REPORT TO FSA SCOTLAND ON

PROJECT S14004

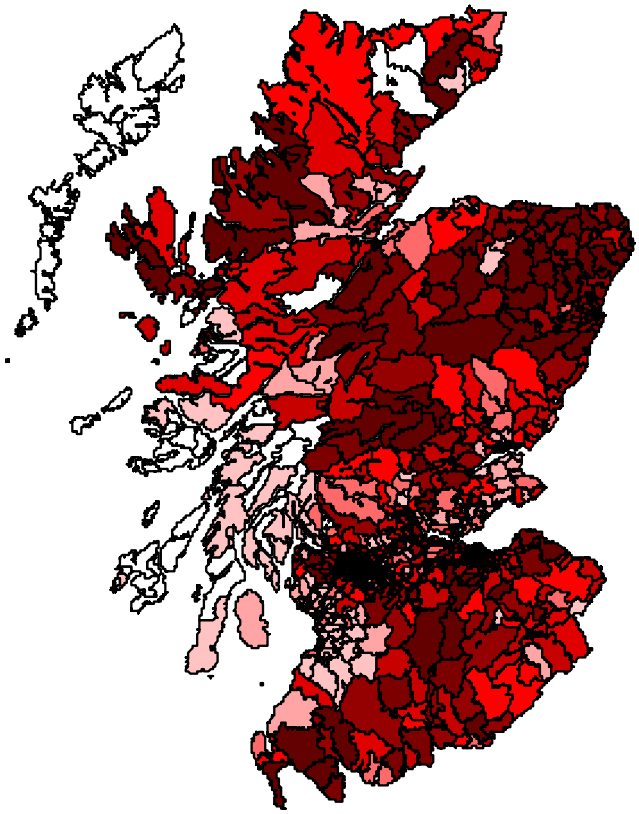
FACTORS ASSOCIATED WITH GEOGRAPHIC AND TEMPORAL VARIATION IN CAMPYLOBACTERIOSIS IN HUMANS

Contractors:

University of Aberdeen

University of Glasgow

Health Protection Scotland



Food Standards Agency
food.gov.uk
Buidheann Inbhe-Bìdhe

2. SUMMARY

Campylobacter spp. is the largest cause of bacterial gastrointestinal infection in the developed world (Blaser, 1997). In Scotland, where this study was based, reported infection rates increased in the 1990's peaking at 121 cases per 100,000 population in 2000. Subsequently they fell to 98 per 100,000 by 2006, but had risen again to 123/100,000 in 2009 (HPS, 2010). It therefore remains a substantial cause of illness. The FSA strategic plan for 2006-2010 included the goal of achieving further reductions in foodborne disease, beyond the 20% reduction achieved in 2006. However, the sustained rates of *Campylobacter* infection suggests that reduction in incidence requires an improved understanding of its epidemiology.

An important feature of the occurrence of *Campylobacter* infection is the number of reported cases per hundred thousand people varies considerably from region to region, in particular with some health boards observed to report more cases than others. This project used a combination of information on the distribution of potential risk factors for *Campylobacter* infection with information on the location of cases, to improve our understanding of the major risks for the acquisition and reporting of human campylobacteriosis. This project also exploited strain-typing data from human cases and from animal reservoirs, providing the opportunity to both identify and determine the relative importance of the major routes by which people become infected.

This project used data collected from human cases between 2000 and 2006, giving the age, sex, timing and location of cases (health board and postcode sector). In combination with data on the distribution of potential risk factors – private water supplies, animal densities, and measures of deprivation – these data were used to develop statistical models that captured the geographic distribution of cases and identified the major risk factors for infection. Mapping tools were also used to examine the spatial distribution of cases to identify clusters, both with and without strain typing.

This study found that there are real differences in the geographic distribution of *Campylobacter* infections within Scotland caused by differences in exposure to infection. Deprivation was found to be a protective factor, with higher rates of *Campylobacter* infection reported in less deprived areas, a feature that was attributable to reduced overseas travel. At least part of the difference is likely to be a result of real differences in rates of infection, although some may be due to differences in ascertainment.

Cluster analysis identified only a small percentage (2.3%) of cases to be associated with a cluster. Similar analyses at the resolution of the household, showed only a small percentage of cases to be associated with household outbreaks (3%). These results confirm that the majority of cases are sporadic.

Under-reporting of *Campylobacter* incidence was found in some health board areas. The reasons for the difference in reporting rates between health boards are unclear but it does not appear to be attributable to the microbiological methods. It seems unlikely that the differences are real (ie that these regions actually have lower *Campylobacter* rates) because the changes are abrupt across the regional boundaries. Accurate quantification of the level of under reporting would aid further studies.

The majority of cases can be associated with poultry or ruminant attributed strains. The strains have differing spatial and temporal distributions and differing epidemiologies: ruminant strains are more common in children, more frequently acquired in summer, are associated with the presence of sheep. Though proximity to animal reservoirs was examined as a potential risk factor our analyses did not identify a significant risk, other than in children living in rural areas who had an enhanced risk of infection in areas of higher sheep density. Likely sources of infection are environmental for ruminant associated strains and consumption of contaminated food for poultry associated strains. The results from this project together with the FSAS project S14006 clearly show that retail chicken as well as ruminants are important sources of human campylobacteriosis.

3. INTRODUCTION

3.1 Background to study

3.1.1 Epidemiology and risk factors for *Campylobacter* infection

Campylobacter are Gram-negative bacteria that live commensally in the gastrointestinal tracts of a wide range of animals and birds, including farmed species and companion animals. Some *Campylobacter* species are also zoonotic human pathogens. A typical human infection consists of a self-limiting bout of diarrhoea, abdominal cramps and fever lasting about five days. *Campylobacter* infection was implicated in causing human enteritis in the late 1970s and is now the largest cause of bacterial gastrointestinal infection in the developed world (Blaser, 1997). According to WHO estimates, *Campylobacter*-related illness affects around 1% of populations in developed countries every year.

The epidemiology of human *Campylobacter* infections is poorly understood but it is widely acknowledged that most cases are sporadic with relatively few general outbreaks. Where outbreaks do occur, poultry (O'Brien et al., 2002) and private water supplies (Said et al., 2003) have been identified as probable vehicles of infection.

In an effort to explain the causes of sporadic disease, case control studies have been carried out but care should be taken in interpreting these findings because in general, only a small proportion of cases are associated with causal factors and immunity in the controls may confound the results (Cowden, 1992). Poultry eaten cooked (Harris et al, 1986), rare (Friedman et al, 2000) or consumed in a commercial food establishment (Rodrigues et al, 2001; Friedman et al, 2000) have been demonstrated to be risk factors. However, some studies have shown that consumption of chicken at home is protective (Adak et al, 1995; Ikram et al, 1994). Other risk factors include different foods (barbecued beef (Kapperud et al, 1992), raw milk, (Studahl et al, 2000), bird pecked milk (Lighton et al, 2001), contact with pets or farm animals (Adak et al, 1995; Neal and Slack, 2000), and foreign and domestic travel (Neal and Slack, 1997). Recent research in the UK indicates that foreign and domestic travel is associated with 20% and 13% of cases, respectively (the *Campylobacter* Sentinel Surveillance Scheme Collaborators, 2003).

A study of *Campylobacter* cases in Gloucestershire (O'Neill et al, 2004) demonstrated that 9% of cases were travel associated and that higher incidence was observed in (a) the 0-4 year age group in rural areas compared with urban areas, (b) affluent areas and (c) areas with high percentage of private water supplies. A similar study in Grampian (Miller et al, 2004b) also showed a higher incidence in young children living in rural areas with a 50% higher incidence than their urban counterparts.

These studies point to the need for a coherent, large scale statistical analysis of *Campylobacter* incidence that can combine information on sources of infection with the geographic location, demography, proximity to potential reservoirs of infection, economic status and lifestyle of the cases.

3.1.2 Campylobacter in Scotland

In Scotland, reported *Campylobacter* rates increased in the 1990's peaking at 121 cases per 100,000 population in 2000. Subsequently they fell to 98 per 100,000 by 2006, but had risen again to 123/100,000 in 2009 (Pollock et al., 2010.). Furthermore, there is an underlying spatial variation in incidence that we sought to address. The incidence of *Campylobacter* varies significantly between Health Board areas in Scotland (Fig. 3.1) but it is not known why these differences exist.

Reporting bias is one possible explanation. In England and Wales it is estimated that for every case that is reported to national surveillance another 7.6 go unreported (Wheeler et al, 1999) and it seems likely that there is a similar situation in Scotland. Differences in ascertainment could occur at the level of the patient, the GP practice, or the testing laboratory.

In Scotland, it is routine practice for the clinical diagnostic laboratories to report all laboratory confirmed cases of *Campylobacter* infection to the Public Health Team in the NHS Board. These data, which include name, date of birth, address and postcode, are made available to HPS by the Public

Health Teams. The extent to which additional epidemiological information from cases is collected varies between NHS boards. In a number of boards an enteric questionnaire is sent to all cases of *Campylobacter* infection either directly by the Public Health Team or by the local Environmental Health Officers. The content of these enteric questionnaires vary between boards but generally collect information on a range of potential risk factors including history of overseas travel prior to onset of symptoms, pet and farm animal contact, drinking water from a private water supply and details of recent food consumption.

Enhanced surveillance in Lothian reported higher incidence in affluent areas and also from those areas with a high percentage of private water supplies. However, to date there has been no concerted effort to use state of the art statistical techniques together with data visualisation and mapping approaches to address the situation in Scotland as a whole. This sets the nation some way behind the situation in England where there has been significant investment of effort at a number of levels in both human and veterinary aspects.

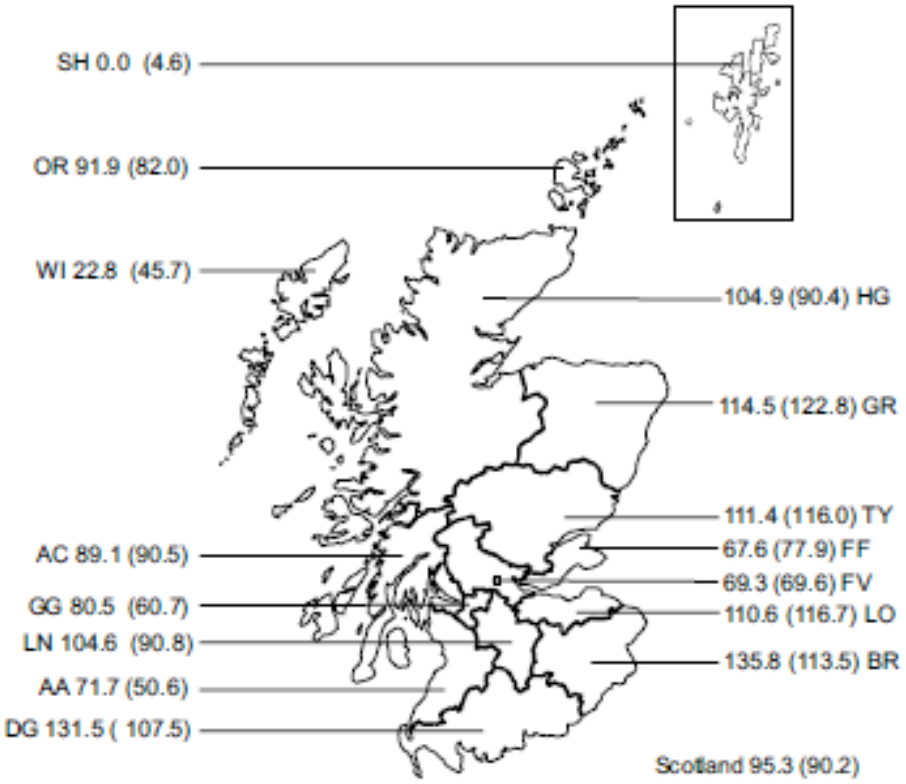


Figure 3.1 Rates per 100 000 population of reports of *Campylobacter* identifications to HPS, 2006 (2005).

To address the aims of the FSAs strategic plan for 2006-2010 to further reduce the incidence of foodborne disease requires an improved understanding of the geographic distribution of *Campylobacter* infection in Scotland. The results of existing studies point to a number of potential risk factors that might explain the observed georgraphic variation in incidence of infection:

Urban – rural It has been established that there is a higher rate of incidence of *Campylobacter* infection in young children in rural areas compared with urban areas (Miller et al, 2004b, O'Neill et al, 2004). It has been hypothesised that the reason for this is greater exposure to *Campylobacter* from the rural environment (e.g. farm animals and private water supplies).

Deprivation It has also been demonstrated that *Campylobacter* incidence is greatest in areas of least deprivation, for example in Lothian (personal communication Alison Smith-Palmer) and in

Gloucestershire (O'Neill et al, 2004). Deprivation could influence other epidemiologically important factors such as food choices (eg restaurant eating, home-cooked or ready prepared), uptake of health-care services and travel.

Animal Reservoirs Chicken is viewed as the predominant source of human infection. However, *Campylobacter* has been found wide range of animals and birds, including farmed species and companion animals which may be contributing to human infection. Proximity to different livestock species is therefore a potential risk factor for human infection

Private water supplies
Private water supplies have been implicated in outbreaks of a number of gastrointestinal diseases (Smith et al, 2006). The case-control study focusing on the Aberdeenshire area running in parallel with this project has identified a significantly enhanced risk of *Campylobacter* infection amongst those with private water supplies.

3.1.3 Typing approaches

One of the problems in understanding the epidemiology of *Campylobacter* has been the lack of a suitable typing method. Strain diversity is known to remain high throughout the infection chain: from sources through infection vehicles to human clinical cases, consistent with the existence of a very large pool of strains. According to this view, the main features of *Campylobacter* transmission will be evident only at large spatio-temporal scales, *e.g*., at a national scale. Most of the studies conducted to date on *Campylobacter* molecular epidemiology are small-scale because the strain typing methods used are unsuitable for high sample throughput, and this makes it difficult for any one study to yield national-scale conclusions. A larger-scale picture could emerge from combining raw datasets across studies but this is difficult to achieve because most of the strain typing methods used yield non-transferable data. These limitations of strain typing methods explain the dearth of studies conducted at national scales and covering a wide range of reservoir hosts and clinical cases. This new typing approach has the potential to add substantial power to analysis of the geographic distribution of cases.

3.2 Aims and objectives

The aim of the project was to provide the first comprehensive study of geographic variation in *Campylobacter* incidence within and between regions in Scotland. Using a combination of epidemiological and typing data the project characterised the spatial and temporal distribution of cases and developed statistical models that quantified the roles of putative risk factors – deprivation, private water supplies, livestock reservoirs, geographic location, urban versus rural living.

Our specific objectives were to:

- Provide a descriptive analysis of geographic and demographic distribution of cases of campylobacteriosis in humans, both with and without MLST profiling
- Identify appropriate statistical models that best describe spatial heterogeneity and variation
- Identify modifiable risks that are associated with an increased risk of campylobacteriosis in humans, stratified by MLST profile

4. MATERIAL AND METHODS

4.1 Ethical approval

Ethical approval for the study was obtained from Multi-centre Research Ethics Committee for Scotland A (MREC). Amendments to the study protocol were submitted to MREC for approval as necessary.

The NHS R&D Application Form was submitted to the Research and Development committee in each of the mainland NHS boards for approval before data collection started.

In all NHS Boards the extraction of information on *Campylobacter* cases was conducted with the approval of the Consultant in Public Health Medicine and access to the information was only possible through the assistance of the Public Health Teams.

4.2 Data Sources and Grouping

4.2.1 Human case data

It is routine practice for the clinical diagnostic laboratories to report all laboratory confirmed cases of *Campylobacter* infection to the Public Health Team in the NHS Board. HPS obtained data from the Public Health Teams for all laboratory confirmed cases reported from the start of 2000 to the end of 2006. This included name, date of birth, address and postcode. However where postcode was not available from the extracted data, but the address was, then the postcode was obtained using Quick Address Pro 3.15 and Royal Mail postcode books.

