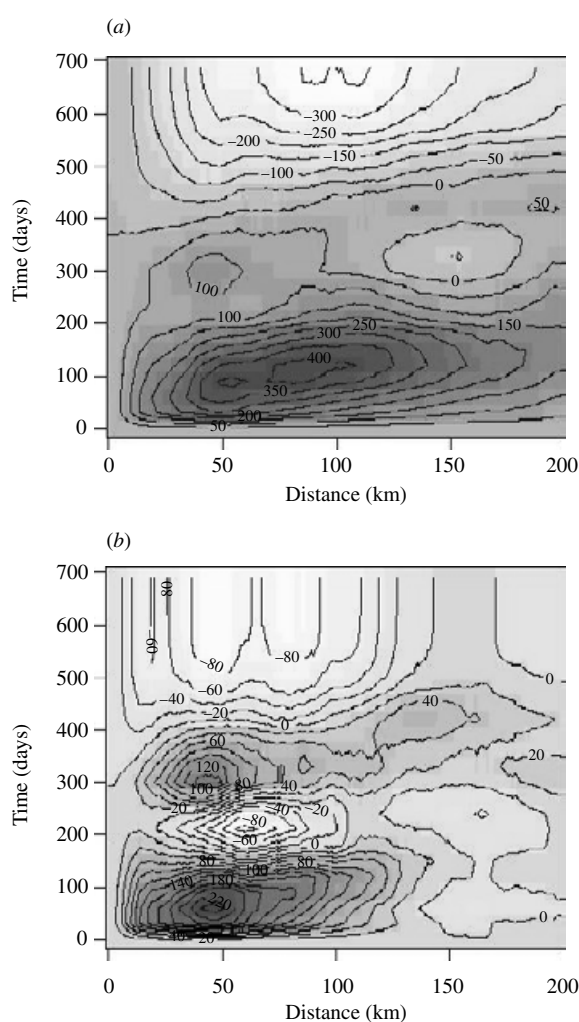
Fig. 3. Graduated shading and contour plot of the *D* value at various distances and times from a sporadic case of *E. coli* O157. The darkest areas represents the highest *D* value, the lightest areas the lowest *D* value. The *D* value is similar to the attributable risk and represents the increase in probability of the presence of a second case due to being within a certain distance and time of an index case. (*a*) Whole dataset. (*b*) Dataset with cases in the Aberdeen area removed.

identifiable from July 1997 to September 1998, with 61 rather than the six expected cases ($P<0.001$). The second included the purely spatial cluster identified in the Glasgow area, but was more extensive, including much of the West of Scotland. This cluster lasted from November 1997 to January 1999, and had 36, rather than the 107 cases expected ($P<0.001$).

K-function analysis

Figure 3*a* is a contour plot of the *D* values, analogous to risk difference, or attributable risk. There was an increase in *D* value up to a distance of around 100 km with a decrease at further distances and some evidence of an increase in *D* value at a distance of 50 km and around 300 days. Removing the Aberdeen data (Fig. 3*b*), an increase in *D* value was observed from distances up to around 50 km, and a subsequent decrease in *D* value as distance increased to around 100 km. The increase in *D* value extended in the time dimension to about 100 days. There was a second rise in *D* value at about 300 days and 50 km.

DISCUSSION

This study describes significant variation in the spatial and temporal distribution of sporadic cases of *E. coli* O157 infections of humans in Scotland that are statistically associated with human and animal population characteristics. By applying a range of spatial and temporal statistical techniques we have identified geographical and temporal patterns. It is worthy of note that where different techniques were assessing the same or similar variables, the results and inference were broadly the same.

First, considering purely spatial aspects, two clusters were identified; one of greater than expected

prevalence and one of lower than expected prevalence. This finding, from the application of the spatial-scan statistic, was supported by the GLM approach. Spatial patterns can be present in two forms. In the first there is some spatial trend across the data, whilst, in the second, local effects produce local regions of high or low prevalence. It is important that any global spatial trend be accounted for before local effects are determined because the former may mask the latter [20]. This is particularly important in a disease such as that caused by human *E. coli* O157 infection, where the geographic origin of the infection may not be local to the position at which the case is reported. In order to determine whether or not trends in infection occurred in either the South–North direction or the West–East direction, the position of the postcode district centroid was used in a Poisson model. Any trend in a direction at an angle to these axes would be identified as a trend along these axes, possibly with an interaction term.

