

Robustness of Statistical Models for *Escherichia coli* O157 in Humans in Scotland

Innocent, G.T.¹, Mellor, D.J.¹, Locking, M.E.², Reilly, W.J.², Ternent, H.E.¹, McEwen, S.A.³, and Reid, S.W.J.¹

¹Institute of Comparative Medicine, Faculty of Veterinary Medicine, University of Glasgow, Bearsden Road, Glasgow G61 1QH, UK; ²Health Protection Scotland, Clifton House, Clifton Place, Glasgow, G3 7LN, UK;

³Department of Population Medicine, Ontario Veterinary College, University of Guelph, Ontario, Canada.

Abstract

Data were available for sporadic human cases of *Escherichia coli* (*E. coli*) O157 for the periods 1996-1999 and 1999-2003. Generalised linear models were fitted to both data sets, using a Poisson model. Spatial position of cases, human population density, cattle population density and the ratio of cattle to human population, were used as explanatory variables. In both data sets period per capita case rates increased in both a northerly and easterly direction, and increased with increase in ratio of cattle to human population. However, the coefficients for the resulting models were so similar that the veracity of the analysis warranted investigation. In particular whether the known high incidence in the north-east, the low incidence in the south-west area, and the high cattle density in the north-east were driving the model fit. Models were therefore refitted split by human population density. These models also had similar coefficients. The authors believe, therefore, that this represents a true model of the data, and the fit is not being forced by a single area, or by any subgroup of cases.

Introduction

Cattle have been identified as important reservoirs of *E. coli* O157, and beef consumption has been recognised as a risk factor for human infection (Gannon, 1999). The risk factors for sporadic disease are poorly understood but it is increasingly recognised that environmental exposure is important (SEHD/FSA(S) 2001). A case control study on sporadic cases conducted by the Scottish Centre for Infection and Environmental Health (SCIEH, now Health Protection Scotland, HPS), identified contact with animal faeces, other than pet faeces, as a significant positive risk factor for disease (Locking and others, 2001). Similar studies elsewhere have demonstrated similar associations (e.g. Trevena and others 1999).

Studies have shown an association between cattle density or farming intensity and risk of human disease (Michel and others 1999, Valcour and others 2002, Innocent and others 2005). However, it is difficult to distinguish between “rurality” and indicators of cattle density. Furthermore much of the spatial aspect of such models may be driven by the spatial distribution of rural areas within the regions considered.

Objective

To investigate the robustness of spatial models for human cases of *E. coli* O157 in Scotland by repeating analyses using subsets of the data and comparing the results of these analyses to the results using the full data set.

Methods

Data on reported human cases of *E. coli* O157, during the periods 1996-1999 and 1999-2003 were obtained from HPS. Postcode sector of residence was obtained by the HPS enhanced surveillance system (Locking and others, 2005) which contains demographic and epidemiological follow-up details for all confirmed cases. Cases linked to general outbreaks (i.e. incidents involving members of more than one household) are differentiated from sporadic cases. Postcode sector was used to

geocode the data in order that patient confidentiality would be maintained. 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 around 900 postcode sectors in Scotland.

Human population and digital boundary data were based on the 2001 census and retrieved, at post-code sector resolution, from EDINA (<http://edina.ac.uk/>). Summarized data of cattle population, again at postcode sector resolution, were obtained from the Department of Environment, Farming and Rural Affairs (DEFRA) 2000 agricultural census. Cases and human and cattle populations occurring within a postcode district were spatially referenced to the postcode district centroid.

A Poisson model was fitted to the data using a generalized linear model (GLM) with the number of cases for each post code 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, viz., the cattle population density, the human population density, and the number of cattle per person. Trends in two orthogonal axes were also considered (West-East and South-North).

In order to test the robustness of the model the second data set, that consisting of cases from 1999-2003, was split into “rural” and “non-rural” areas to compare models. Since no standard definition of rurality exists, it was decided to base rurality on human population density. Two approaches were considered.

In the first the post code sectors were split into five more-or-less equal-sized groups (pentiles) by human population density. Each data set represented only one fifth of the whole data set, and were therefore considered unlikely to give a stable model. Therefore, the model of best fit to the whole data set was fitted to each of the pentiles in turn, and the coefficients examined to determine by how much they changed.

In the second analysis only rural areas were considered. However the definition of rural was altered to include only the first pentile, the first two pentiles, the first three, four and finally all five pentiles, respectively. In these analyses the data sets were sufficiently large to produce reasonable models by themselves, and so in this approach the best fit model was fitted, and coefficients considered.

Results

The results from the two best fit models to the two data sets are given in Table 1. As can be seen the models were similar, and the coefficients were in close agreement.

Table1: Coefficients of best fitting model to human sporadic case data 1996-1999 and 1999-2003.

<i>Coefficient</i>	<i>1996-1999</i>	<i>1999-2003</i>
Intercept	-14	-13
South-North (Km)	7.2×10^{-3}	7.0×10^{-3}
West-East (Km)	2.6×10^{-3}	2.5×10^{-3}
Cattle population density (Km ⁻²)	3.3×10^3	NA
Cattle:human population	2.9×10^{-1}	2.6×10^{-1}

When we considered the models fit to each pentile of human density the results in Table 2 were obtained. The coefficients for each model were similar, particularly for those coefficients that were significant. It should be noted that each pentile has only one fifth of all areas available for consideration, and so the power of the analysis is reduced.