The data was checked for duplicate entries and one of these entries were removed from the dataset . A duplicate was defined as two samples from a case with the same name and date of birth with a report date within a 4-week period.

The routine collection of additional epidemiological information from cases varies between NHS boards. In a number of boards an enteric questionnaire is sent to all cases of *Campylobacter* infection either directly by the Public Health Team or by the local Environmental Health Officers. The content of these enteric questionnaires vary between boards but generally collect information on a range of potential risk factors including history of overseas travel prior to onset of symptoms, pet and farm animal contact, drinking water from a private water supply and details of recent food consumption. Where enteric forms were available for the period 2000-2006, they were borrowed from the NHS board and the information on travel, animal contact and private water supply use was extracted and added to the basic demographic information collected on all cases from the NHS Board. Linkage to cases was based on name, address, date of birth and report date.

Names, dates of birth and full address were removed from the data at HPS before being shared with the study partners at the University of Aberdeen and University of Glasgow. Additionally the postcode of cases was limited to the level of postcode sector (eg AB11 6** or G12 8**).

4.2.2 MLST data

4.2.2.1 Matching human MLST data to epidemiological case data

During July 2005 to September 2006, 5831 cases of human campylobacteriosis were reported by national surveillance in Scotland. Under FSAS project S14006, clinical *Campylobacter* isolates were submitted from public health bacteriology laboratories and typed by MLST as previously described (Sheppard et al, 2009). Linkage between epidemiological data (date of birth, sex and date of report) and MLST type was conducted at HPS and achieved for 3713 cases. The pie charts (Fig. 4.1, Table 4.1 and 4.2) display the relative abundance of sequence types for human cases of *C. jejuni* and *C. coli* respectively

(a) (b)

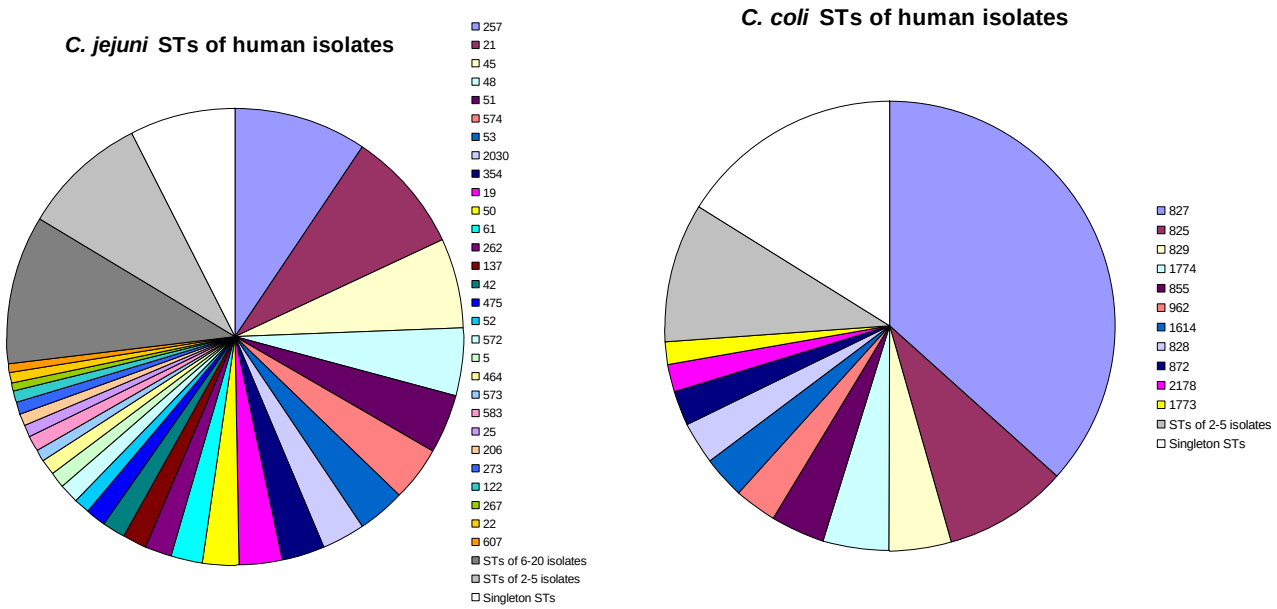


Figure 4.1 Relative abundance of human sequence types: (a) *C. jejuni* and (b) *C. coli.*

Table 4.1 Frequencies of *C. jejuni* sequence types found in human clinical samples.

ST	**No. of human isolates (%)**	**ST**	**No. of human isolates (%)**
257	330 (9.3)	572	43 (1.2)
21	307 (8.7)	5	40 (1.1)
45	233 (6.6)	464	39 (1.1)
48	168 (4.7)	573	36 (1.0)
51	143 (4.0)	583	35(1.0)
574	137 (3.9)	25	31 (0.9)
53	121 (3.4)	206	30 (0.8)
2030	110 (3.1)	273	30 (0.8)
354	109 (3.1)	122	27 (0.8)
19	105 (3.0)	267	24 (0.7)
50	87 (2.4)	22	22 (0.6)
61	80 (2.3)	607	21 (0.6)
262	68 (1.9)	STs of 6-20 isolates	374 (10.6)
137	63 (1.7)	STs of 2-5 isolates	319 (9.0)
42	52 (1.5)	Singleton STs	265 (7.5)
475	52 (1.5)		
52	43 (1.2)	**Total**	**3554 (100)**

Table 4.2 Frequencies of *C. coli* sequence types found in human clinical samples.

ST	No. of human isolates (%)	ST	No. of human isolates (%)
827	140 (35.5)	828	11 (2.8)
825	34 (8.6)	872	10 (2.5)
829	17 (4.3)	2178	7 (1.8)
1774	17 (4.3)	1773	6 (1.5)
855	15 (3.8)	STs of 2-5 isolates	39 (10.2)
962	12 (3.1)	Singleton STs	61 (15.5)
1614	12 (3.1)	**Total**	**381 (100)**

4.2.2.2 MLST data from animal and food sources

Environmental and retail food isolates of *Campylobacter* were typed by MLST in FSAS project S14006. Figure 4.2 and Table 4.3 shows the distribution of *C. jejuni* sequence types for each animal reservoir species. The distribution of *C. coli* species (this also includes strains from PubMLST database) is presented in Figure. 4.3 and Table 4.4.

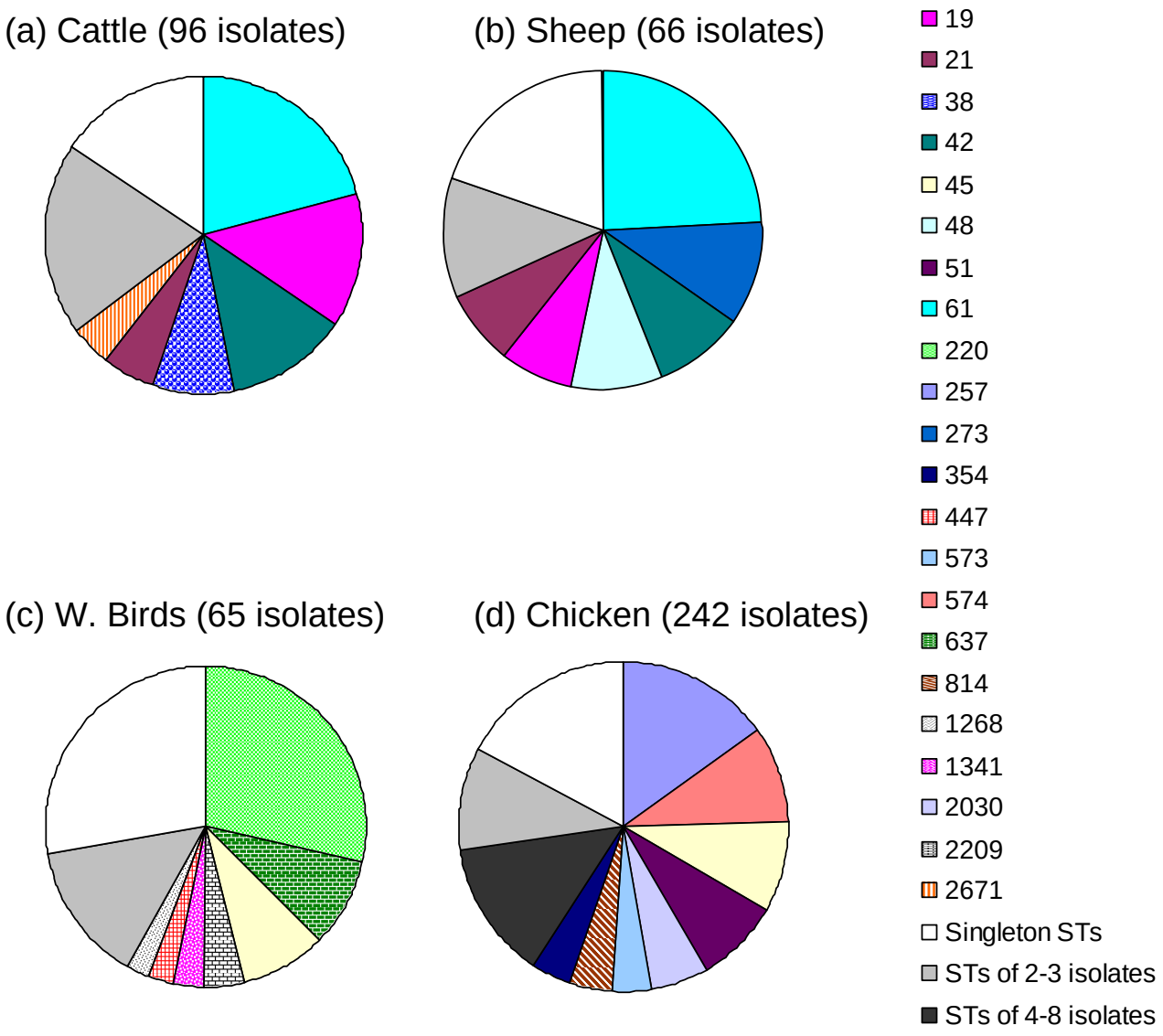


Figure 4.2 Frequencies of *C. jejuni* sequence types found in (a) cattle, (b) sheep, (c) wild birds and (d) retail chicken.

Table 4.3 Frequencies of *C. jejuni* sequence types found in (a) cattle, (b) sheep, (c) wild birds and (d) retail chicken.

ST	**No. of cattle (%)**	**No. of sheep (%)**	**No. of birds (%)**	**No. of retail chicken (%)**
19	13 (13.5)	5 (7.6)		
21	5 (5.2)	5 (7.6)		
38	8 (8.3)			
42	12 (12.5)	6 (9.1)		
45			14 (8.5)	22 (11.0)
48		6 (9.1)		
51				20 (10.0)
61	20 (20.8)	16 (24.2)		
220			47 (28.5)	
257				36 (18.0)
273		7 (10.6)		
354				
447			4 (2.4)	
573				10 (5.0)
574				23 (11.5)
637			15 (9.1)	
814				10 (5.0)
1268			4 (2.4)	
1341			5 (3.0)	
2030				
2209			7 (4.2)	
2671	4 (4.2)			
Singleton STs	15 (15.6)	13 (19.7)	46 (27.9)	42 (21.0)
STs of 2-3 isolates[a]	19 (19.8)	8 (12.1)	23 (13.9)	24 (12.0)
STs of 4-8 isolates[a]				33 (16.5)
Total	**96 (100)**	**66 (100)**	**165 (100)**	**200 (100)**

[a]Does not include ST's specifically mentioned in the table.

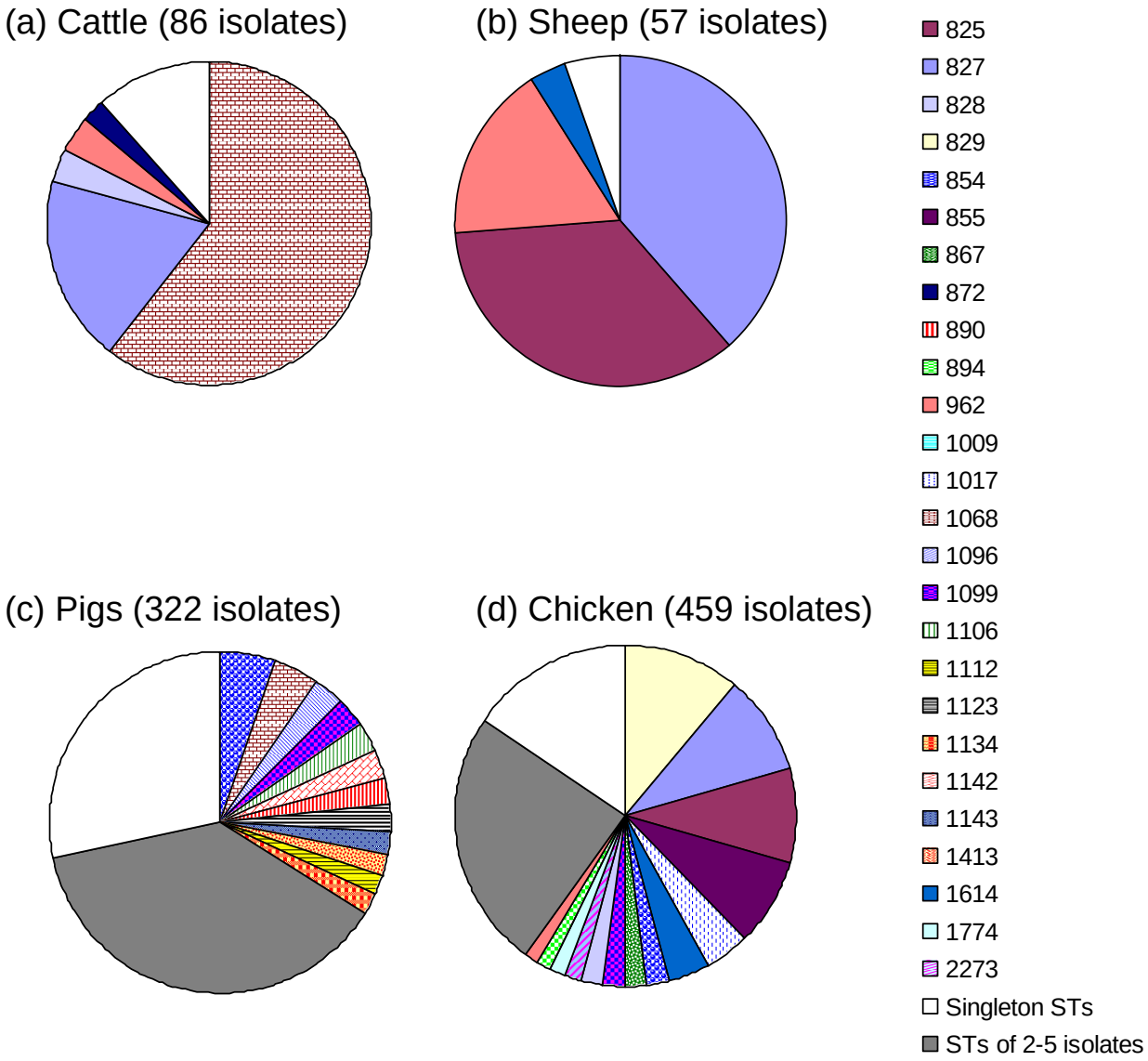


Figure 4.3 Frequencies of *C. coli* sequence types found in (a) cattle, (b) sheep, (c) pigs and (d) chicken.

Table 4.4 Frequencies of *C. coli* sequence types found in (a) cattle, (b) sheep, (c) pigs and (d) retail chicken.