The analyses presented here demonstrate a significant trend from lower prevalence in the Southwest to higher prevalence in the Northeast. There was no demonstrable spatial autocorrelation of numbers of cases in postcode districts once these trends were accounted for, although there did appear to be evidence of high numbers of cases in the Aberdeen area, and low numbers of cases around Glasgow. This may represent a true trend, or may represent the model that best fits the two clusters, which in themselves represent a South–West, North–East axis. However, given that the introduction of an interaction term was not significant, it suggests that the two clusters were not the only areas contributing to this effect. Furthermore, this trend remained when Northeastern postcodes were removed from the model.

The two spatio-temporal clusters identified by the space–time scan statistic, represent a cluster of increased numbers of cases in the Northeast, and a cluster of fewer than expected cases in the Southwest. However, one must be careful in interpreting these results. The presence of these spatio-temporal clusters may have little or no biological significance. However, the *K*-function analysis supports the existence of spatio-temporal clustering. The two peaks in the *D* value may represent an increase in risk associated with time of year, and the increase in spatial association might be expected due to the presence of conurbations and the accompanying increased population density. There was some indication that there was a more extensive temporal cluster and it would be

interesting to investigate the possibility of a meta-population epidemic, i.e. one that crosses several population groups such as towns, although more extensive time series would be required to explore this possibility.

The development of models that included other putative explanatory variables must be interpreted with care. Whilst it is tempting to consider the human and cattle population variables as risk factors, it is important to note, first, that the number of explanatory variables and factors offered to the models was small and, second, these measures are crude estimates of biological variables and may be surrogate or proxy measures for a combination of unmeasured or unknown variables. However, the models displayed a high degree of similarity regardless of whether the Northeastern postcode districts were included. In particular, in all models, an increase in the numbers of cattle per person was accompanied by an increase in the prevalence of sporadic cases.

The parameters that best modelled the data in the absence of Northeastern postcode districts, although similar to those when they were present, included the human population density as a protective factor, rather than the cattle population density as a risk factor. The addition of the Northeastern postcode districts altered this model less than removing the Northeastern postcodes from the model that was the best fit for the whole dataset. This suggests that the model including human population density as a protective factor is a more robust model for the occurrence of sporadic *E. coli* O157. Importantly, the fact that the trend persisted in the absence of the Northeastern postcode districts suggests that the possible detection bias due to the national reference laboratory for *E. coli* O157 being located in Aberdeen at the time of the enhanced surveillance may not be a significant issue.

However, there are several considerations that should be taken into account in the interpretation of these results. Michel et al. [16] were concerned about the potential diluting effect that metropolitan centres might have on the case rate at Canadian county level. They amalgamated metropolitan areas, defined as having >1500 people per km² and a total population of >100 000, and analysed them as separate areas, thus reducing the heterogeneity of their data at county level. In the present study, none of the districts met both of these criteria. It was assumed that a high degree of data homogeneity was achieved by the use of postcode districts as the spatial unit, which tend to

delineate areas of similar human population density. However, the use of these relatively small spatial units may result in other errors, particularly with regard to the location of human cases of *E. coli* O157. Cases are allocated to the district in which the patient resides, which is not necessarily the district in which infection occurred. The relatively small size of many districts, particularly those close to urban centres, may result in cases being recorded some distance from the point of infection. In large spatial units, such as the counties used by Michel et al. [16], the likelihood of cases being recorded in the same location as infection occurred is greater, although there is likely to be less homogeneity of exposure over the whole area. Edge effects were not accounted for and there is a possibility that a farm located in one postcode district might have cattle in another, neighbouring or otherwise. Furthermore, the spatial models that were used in this study consider only straight-line distances. This shortcoming might have significant effects in coastal or mountainous regions where geographical features might create biological boundaries.