Table 2: Coefficients from the same model fitted to each of the five pentiles of human population density. Pentile 1 is the least dense fifth, pentile 5 the most densely populated post code sectors. Numbers in bold type are significant in the model at the 5% level.

<i>Coefficient</i>	<i>Pentile 1</i>	<i>Pentile 2</i>	<i>Pentile 3</i>	<i>Pentile 4</i>	<i>Pentile 5</i>
Intercept	-12	-10	-10	-20	-13
Cattle:human population	1.4x10⁻¹	2.2x10⁻¹	-8.3x10 ⁻¹	6.6	-11
West-East (Km)	1.1x10⁻²	3.2x10⁻³	-2.4x10 ⁻³	9.5x10⁻³	3.5x10 ⁻³
South-North (Km)	1.1x10⁻³	1.2x10 ⁻³	4.1x10⁻³	1.1x10⁻²	4.6x10 ⁻³

Similarly, fitting the best fit model to the least populated fifth, two fifths, three fifths etc. the results in Table 3 were obtained. Once more this results were similar across all analyses. Although the models varied there was consistency in that there was an increase from West to East, from South to North and an increase in “rural” areas.

Table 3: Coefficients from the best fit model to rural areas only. Rural areas were defined as the 20%, 40%, 60% 80% or 100% of postcode sectors with the lowest human population density.

<i>Coefficient</i>	<i>20%</i>	<i>40%</i>	<i>60%</i>	<i>80%</i>	<i>100%</i>
Intercept	-12	-11	-11	-12	-13
West-East (Km)	9.5x10 ⁻³	7.8x10 ⁻³	6.4x10 ⁻³	7.3x10 ⁻³	7.0x10 ⁻³
South-North (Km)	1.6x10 ⁻³	8.5x10 ⁻⁴	1.4x10 ⁻³	2.0x10 ⁻³	2.5x10 ⁻³
Human population density	-1.3x10 ⁻²	-6.8x10 ⁻³	-1.2x10 ⁻³	-2.7x10 ⁻⁴	
Cattle population density	1.5x10 ⁻²	3.1x10 ⁻³			
Cattle:human population		1.1x10 ⁻¹	1.8x10 ⁻¹	2.0x10 ⁻¹	2.6x10 ⁻¹

Discussion

The similarity of all models presented indicates that the models are fairly robust. It is interesting to note that the model from the middle pentile gave very different coefficient estimates to all other models, although they were not significant. Furthermore, the most densely populated pentile indicated that cattle population density was protective, although this was again not significant. This is almost certainly due to there being very few postcode sectors in this pentile with cattle present.

Overall these models indicate that there is a trend both West to East and South to North in the period prevalence of human *E. coli* O157 sporadic cases. Furthermore, particularly in areas of low human population density, the human disease is more common in areas with high numbers of cattle present. However we would caution against making the assumption of causality. Areas that are ideal for cattle farms may, in turn, have other attributes that make them a high risk for human disease.

References

Innocent, GT, Mellor, DJ, McEwen, SA, Reilly, WJ, Smallwood, J, Locking, ME, Shaw, DJ, Michel, P, Taylor, DJ, Steele, WB, Gunn, GJ, Ternent, HE, Woolhouse, MEJ, Reid, SWJ. (2005) Spatial and temporal epidemiology of sporadic human cases of *Escherichia coli* O157 in Scotland, 1996-1999. *Epidemiol Infect*; 133(6): 1033-1041.

Locking ME, O'Brien SJ, Reilly WJ, Wright EM, Campbell DM, Coia JE et al (2001). Risk factors for sporadic cases of *Escherichia coli* O157 infection: the importance of contact with animal excreta. *Epidemiol Infect*; 127: 215-220.

Locking M, Allison L, Rae L, Pollock K, Hanson M (2005). VTEC in Scotland 2004. *Health Protection Scotland Weekly Report*; 39 (51-52): 290-295.
<http://www.show.scot.nhs.uk/scieh/PDF/pdf2005/0551.pdf>

Michel P, Wilson JB, Martin SW, Clarke RC, McEwen SA, Gyles CL. (1999) Temporal and geographical distributions of reported cases of *Escherichia coli* O157:H7 infection in Ontario. *Epidemiol Infect*; 122(2): 193-200.

Scottish Executive Health Department/Food Standards Agency (Scotland) (2001). Report of the E.coli O157 Task Force. Edinburgh: The Stationery Office, July 2001.
<http://www.food.gov.uk/news/newsarchive/2001/oct/ecolitask>

Trevena WB, Willshaw GA, Cheasty T, Domingue G, Wray C. (1999) Transmission of Vero cytotoxin producing *Escherichia coli* O157 infection from farm animals to humans in Cornwall and West Devon. *Commun Dis Public Health*; 2(4): 263-268.

Valcour JE, Michel P, McEwen SA, Wilson JB. (2002) Associations between Indicators of Livestock Farming Intensity and Incidence of Human Shiga Toxin-Producing *Escherichia coli* Infection. *Emerg Infect Dis*; 8(3): 252-257.

Acknowledgements

This work was funded by the Wellcome Trust under the International Partnership Research Awards in Veterinary Epidemiology (IPRAVE) initiative, and the authors are grateful for the contributions, assistance and advice of the other members of the IPRAVE consortium. Patients, their relatives, public health teams and other contributors to HPS enhanced surveillance are also gratefully acknowledged.