ST	**No. of cattle (%)**	**No. of sheep (%)**	**No. of pigs (%)**	**No. of retail chicken (%)**
825		20 (35.1)		41 (8.9)
827	16 (18.6)	22 (38.6)		43 (9.4)
828	3 (3.5)			8 (1.8)
829				51 (11.1)
854			17 (5.3)	10 (2.2)
855				39 (8.5)
867				10 (2.2)
872	2 (2.3)			
890			8 (2.5)	
894				6 (1.3)
962	3 (3.5)	10 (17.5)		6 (1.3)
1009				10 (2.2)
1017				19 (4.1)
1068	52 (60.5)		13 (4.0)	
1096			10 (3.1)	
1099			9 (2.8)	
1106			9 (2.8)	
1112			6 (1.9)	
1123			8 (2.5)	
1134			6 (1.9)	
1142			9 (2.8)	
1143			7 (2.1)	
1413			7 (2.1)	
1614		2 (3.5)		17 (3.7)
1774				7 (1.5)
2273				8 (1.7)
Singleton STs	10 (11.6)	3 (5.3)	91 (28.3)	72 (15.7)
STs of 2-5[a] isolates			122 (37.9)	112 (24.4)
Total	**86 (100)**	**57 (100)**	**322 (100)**	**459 (100)**

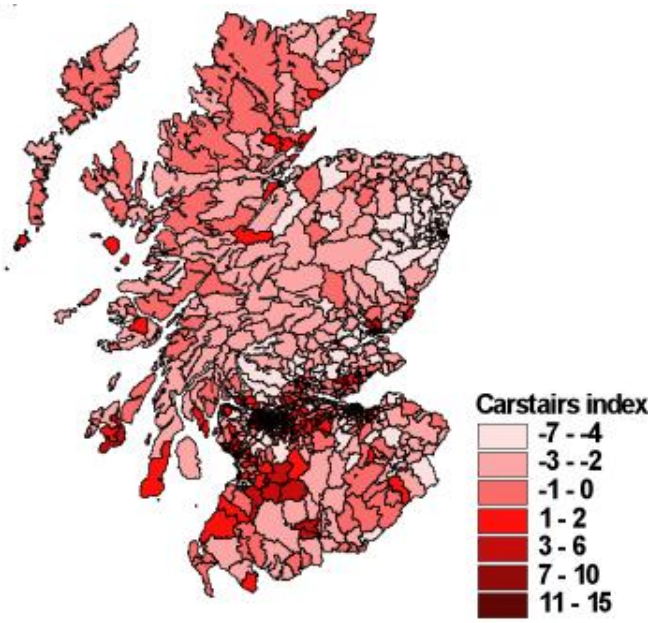
[a]Does not include ST's specifically mentioned in the table.

4.2.3 Deprivation indices

Representative socio-economic indices (Carstairs, DEPCAT, SIMD, MOSAIC) were collected from external organisations (2001 census, Scottish Government, Experian) and collated at postcode sector level. These were mapped across Scotland and used as “risk factors” in univariate regression models, but as a consequence of their close agreement in categorising regions, only the Carstairs index was used in the final analysis (multivariate regression models). A description of these socioeconomic indices is given in the appendix.

The Carstairs index was obtained from MRC Social & Public Health Sciences Unit (Carstairs scores for Scottish postcode sectors from the 2001 Census, Philip McLoone’s report, University of Glasgow, http://www.msoc-mrc.gla.ac.uk/). It is a socio-economical index for the Scottish population that was derived by combining selected variables (e.g. number of cars owned per household, male unemployment, overcrowding etc.) to generate indices at postcode sector level. The Carstairs score ranges between -7 to +15, with the lowest scores representing the most affluent regions of Scotland and the highest the most deprived ones. The spatial distribution of this index is shown in Figure 4.4.

(a)



(b)

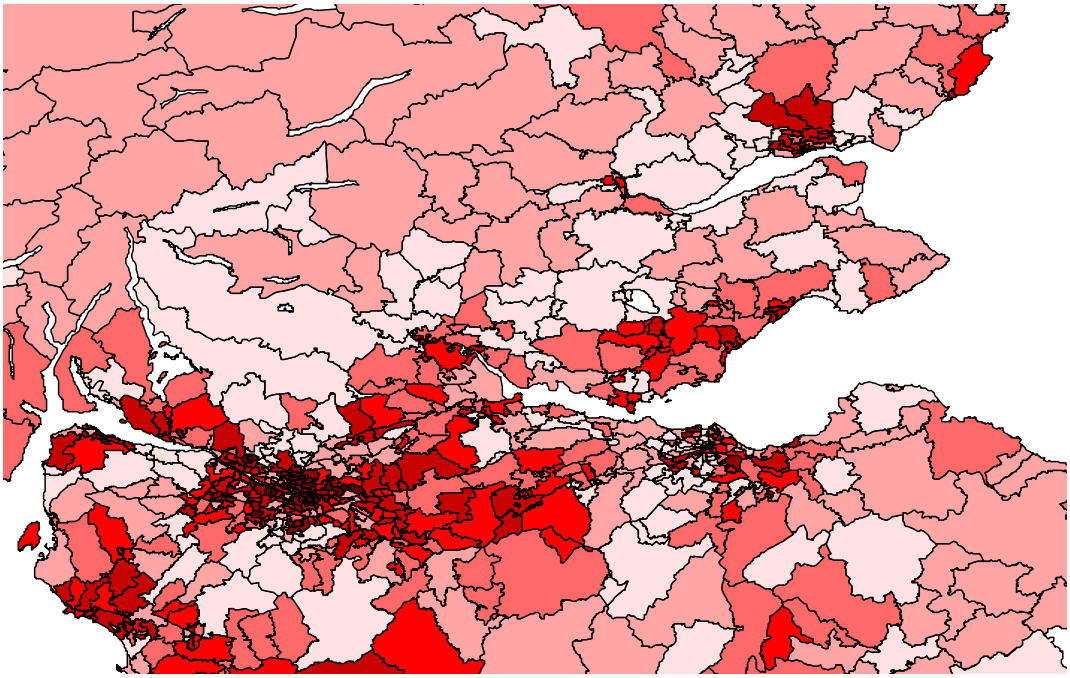


Figure 4.4. Distribution of the Carstairs index across (a) mainland Scotland and (b) Central belt at postcode sector level.

4.2.4 Animal densities

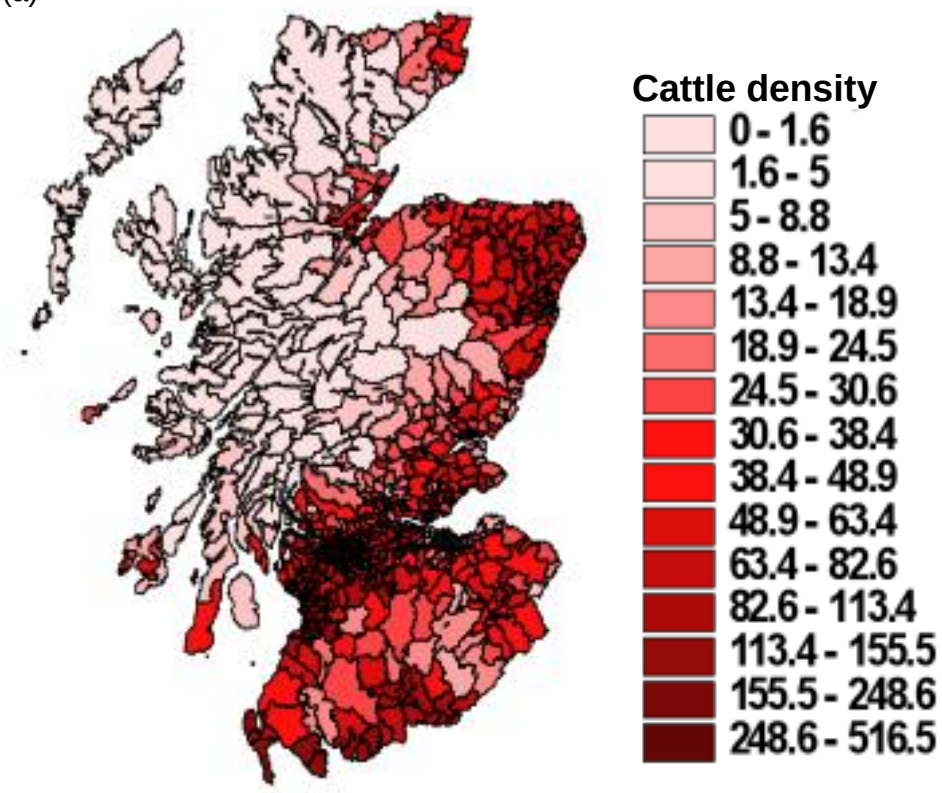
Farm animal data were obtained from the Scottish agricultural census (EDINA, http://edina.ac.uk/agcensus, 2004 estimates). Table 4.5 shows the number of animals by species in Scotland. The number of animals for nine categories (cattle, sheep, pigs, broilers, ducks, geese, total poultry, horses, deer) were downloaded at 2 x 2 km^2 spatial resolution and integrated at postcode sector level. Maps of animal densities were drawn in ArcView 3.3 (Environmental Systems Research Institutes, Inc., http://www.esri.com/software/arcview) (Fig. 4.5). These data were used as "risk factors" in the regression models.

Table 4.5 Farm animal numbers in Scotland (Source: Agricultural census, 2004).

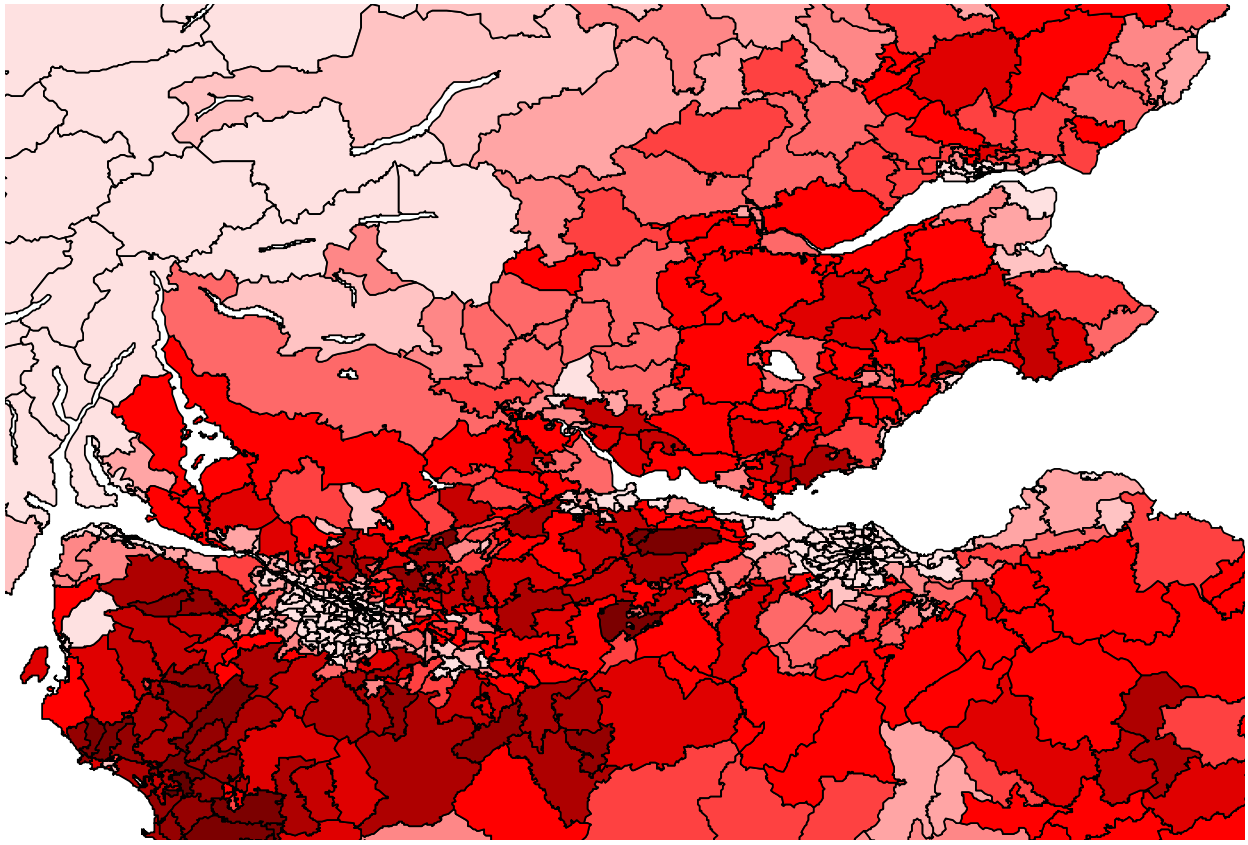
Animal	Number
Cattle	1953556
Pigs	469984
Broilers	10713067
Ducks	9685
Geese	2879
Total Poultry	15855521
Sheep	7860254
Horses	20206
Deer	6239

Number of animals/km^2 - Cattle

(a)

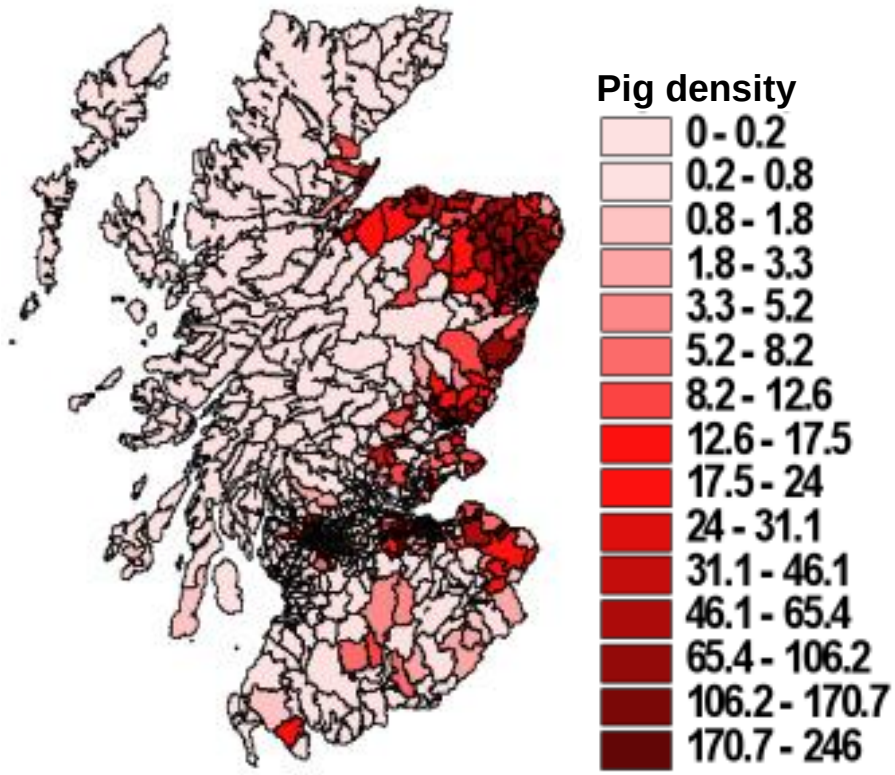


(b)