Discrepancies between the times at which the different datasets were collected, ranging from 1991 for human population data to 2000 for cattle population data, were also considered to be a potential source of error. Examination of cattle population figures from the Scottish Environment and Rural Affairs Department (SERAD) covering this period showed that between 1991 and 1999 the total cattle population in Scotland had decreased by 3·7% from approximately 2 107 870 to 2 029 580. However, the observed decrease in cattle population was not constant; rather it has fluctuated throughout the decade. The elementary calculation of cattle population density at the postcode district level is in itself a serious oversimplification of the true distribution of cattle throughout the area, nor does it take into account age structure, management types or seasonal management of herds, for example calving patterns, or the effect that these factors may have on seasonal distribution of shedding *E. coli* O157. Cattle may be grazed or indeed housed some distance from the identified farm of origin, possibly in neighbouring postcode districts. A further source of error is related to the time at which census returns are completed. The census represents a 'snapshot' of the true farm population, which is varying in time, dependent upon sales patterns and management practices and it is clear from preliminary movement studies that the population in Scotland is dynamic.

It is unclear how far these spatial and temporal caveats impact on the anomaly of the highest prevalence of shedding in cattle occurring in winter and spring [8] and yet the highest prevalence of human cases occurs in summer. If the premise that cattle are the major source of *E. coli* O157 is accepted and the trend for the majority of outbreaks being environmental holds true for sporadic human infections, then the direct link between cattle and humans requires closer scrutiny. Certainly, the cattle and human population densities could be argued as measures for rural–urban exposure. In this biological model, supported by Valcour et al. [17], there is a need for greater focus on the behavioural attributes of humans and in particular on how individuals interact with environment in which *E. coli* O157 might occur. Such exposures, whether by direct contact with animal faeces [14] or through contaminated water [17], are plausible. The role of other animal species, especially sheep, remains untested in our models.

In conclusion, significant clusters of sporadic *E. coli* O157 infections in humans in Scotland do exist and the ratio of cattle population to human population would appear to be a small but significant risk factor in determining the likelihood of *E. coli* O157-related disease in people. Whether this is a direct effect or a proxy measure of some other cause within the rural environment will require further research. The presented techniques allow the association of both easily identified and hidden patterns within the data to environmental markers of interest. This allows a comprehensive model of parameters that are believed *a priori* to be sufficiently important to be investigated. Using these methods it is intended to extend the investigation to include such other possible indicators of exposure risk, where these are available. An important addition to future studies would be the inclusion of district level prevalence of cattle shedding of *E. coli* O157.

REFERENCES

1. **Coia JE, Sharp JCM, Campbell DM, Curnow J, Ramsay CN.** Environmental risk factors for sporadic *Escherichia coli* O157 infection in Scotland: results of a descriptive epidemiology study. J Infect 1998; **36**: 317–321.

2. **Chapman PA.** Verocytotoxin-producing *Escherichia coli*: an overview with emphasis on the epidemiology and prospects for control of *E. coli* O157; *Escherichia coli* as an intestinal pathogen. Food Control 1995; **6**: 187–193.

3. **Reilly WJ.** Task force on *E. coli* O157 final report. Glasgow: Scottish Centre for Infection and Environmental Health, 2001.

4. **Smith RA, Griffin DD, Dargatz DA.** The risks and prevention of contamination of beef feedlot cattle: the perspective of the United States of America. Rev Sci Tech 1997; **16**: 359–368.

5. **Hancock DD, Besser TE, Rice DH, Ebel ED, Herriot DE, Carpenter LV.** Multiple sources of *Escherichia coli* O157 in feedlots and dairy farms in the Northwestern USA. Prev Vet Med 1998; **35**: 11–19.

6. **Van Donkersgoed J, Graham T, Gannon V.** The prevalence of verotoxins, *Escherichia coli* O157:H7, and Salmonella in the feces and rumen of cattle at processing. Can Vet J 1999; **40**: 332–338.