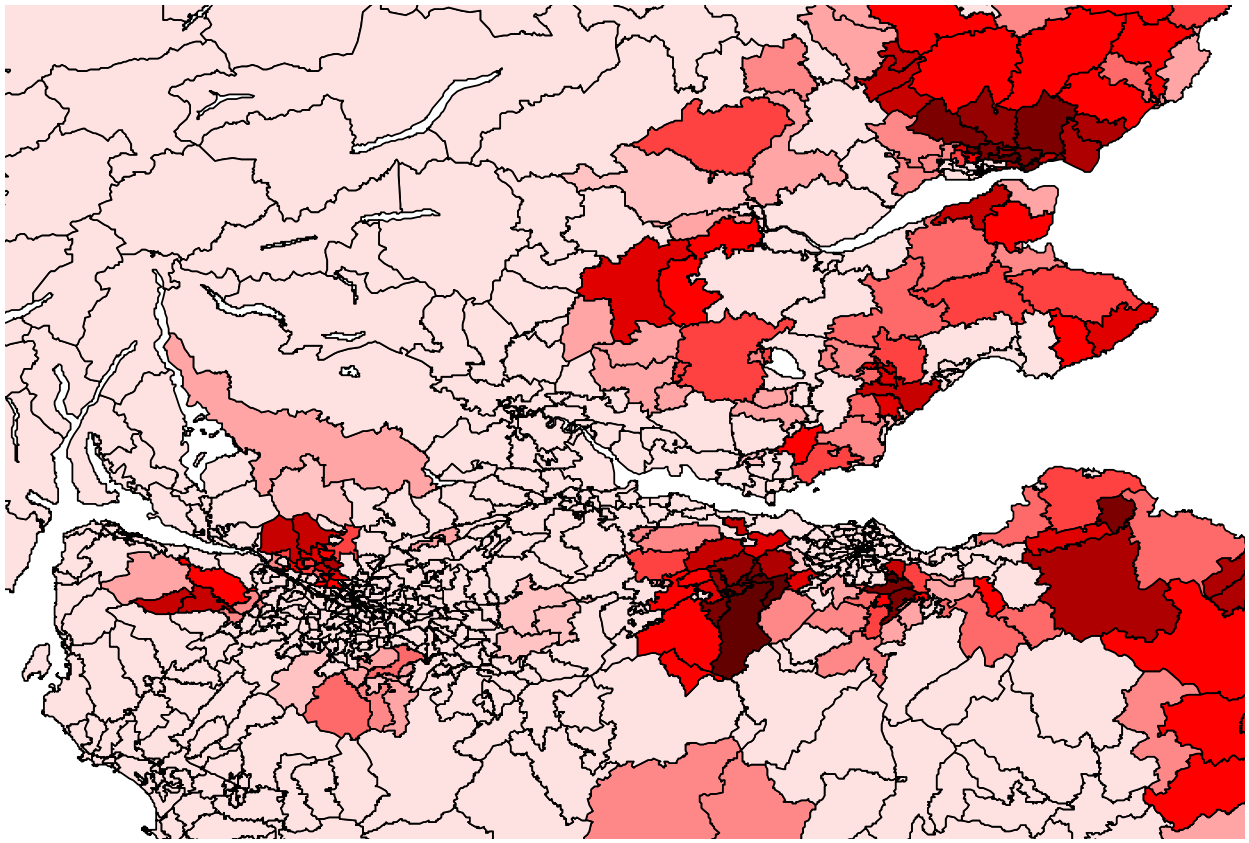


Number of animals/km^2 - Pigs

(c)

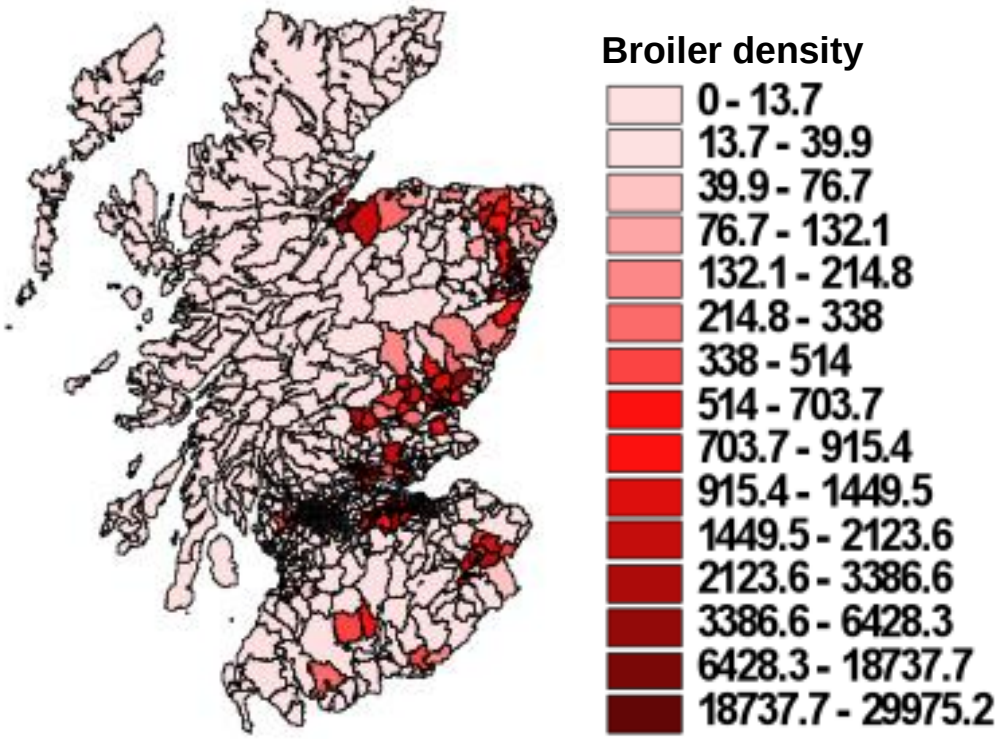


(d)

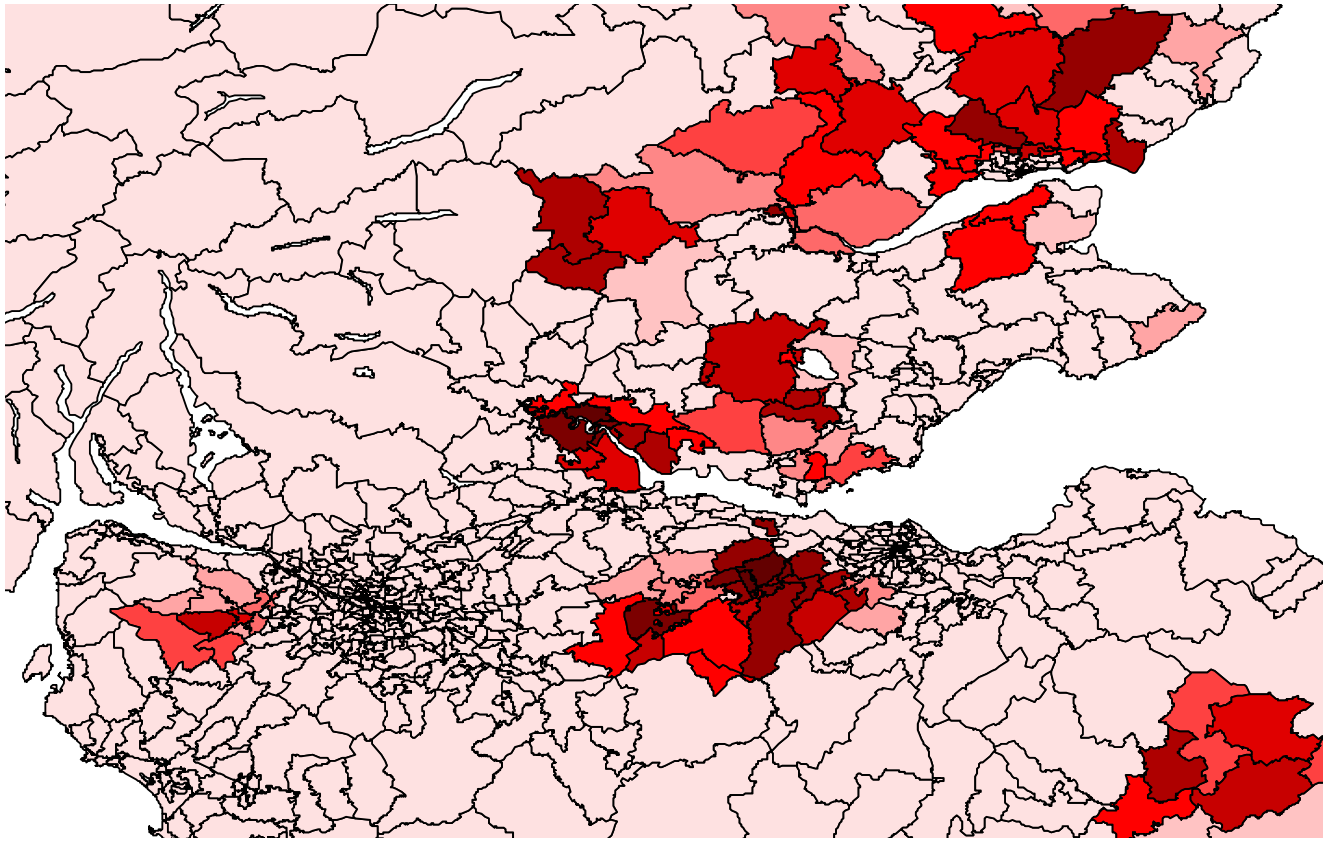


Number of animals/km^2 - Broilers

(e)

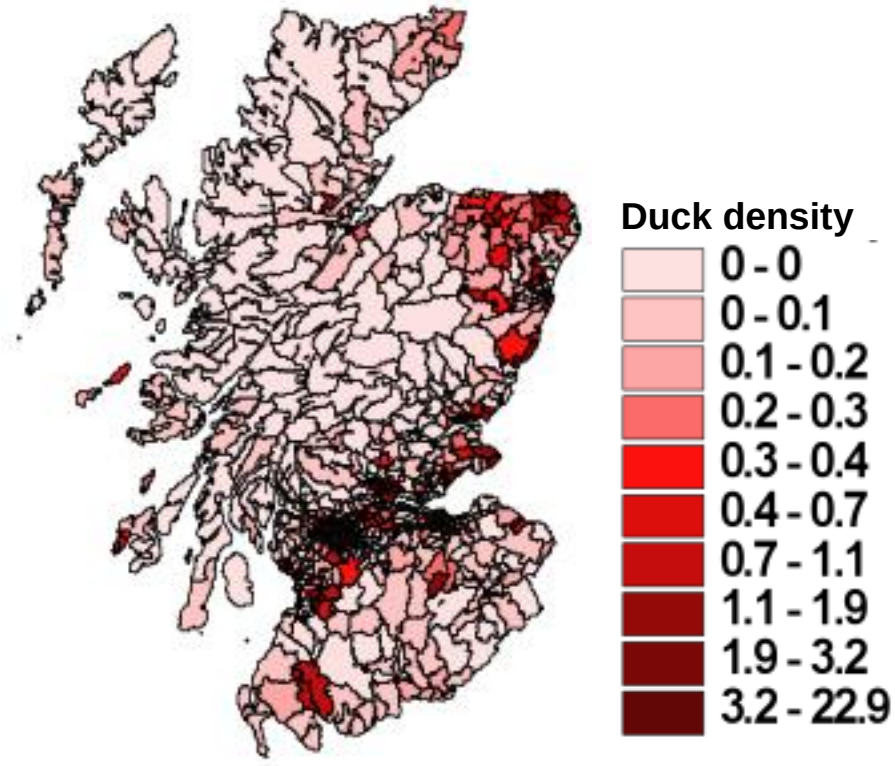


(f)

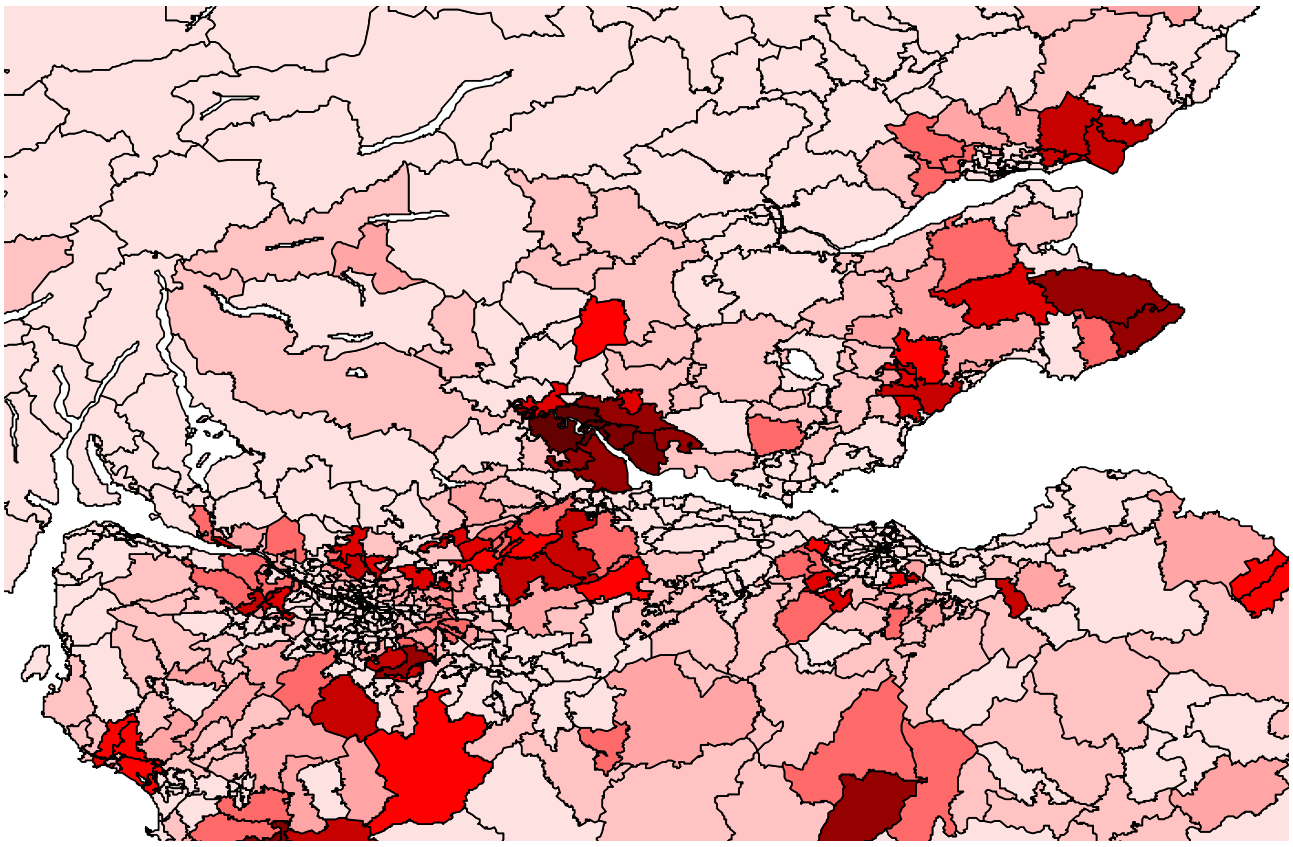


Number of animals/km^2 - Ducks

(g)

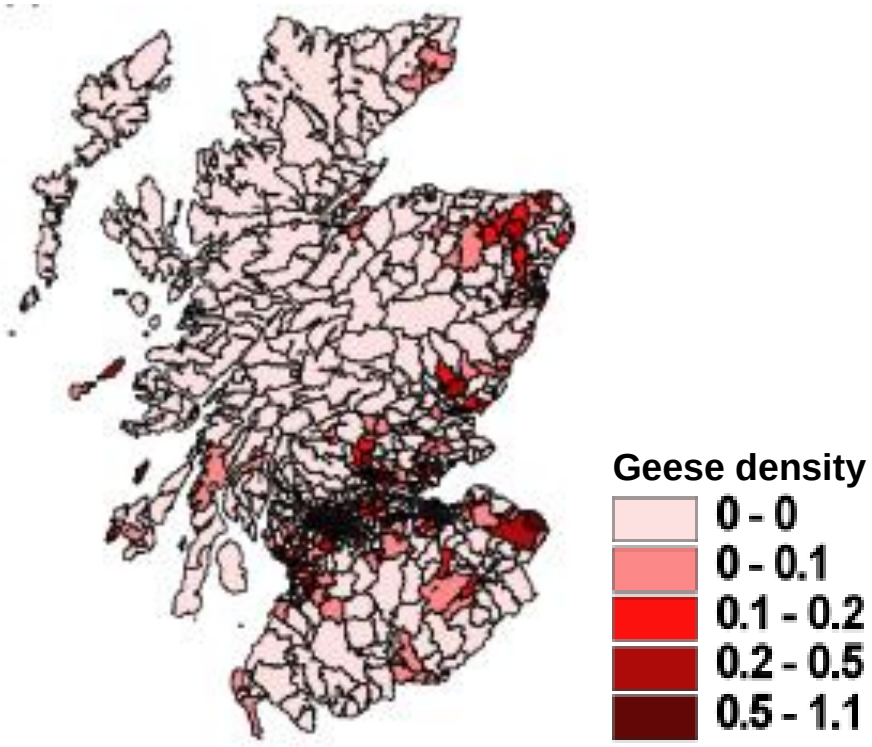


(h)

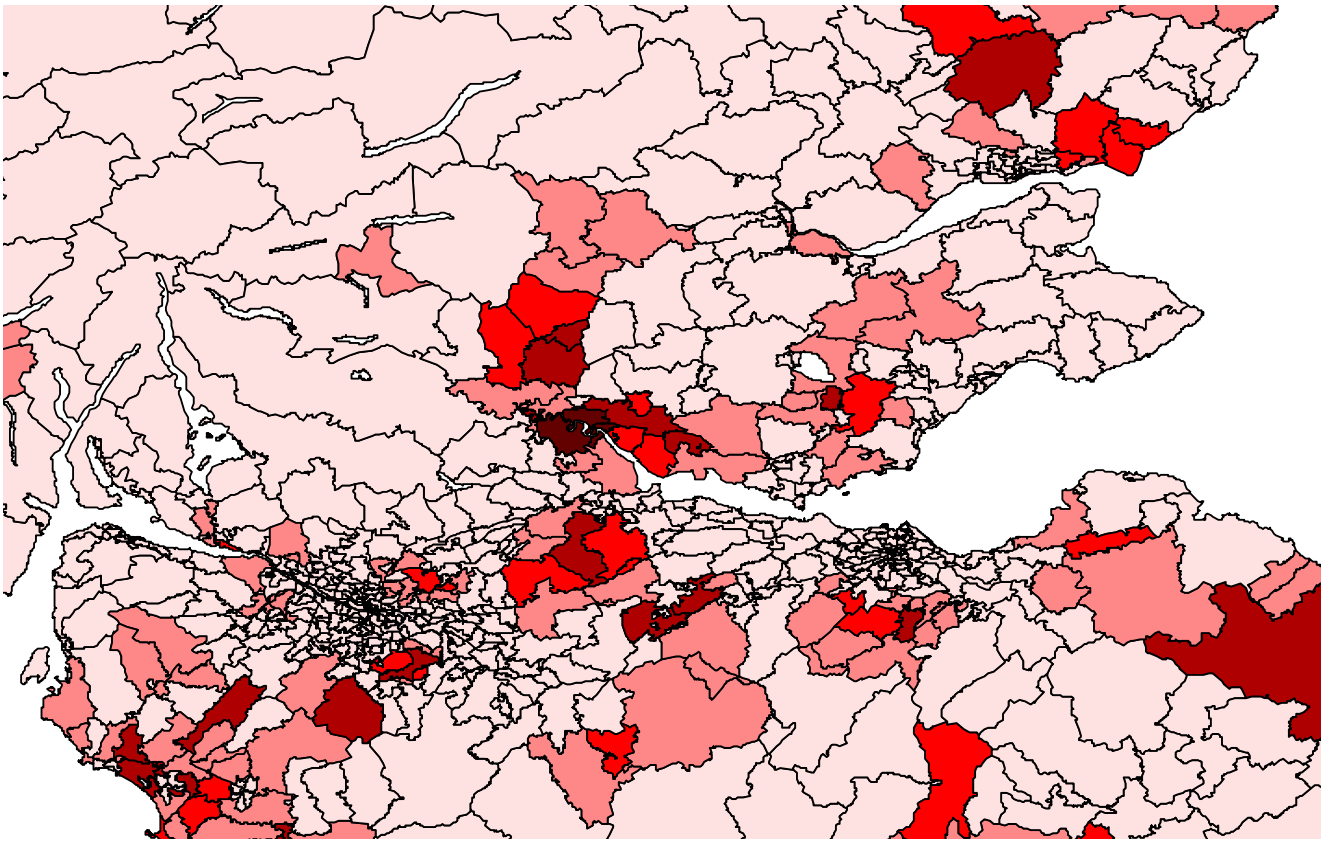


Number of animals/km^2 - Geese

(i)

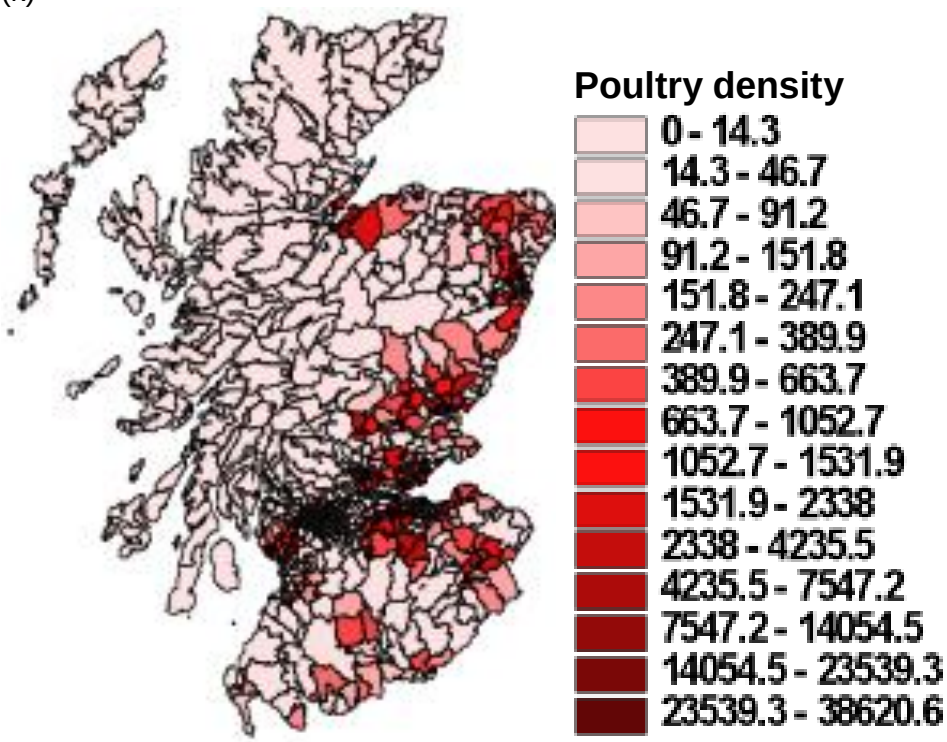


(j)

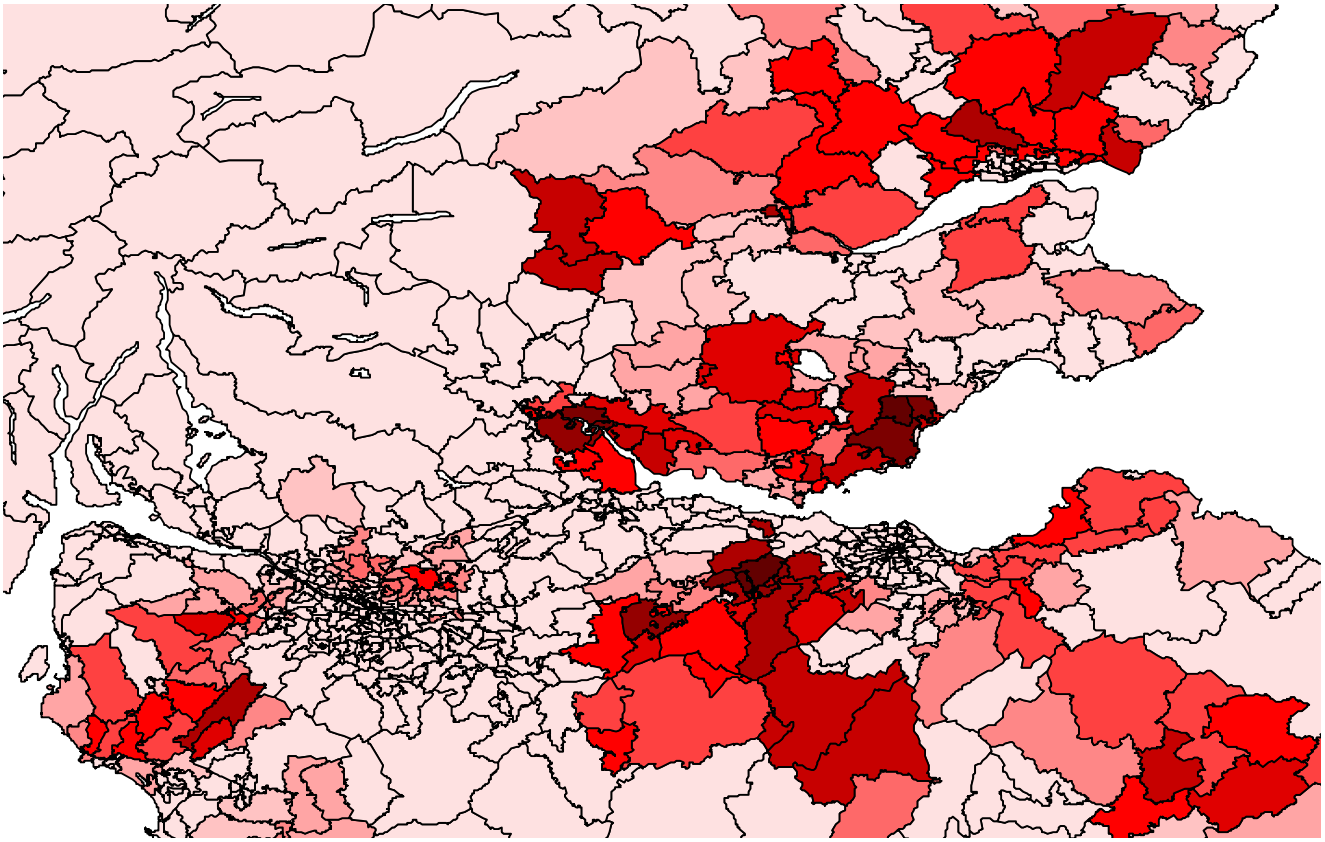


Number of animals/km^2 - Poultry

(k)

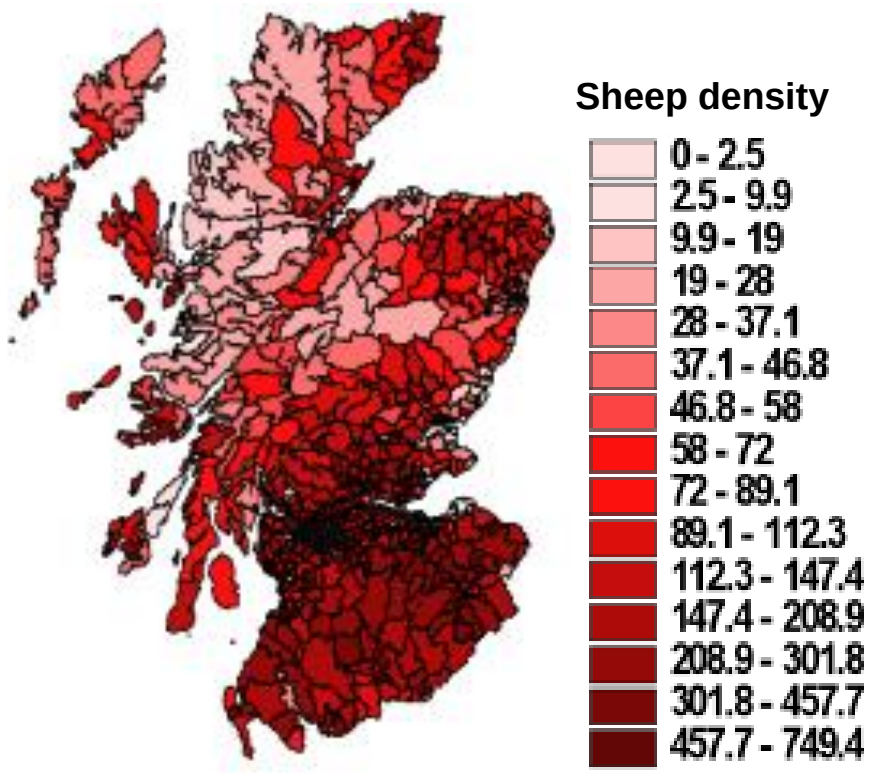


(l)

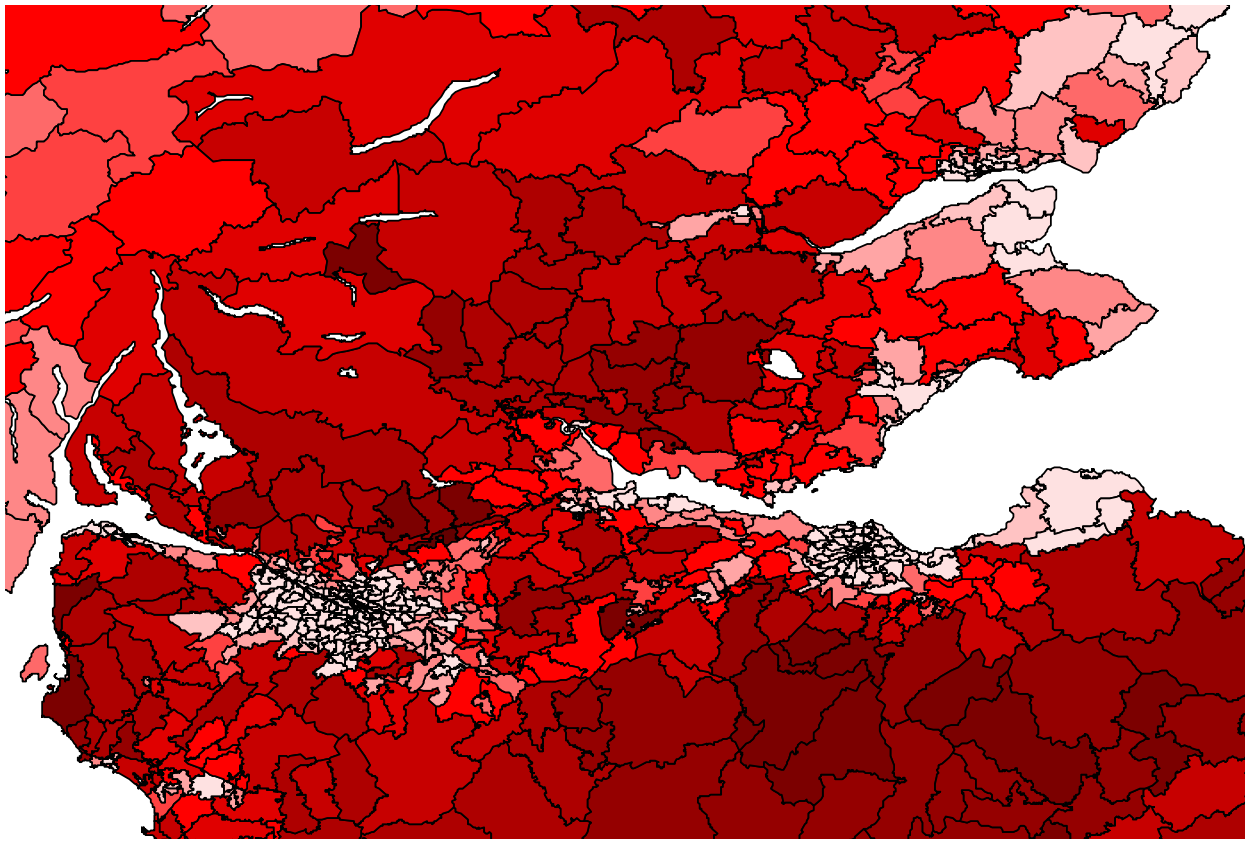


Number of animals/km^2 - Sheep

(m)