7. **Chapman PA, Siddons CA, Cerdan Malo AT, Harkin MA.** A 1-year study of *Escherichia coli* O157 in cattle, sheep, pigs and poultry. Epidemiol Infect 1997; **119**: 245–250.

8. **Synge B, Gunn GJ, Ternent HE, et al.** Preliminary results from epidemiological studies in cattle in Scotland. Zoonotic infections in livestock and the risk to public health, VTEC O157. Edinburgh: Royal College of Physicians, 28 June 2000.

9. **Paiba GA, Gibbens JC, Pascoe SJS, et al.** Faecal carriage of verocytotoxin-producing *Escherichia coli* O157 in cattle and sheep at slaughter in Great Britain. Vet Rec 2002; **150**: 593–598.

10. **Bell RG.** Distribution and sources of microbial contamination on beef carcasses. J Appl Microbiol 1997; **82**: 292–300.

11. **Elder RO, Keen JE, Siragusa GR, Barkocy-Gallagher GA, Koohmaraie M, Laegreid WW.** Correlation of enterohemorrhagic *Escherichia coli* O157 prevalence in feces, hides, and carcasses of beef cattle during processing. Proc Natl Acad Sci USA 2000; **97**: 2999–3003.

12. **McEvoy JM, Doherty AM, Finnerty M, et al.** The relationship between hide cleanliness and bacterial numbers on beef carcasses at a commercial abattoir. Lett Appl Microbiol 2000; **30**: 390–395.

13. **Parry SM, Salmon RL, Willshaw GA, Cheasty T.** Risk factors for and prevention of sporadic infections with vero cytoxin (shiga toxin) producing *Escerichia coli* O157. Lancet 1998; **351**: 1019–1022.

14. **Trevena WB, Willshaw GA, Cheasty T, Domingue G, Wray C.** Transmission of vero cytotoxin producing *Escherichia coli* O157 infection from farm animals to humans in Cornwall and West Devon. Commun Dis Public Health 1999; **2**: 263–268.

15. **Wang G, Zhao T, Doyle MP.** Fate of Enterohemorrhagic *Escherichia coli* O157:H7 in bovine faeces. Appl Environ Microbiol 2001; **62**: 2567–2570.

16. **Michel P, Wilson JB, Martin SW, Clarke RC, McEwen SA, Gyles CL.** Temporal and geographical distributions of reported cases of *Escherichia coli* O157:H7 infection in Ontario. Epidemiol Infect 1999; **122**: 193–200.

17. **Valcour JE, Michel P, McEwen SA, Wilson JB.** Associations between indicators of livestock farming intensity and incidence of human shiga toxin-producing *Escherichia coli* infection. Emerg Infect Dis 2002; **8**: 252–257.

18. **Census Dissemination Unit** (http://census.ac.uk/cdu/). Manchester Information & Associated Services (MIMAS). University of Manchester. Accessed 14 Sept. 2001.

19. **UK Borders** (http://edina.ac.uk/ukborders/). Data Library Services for the University of Edinburgh (EDINA). Accessed 14 Sept. 2001.

20. **Bailey TC, Gatterell AC.** Interactive spatial data analysis. Harlow: Longman, 1995.

21. **Kulldorff M, Nagarwalla N.** Spatial disease clusters: detection and inference. Stat Med 1995; **14**: 799–810.

22. **Kulldorff M, Rand K, Gherman G, Williams KG, DeFrancesco D.** SaTScan, version 2.1, software for calculating the spatial, temporal and space-time scan statistics. National Cancer Institute, 1998.

23. **National Cancer Institute** (http://www3.cancer.gov/prevention/bb/satscan.html). Biometry Research Group. Accessed 9 March 2004.

24. **Diggle PJ, Chetwynd AG, Haggkvist R, Morris SE.** Second-order analysis of space-time clustering. Stat Methods Med Res 1995; **4**: 124–136.

25. **Ripley BD.** The second-order analysis of stationary point processes. J Appl Prob 1976; **13**: 255–266.