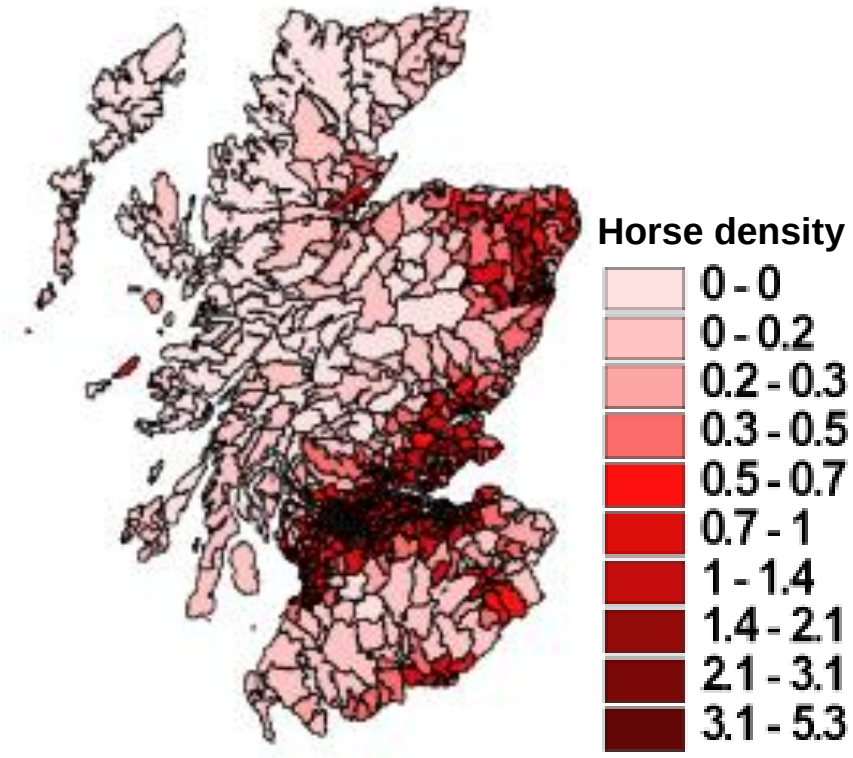


(n)

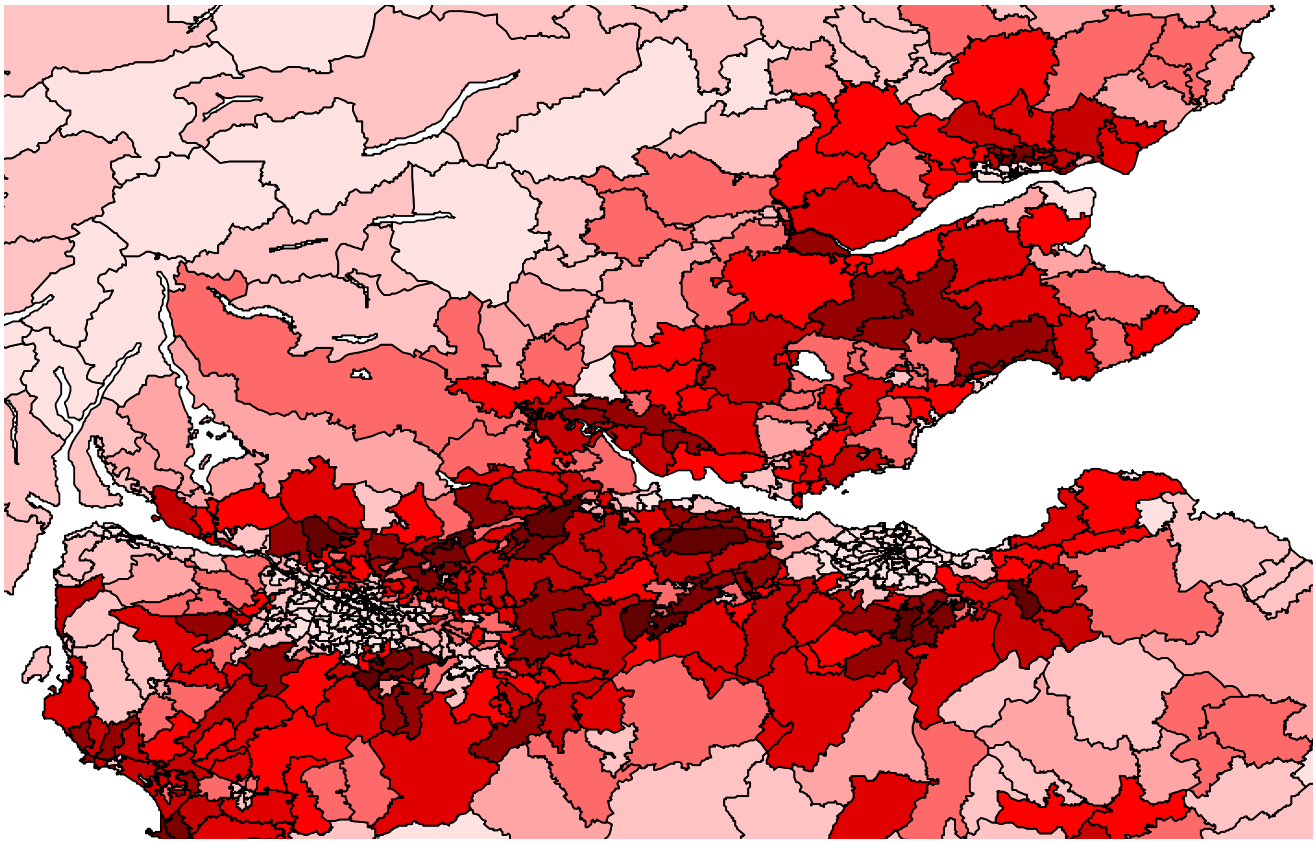


Number of animals/km^2 - Horses

(o)

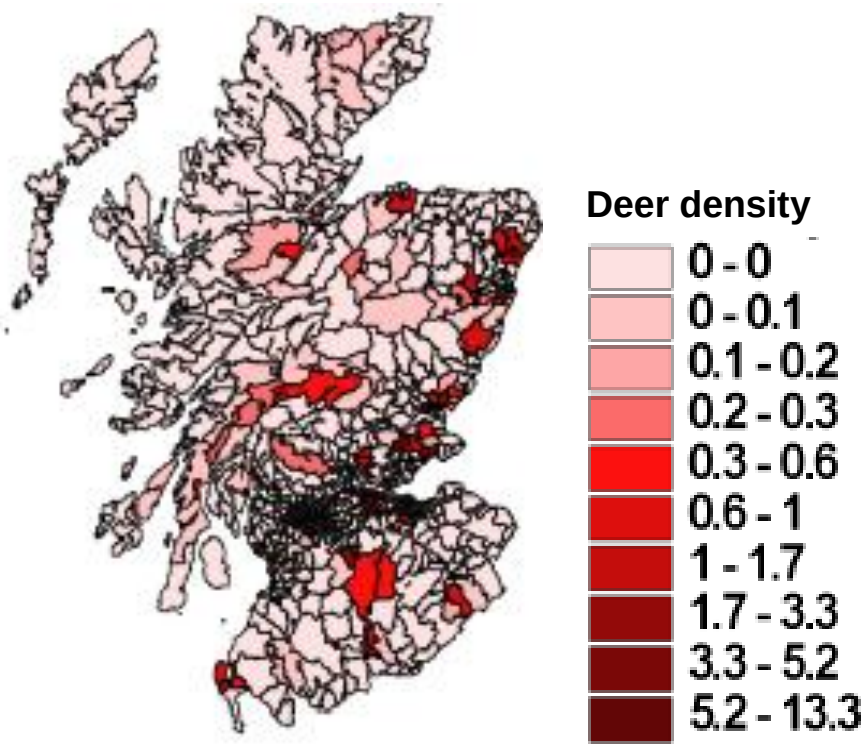


(p)



Number of animals/km^2 - Deer

(q)



(r)

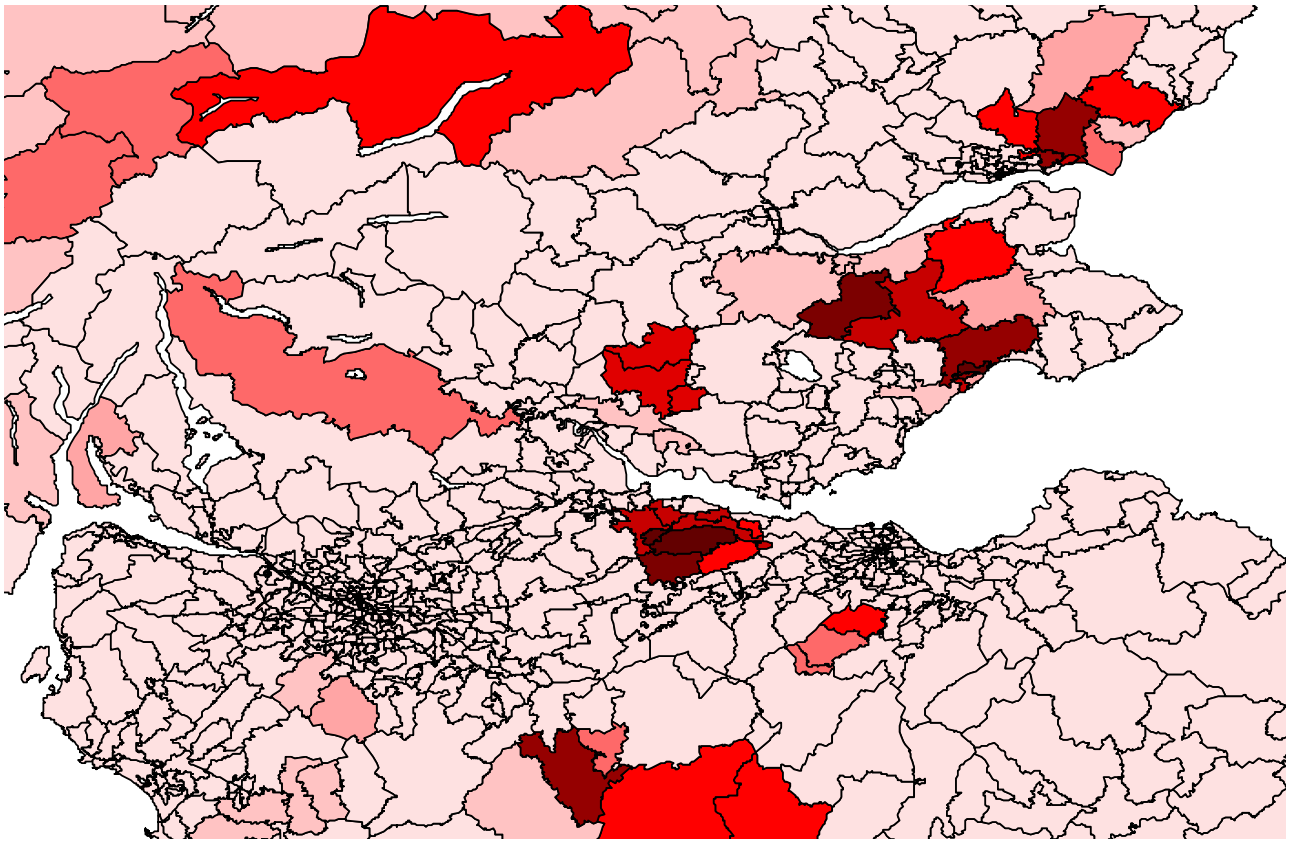
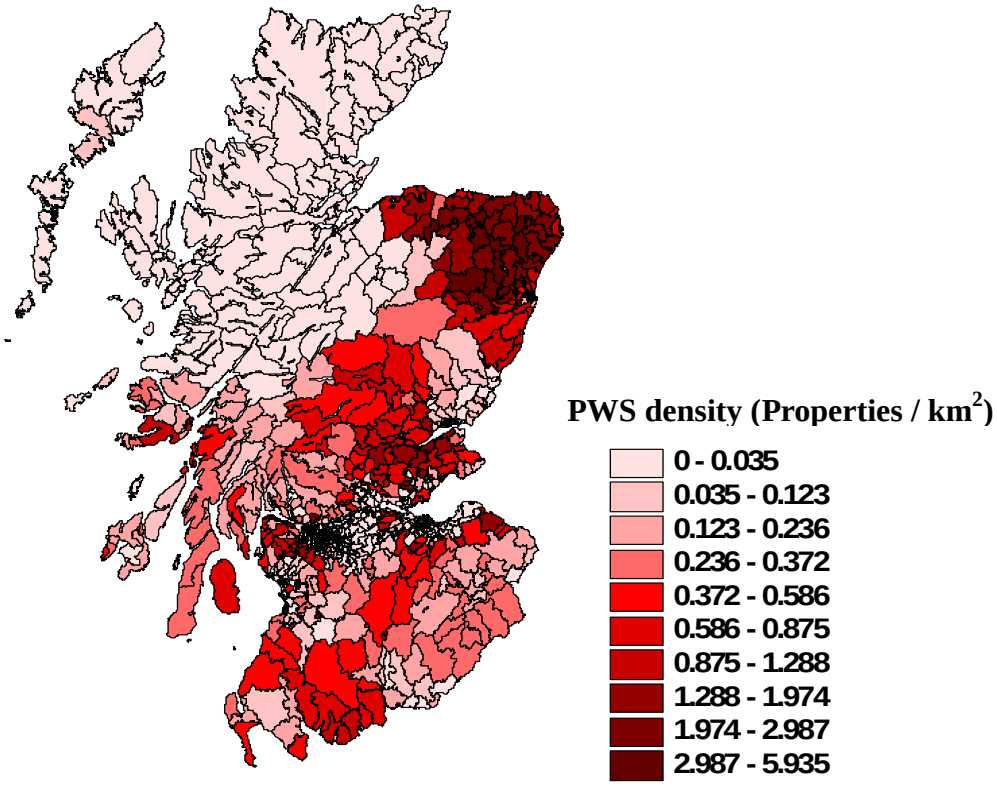


Figure 4.5 Distribution of farm animal data (number of animals / km^2) across mainland Scotland and central belt at postcode sector level: (a-b) cattle, (c-d) pigs, (e-f) broilers, (g-h) ducks, (i-j) geese, (k-l) total poultry, (m-n) sheep, (o-p) horses and (q-r) deer.

4.2.5 Private water supplies

Information on the number of properties on a private water supply (PWS) (reports available for 2007) was obtained from the water authorities, which are under the jurisdiction of the Scottish local authorities. The location of each property supplied by a PWS was obtained for 30 out of the 32 local authorities in various formats (e.g. address, coordinates and postcode). The data were integrated together at postcode sector level and displayed in map format (Fig. 4.6). The two missing authorities were Edinburgh and Highland. For Highland the addresses were incomplete for more than 50 % of the properties on PWS and therefore were unable to be included in the mapping. A summary of the PWS status in 2007 is given in Table 4.6.

(a)



(b)

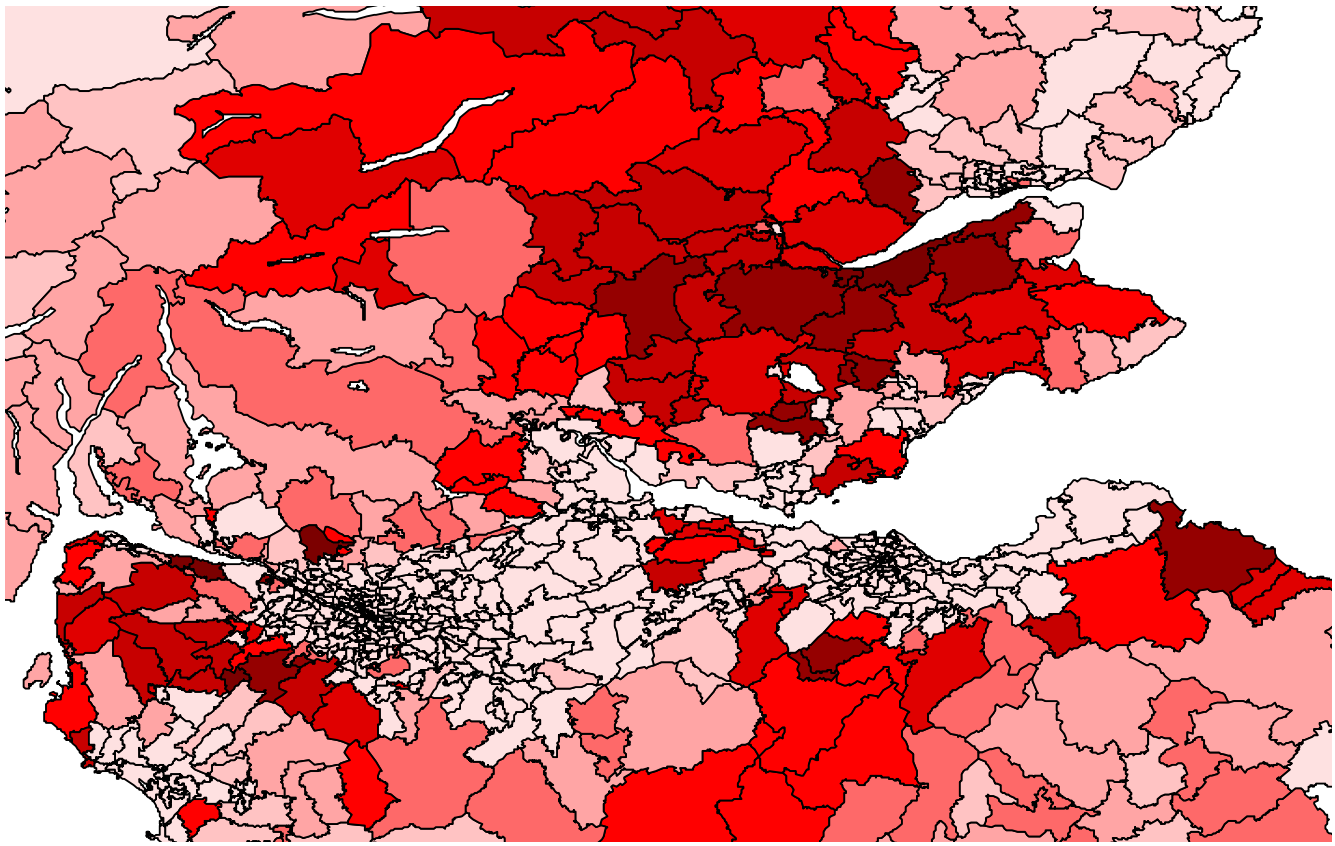


Figure 4.6. Density of properties on Private Water Supply (Properties/km^2) across Scotland (a) and central belt (b) at postcode sector level. Note Highland and Edinburgh are omitted due to data unavailability.

Table 4.6 Number of properties on private water supplies for the 32 city councils (2007 register)

Council	Number of properties on PWS
Aberdeen City	151
Aberdeenshire	11593
Angus	204
Argyll & Bute	1540
Clackmannanshire	104
Comhairle nan Eilean Siar	84
Dumfries and Galloway	2474
Dundee City	1
East Ayrshire	157
East Dunbartonshire	48
East Lothian	427
East Renfrewshire	117
Edinburgh City	~50**
Falkirk District	11
Fife	865
Glasgow City	-*
Highland	7712***
Inverclyde	96
Midlothian	223
Moray	2212
North Ayrshire	493
North Lanarkshire	18
Orkney	125
Perth and Kinross	3741
Renfrewshire	157
Scottish Borders	1176
Shetland Islands	101
South Ayrshire	470
South Lanarkshire	532
Stirling	668
West Dunbartonshire	21
West Lothian	107

*no properties on private water supplies
**approximately 50 properties on PWS for which data was not obtained
***incomplete addresses for >50% of the properties

4.2.6 Human population density and boundary data

Human population data were collected from the 2000 Scottish population estimate (General Register Office for Scotland, http://www.gro-scotland.gov.uk/files/00sap5.xls). The distribution of the Scottish population is presented in Figure 4.7. In the current study the population is defined as urban unless the population density is <200 /km^2 when it is classified as rural.

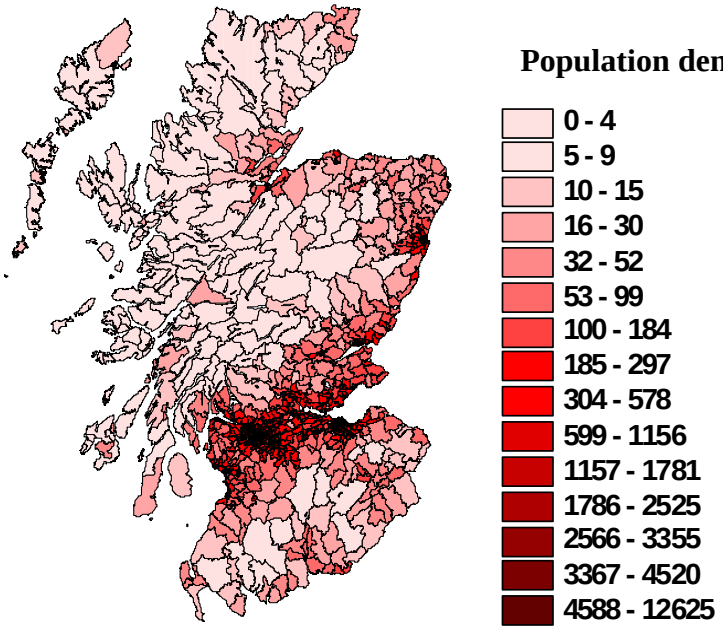


Figure 4.7. Population density (/ km^2) across Scotland at postcode sector level.

The age distribution of the population is given in Figure 4.8.

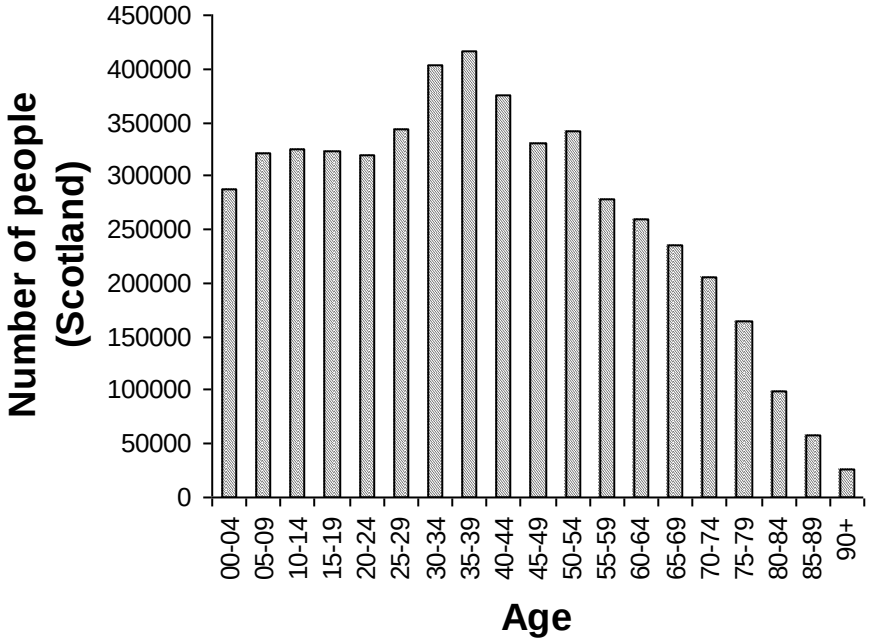


Fig.4.8 The Scottish population stratified by five years age groups in 2000.

4.3 Descriptive statistics

The incidence of human *Campylobacter* infections was determined at national, health board and postal sector level (2000-2006). The incidence was also stratified by both age and gender. Analysis of variance was performed using Microsoft Excel to identify differences in health board incidence during the period of the study.

4.4 Reporting Bias

Statistical epidemiology and analysis of the microbiological methods were performed and used to evaluate if there was a bias in the reporting of *Campylobacter* cases between Scottish health boards (HBs). The scope of this study did not allow the full reporting pyramid to be examined and therefore did not cover GP referrals or patients going to a GP. This component is being covered in another FSA funded study looking at infectious intestinal disease.

4.4.1 Statistical epidemiology

To investigate putative reporting differences at the level of health board it was necessary to analyse the data at a higher spatial resolution (postcode sector). Maps were generated in ArcView 3.3 for visualisation purposes (Environmental Systems Research Institutes, Inc., http://www.esri.com/software/arcview). In order to determine if the bias was at health board level the average incidence was determined from postcode sectors at the border within the health board and compared with the average of those just outside the border. To check if the difference in incidence was significant a randomisation test involving shuffling of the incidence in postcode sectors (Manly 2007) was performed using 10,000 Monte Carlo iterations. Correction for multiple comparisons was performed by the Bonferroni technique.

4.4.2 Analysis of the microbiological methods

Microbiological approaches to explain the reporting bias between health boards involved expert analysis of the laboratory methods for detection of *Campylobacter*. It was hypothesised that variation in laboratory practice might, in part, be responsible for regional differences. Twenty seven hospitals from 13 health boards were contacted (by email and phone) requesting information on laboratory sampling strategy i.e. whether all submitted human stool samples were tested for *Campylobacter* in their laboratory and on the methods employed for isolation.

It was postulated that differences in reporting rates between health boards might be similar for the different gastrointestinal pathogens. Hence, information from HPS detailing the incidence of the main gastrointestinal bacterial pathogens (*Campylobacter, E. coli O157, Salmonella, Cryptosporidium*) for the 12 Scottish mainland health boards was collated for the years 2002-2006. These were then ranked (health board with highest incidence – rank 1, health board with lowest incidence – rank 12) and compared semi-quantitatively.

4.4.3 Analysis of health board as a risk factor

Health board was included as a putative risk factor in the statistical regression models (see Section 4.4). This allows us to distinguish whether there are significant differences between health boards once other risk factors have been accounted for.

4.5 Risk factor modelling

4.5.1 Overview of risk factors and model development

The data collected that relate to the putative risk factors for *Campylobacter* infections are screened using a Generalised Linear Model (GLM) with a Poisson distributed outcome. Specifically, as the *Campylobacter* data are case rate data the appropriate model is a Poisson regression model. However, one of the key assumptions of statistical analysis is that each data point is independent of every other data point in the dataset. As these data are spatially aggregated (based upon postcode sectors), this may not be the case – two postcode sectors that are close to one another are more likely to have similar rates of *Campylobacter* infection compared to two that are far apart – therefore, these data exhibit spatial dependency. This can be taken into account by employing a class of model called spatial regression model in which the spatial structure of the data – in this case the neighbouring postcode sectors is included as a random effect. Essentially, the model smoothes over the spatial structure of the data. For these analyses the spatial structure was modelled using Gaussian Markov random fields fitted using Integrated Nested Laplace Approximation (Rue et al. 2009).

The initial model was developed to relate the incidence rate of *Campylobacter* infections to the following putative risk factors: deprivation, population density (as an indicator of rurality), the density of cattle, sheep and poultry, the density of private water supplies, and the location of the postcode sector in which the *Campylobacter* case resided. Due to the limited availability of data, the effect of travel could not be included in this analysis.

Specifically, the following risk factors were included in the statistical model for initial screening:

- The Carstairs deprivation score (Carstairs and Morris, 1990).
- The location of the case (which is assumed to be the centre of the associated postcode sector)
- Population density (people / km^2) of the postcode sector.
- Density of cattle, sheep and poultry (head / km^2) in the postcode sector.
- Density of private water supplies (supplies / km^2) in the postcode sector (log_{10} transformed to normalise its distribution).

A further factor to be taken into account is the NHS Health Board of the case. As this relates to the collection of the data rather than a risk factor, health board was accounted for in the model as a random effect. There are 12 health boards in the data, however for these analyses it was treated as a factor with 13 levels. Although there were only 12 health boards in our dataset, the 10-fold difference in case rates between the two regions in Argyll and Clyde (see section 5.4.1 for further details) suggested that it would be appropriate to treat this health board as two separate regions. The 13 regions included in the model (see Fig 3.1 to see their location) were therefore:

- Argyll (Argyll)
- Clyde (Clyde)
- Ayrshire and Arran (AA)
- Borders (BR)
- Dumfries and Galloway (DG)
- Fife (FF)
- Forth Valley (FV)
- Greater Glasgow (GG)
- Grampian (GR)
- Highland (HI)
- Lanarkshire (LN)

- Lothian (LO)
- Tayside (TY)

Animal density data were obtained for cattle, pigs, broilers, ducks, geese, total poultry, sheep and deer. Due to difficulties resulting from either sparsity of the data, or in interpretation of the data (for example very different densities of pigs or poultry might reflect quite different management types and therefore human risks not related straightforwardly to density) only the density for cattle, sheep and total poultry were included as putative risk factors.

The risk factors listed above were screened using a univariate Poisson regression model. Those factors deemed sufficiently significant in the univariate analysis were retained for inclusion in the multivariate model. Specifically, those risk factors with a p-value satisfying p<0.25 were selected for inclusion in the multivariate model.

4.5.2 Statistical model and model selection

The outcome (number of cases - *Y*) was fitted with a Poisson distribution offset by the log of the population of the postcode sector. Thus, the model takes the form:

$$Y_{ij} \sim Poisson(\lambda_{ij})$$

$$\log(\lambda_{ij}) = \beta X_{ij} + H_i + U_{ij} + V_{ij} + \log(O_i)$$

where H_i represents the effect of health board *i*; V_{ij} the spatially structured variation associated with being in postcode sector *j* in health board *i* and U_{ij} the corresponding unstructured variation. X_{ij} represents the matrix of risk factors in each postcode sector in each health board. $\log(O_i)$ is the population adjustment offset. The estimated coefficients and their 95% confidence intervals are transformed to relative risks (RRs) to ease interpretation. A RR whose 95% CIs do not overlap 1 is significant at the 95% confidence interval and the RR represents the proportional change in the outcome with a change of 1 unit in the predictor. INLA was implemented in the INLA package (Martino et al., 2009) for the R statistical environment (R Core Development Team, 2008).

The model with the best explanatory power was selected by removing non-significant (p>0.05) risk factors in turn from the multivariate model starting with the least significant and monitoring the Deviance Information Criteria (DIC) upon each removal, accepting the model with the lower DIC until only significant risk factors remain and the lowest DIC has been found. When removed, the effect of the removal on remaining risk factors was monitored, as a substantial change in the estimate of the associated coefficient would indicate correlation.

To allow for the different case rates in different age-groups (Strachan et al., 2008) and explore potential age-dependent differences in the effect of rurality (Ethelberg et al., 2005), separate models were constructed for those aged under 15 and those aged 15 and over. The differing roles of particular risk factors in these two age groups were assessed by comparing the estimates for the associated coefficients between the two models.

The analyses were repeated for each of the two/three deprivation indices SIMD, Carstairs and DepCat to ascertain robustness of the statistical model.

All analyses were conducted in the R statistical package (R Development Core Team, 2008).

4.6 Spatial analysis

4.6.1 SatScan analysis of the 2000-2006 case dataset

Spatio-temporal clustering was analysed using the scan statistic available in SatScan 8.0.1 (Kulldorff, 2005, http://www.satscan.org). The program uses a likelihood function that is calculated from a cylindrical (space-time) window of variable size and position (Figure 4.9). This basically estimates the ratio of observed to expected number of cases inside the scan window in order to detect clusters that are not likely to appear by chance. This is achieved by calculating the maximum likelihood. The statistical significance for each cluster is obtained by Monte Carlo hypothesis testing, i.e., results of the likelihood function are compared with those obtained from the randomisation of the dataset. In this study, the Poisson probability model was considered for the distribution of cases, the analysis was retrospective and the space-time scan statistic was applied to scan for clusters. A maximum spatial cluster size was set as 0.5% of the population at risk ~25,000 people, to be comparable with the postcode sector with largest population (21,700 people). A summary of the parameters used in the simulation is given in Table 4.7.

All 34173 geocoded cases in the 2000-2006 study period were included for analysis. The detected clusters were and their distribution across Scotland was mapped. The SatScan output file includes descriptors that characterize each individual cluster (position, time frame, number of cases observed, number of cases expected and statistical significance).

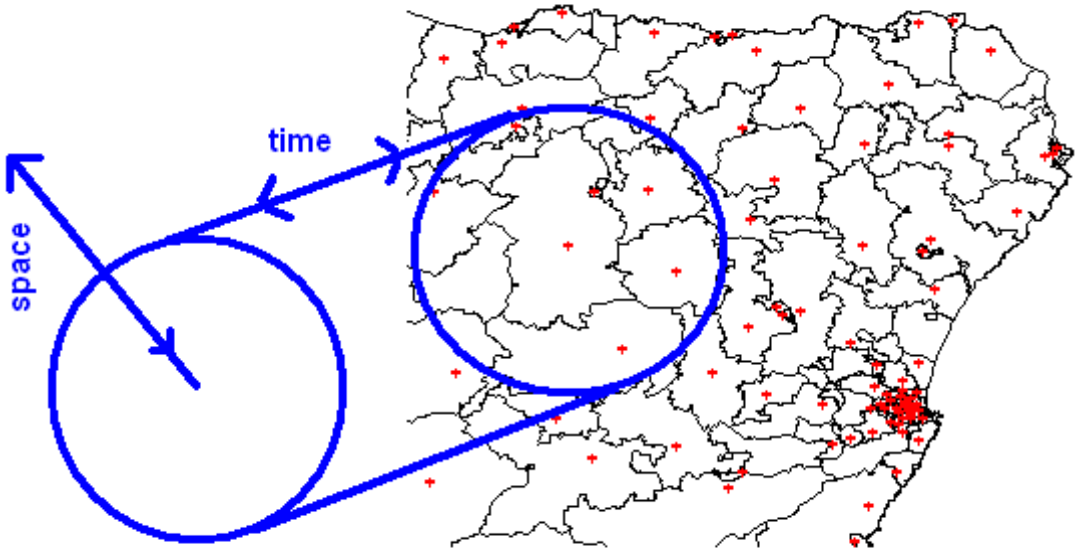


Figure 4.9. The space-time cylinder used by SatScan to search for clusters.

Table 4.7 Parameter settings used in SatScan

Input		
	Time Precision	Day
	Start Date	2000/1/5
	End Date	2006/12/29
	Coordinate	Cartesian
Analysis		
	Type of Analysis	Retrospective Space-Time
	Probability Model	Poisson
	Scan for Areas with	High Rates
	Time Aggregation Units	Day
	Time Aggregation Length	1
	Number of Replications	999
Spatial Window		
	Maximum Spatial Cluster Size	0.5% of population at risk
	Include Purely Temporal Clusters	No
	Window Shape	Circular
Temporal Window		
	Maximum Temporal Cluster Size	60 Days
	Include Purely Spatial Clusters	No
Clusters Reported		
	Criteria for Reporting Secondary Clusters	No Geographical Overlap

4.6.2 K function analysis

K function analysis is a tool for analysing the distribution of cases in space, time, or space-time, relative to a control population (Ripley, 1976; Diggle et al., 1995). It provides a means of quantifying the clustering of cases in space, time, or space and time and differs conceptually from SatScan statistic in the following way: it does not identify specific clusters, but quantifies the tendency of cases to cluster relative to the controls. Here, the control population is the Scottish population. Specifically, if the cases are randomly distributed in space or time, relative to the underlying control populations, the expected value of the k-function is 1.

The technique of K function analysis requires that both cases and controls must be assigned discrete points in space and time. The *Campylobacter* infection data available to this project do not have discrete locations, but are at the level of the postcode sector. To handle this problem, points were assigned some random location within their postcode sector using the maptools package (Lewin-Koh and Bivand, 2009) in the R statistical environment (R Development Core Team, 2008).

Two types of analysis were conducted. In the first, the control population was sampled from the overall population according to the 2000 population estimate; a number of controls equal to the number of cases were sampled and assigned random locations within the postcode sector and a date sampled from the period of study. In the second, the control population was sampled from the distribution of cases predicted by the statistical model. This approach removes spatial trends that result from the spatial localisation of risk factors, allowing better

discrimination between sporadic and outbreak cases.

The K function was calculated for both cases and controls and the ratio gave a measure of the relative spatial association. This process is repeated for 100 iterations to allow for the randomised locations of points and the sampling of controls and the mean of 100 iterations was used. The K function was calculated in the SPLANCS package for R (Rowlingson and Diggle, 2008).

The analyses were conducted for all ruminant and all poultry assigned STs and the three most common individual STs in the human cases: ST21, ST45 and ST257.

4.7 Incorporation of MLST data

4.7.1 Attribution modelling

Attribution by microbial sub-typing is a relatively new area of research. The term "source attribution" has been defined (Pires et al, 2009) as: "…the partitioning of the human disease burden of one or more foodborne infections to specific source, where the term *source* includes animal reservoirs and vehicles (e.g. foods)."

Further, the microbial subtyping methodology, uses the distribution of subtypes in each of the sources and compares this with that found in humans. This can be done in terms of simple proportions (e.g. the Dutch model) or using Bayesian stochastic methods (e.g. STRUCTURE). Currently, there are 5 main techniques for attributing disease on a population level using microbial sub-typing. Three of these methods are used in the current study and these are detailed below.

4.7.1.1 Proportional Similarity Index (PS)

The proportional similarity index is an estimate of the area of intersection between two frequency distributions (Rosef et al, 1985). When comparing the MLST types of *Campylobacter* found in humans and animal reservoirs, it uses the absolute difference of genotype frequencies found in one particular reservoir and humans ($|p_i - q_i|$) to calculate the fraction of similarity of the two frequency distributions, as follows

$$PS = 1 - 0.5\sum_i |p_i - q_i|$$

where p_i and q_i are the frequencies of genotype *i* in a reservoir (e.g. poultry) and humans, respectively. When the frequency distributions are identical PS=1, and PS =0 when the distributions have no common types.

The PS index gives a direct comparison of the similarity in the frequency of sequence types found in a pair of hosts (e.g. in humans and poultry, or humans and cattle, or cattle and sheep), but does not give an estimation of the attribution to each host.

4.7.1.2 The Dutch Model

The Dutch Model (French, 2008) is a straight forward way to estimate the attribution of a particular genotype (e.g. ST) to a reservoir, when the frequency distribution of each type is known for each reservoir. If p_{ij} represents the frequency of type *i* (eg ST 19) in source *j* (e.g. poultry) then the proportion of attribution of type *i* in source *j* is given by

$$\lambda_{ij} = \frac{p_{ij}}{\sum_{j} p_{ij}}$$

where the summation by *j* considers all the reservoirs where data exist (e.g. cattle, sheep, wild birds, poultry etc.).

When applied at ST level this model does not guarantee that all STs will be attributed to sources. This is because human types that are not found in the animal reservoir cannot be attributed.

However, if genetic information exists at multiple loci (e.g. 7 MLST loci for this study), then the Dutch Model can make use of the frequency of each individual allele at each individual locus, and estimate attribution even for STs that are not present in the animal reservoirs.

In particular, at allele level the frequencies $p_{a_{ijk}}$ can be calculated for each allele a_{ijk} of all isolates from the animal reservoirs. Where i is subtype, j source and k the loci number (see Table 4.8 below).

Table 4.8 Notation for the Dutch model at allele level.

Loci1	Loci2	Loci3	Loci4	Loci5	Loci6	Loci7	MLST type
a_{1j1}	a_{1j2}					a_{1j7}	1
a_{ij1}	a_{ij2}					a_{ij7}	i

The attribution score of bacterial subtype *i* in source *j* is

$$\lambda_{ij} = \frac{\prod_{k=1}^{7} p_{a_{ijk}}}{\sum_{j}\left(\prod_{k=1}^{7} p_{a_{ijk}}\right)}$$

where $p_{a_{ijk}} = BetaInv(0.5, 0+1, N_{isolates}+1)$ if its frequency is zero (*BetaInv* fn in Excel). This assumes that we have no prior knowledge of $p_{a_{ijk}}$ and so is maximally noncommittal or conservative.

The Dutch Model does not take into account the uncertainty in the frequency distribution of genotypes. It does not consider any information about the exposure of humans to sources or the viability/virulence of pathogens once they are ingested by humans.

4.7.1.3 Population STRUCTURE

STRUCTURE (Pritchard et al, 2000) is a Bayesian clustering model designed to infer population structure and to attribute individuals to population groups. The program can use MLST genotyping data. Each isolate is attributed on the basis of a training dataset consisting of isolates from known populations (i.e. set USEPOPINFO to 1). The algorithm calculates the frequency of each particular sequence type in each population taking into account the uncertainty due to the sample size. Based on these frequencies the probability to belong to a population group/reservoir is calculated, following multiple iterative steps (MARKOV CHAIN

MONTE CARLO- MCMC) for the estimation of frequencies. The program has the option to consider the allele independent (no-admixture model – independent alleles) and starts with equal frequencies for each isolate type. Following an initial number of MCMC burn-in steps (e.g. 1000) further iterations (e.g. 10000) are used for estimation of the probabilities that an isolate belongs to each particular population being considered (eg cattle, sheep, poultry etc.). To enable the largest reference dataset to be used (often datasets are small due to the cost of typing many isolates) only one ST is selected at a time from the unknown dataset by using the jacknife method. This process is repeated to enable multiple estimations of the same sequence type so that uncertainty in the attribution scores can be determined.

STRUCTURE can be used at ST or allele level, it incorporates uncertainty and takes account of sample size. Hence, in principal it gives a more realistic estimation of the attribution to a specific reservoir than the Dutch Model. Also, like the Dutch Model at allele level it can assign human cases that have STs that are not found in the animal reservoirs. However it is highly time consuming and does not consider any exposure to risk factors or the viability of pathogens.

4.7.1.4 Application of Models

The proportional similarity index was used to compare the MLST types found in human isolates with those found in source reservoirs. This was carried out independently for *C. jejuni* and *C. coli*. This was done in Visual Basic Application under Microsoft Excel. Bootstrapped confidence intervals were obtained by 10,000 Monte Carlo iterations (Manly, 2007).

Analysis of source reservoirs was carried out using the DUTCH model and STRUCTURE at the level of ST and allele and applied to *C. jejuni* and *C. coli* independently. The source reservoirs for *C. jejuni* were cattle, retail poultry, sheep and wild birds, and for *C. coli* were cattle, pigs, retail poultry and sheep.

To determine whether the attribution methods could differentiate between different source reservoirs self attribution was carried out. This involved using a jacknife method to predict the source of an isolate that was unknown to the model but known to the user. This was repeated many times for each of the isolates in the animal dataset so that an average value and confidence intervals could be obtained for the accuracy of the self attribution. Overall, 10,000 iterations were carried out.

The next step involved the assignment of human cases to source reservoir. First, the probability of a human isolate originating from each particular animal source was estimated using the models. This was repeated 10,000 times for all of the human isolates to determine the average probability assignment to each source and the corresponding 95% confidence intervals (CIs).

4.7.2 K function modelling of MLST data

The K function analysis was conducted as for the full dataset, but with the analysis restricted to the cases for which MLST typing was available in the following ways. First, the analysis was restricted by dividing the typed cases into those with poultry and those with ruminant associated STs. Second, the analyses were restricted to the three most common STs in the human cases (257, 21 and 45).

These restricted datasets were used to compare the outputs of the risk factor model for poultry and ruminant strains, to use K function analysis to compare the space-time distribution of poultry and ruminant strains and of individual STs, and to conduct a case-case

analysis of the risks associated with poultry versus ruminant strains.

4.7.3 Urban – rural

4.7.3.1 Incidence ratios

The postcodes sectors in Scotland were designated as either rural (< 200 people/km^2) or urban (>200 people/km^2). Rural to urban incidence ratios stratified by age (0-14+, 15-64+ and >=65 age groups) were calculated for the 2000-06 dataset as well as for the 15 months of the MLST study. Source reservoir attribution was performed for rural and urban human cases independently. This was also stratified by age. The attribution method used was STRUCTURE at allele level.

4.7.3.2 Population density as a risk factor

Population density was included in the statistical regression models as a putative risk factor enabling us to explore the role of rural versus urban living on the risk of *Campylobacter* infection.

4.7.4 SatScan analysis of the 15 months MLST dataset

In order to detect small spatio-temporal distinct outbreaks the SatScan space-time Poisson model was applied to the MLST typed Scottish data obtained during July 2005-September 2006. The scan for clusters was performed for each individual sequence type that had more than two cases, except for doublets separated by more than 60 days. A total of 208 runs were carried out. The parameter settings were identical with those used for the 2000-2006 dataset, as previously described. The detected clusters were saved and plotted using ArcView 3.3. Also, the spread of the clusters during the study period was determined.

4.7.5 Foreign Travel

4.7.5.1 Travel/deprivation as a risk factor

Travel abroad is a potential risk factor for *Campylobacter* infection and may also act as a proxy for deprivation. As only a proportion of the health boards supplied travel data, the data were not sufficient to quantify the potential risk of infection due to travel. However, the data could be used to assess whether our conclusions about the impact of deprivation were being biased by a potential higher rate of travel abroad in less deprived areas. The risk factor modelling was repeated for the Grampian, Lothian and Forth Valley health boards, with and without the cases reporting recent travel abroad.

4.7.5.2 Diversity of MLST types

Data: Linkage of human clinical cases (2000-06) with travel data was carried out. However, as most of the health boards provided only partial linkage of data it was decided to concentrate on the Grampian case-control dataset where linkage to both MLST and travel data had been carried out in a consistent manner.

The Grampian case-control dataset (15 months from 15/7/2005 to 9/11/2006) contained 763 cases with reporting dates from and linked to MLST type. Of these 373 had information supplied indicating whether the individual had travelled in the previous 14 days. Foreign travel was denoted as travel outwith the United Kingdom.

Diversity: Simpsons diversity index was used to determine whether travel associated cases were more diverse than indigenous cases. Briefly, if there is a total of N isolates and if there are n_i of sequence type *i*.

$$Simpsons\ Index = 1 - \sum_i \left(\frac{n_i}{N} \right)^2$$

This will have value 0 when all of the isolates belong to a single ST (i.e. no diversity) and will have a higher value when isolates are spread between different STs.
A web based implementation was used called VDICE accessed at www.hpa-bioinformatics.org.uk/cgi-bin/DICI/DICI.pl.

Common /uncommon types: To determine whether the MLST types of foreign travel cases were more common either at home or abroad odds ratios were calculated and statistical significance was determined by Fisher's Exact Test.

4.7.6 Household outbreaks

The aims of this section are to identify and analyse from an epidemiological point of view household outbreaks in Scotland and to evaluate the utility of MLST as a means of confirming these outbreaks.

Household outbreaks are defined as two or more cases reported within 28 days at the same address with the same family surname. These were identified by searching the database of cases for the 15 month (July 2005 – September 2006) Scottish dataset that also contained a portion of isolates that were typed by MLST. Randomisation tests were performed (Manly 2007) to determine whether this number of outbreaks was more than would be expected by chance from the 2.29 million households in Scotland (http://www.gro-scotland.gov.uk/statistics/publications-and-data/household-estimates-statistics/estimates-of-households-and-dwellings-in-scotland-2007/index.html).

5. RESULTS

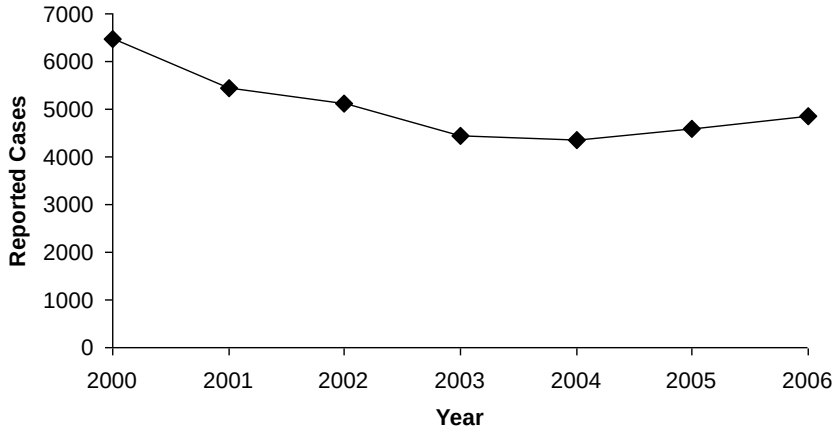
5.1 Descriptive statistics

This section provides the basic descriptive epidemiological statistics of all laboratory confirmed human *Campylobacter* cases obtained from the Public Health Teams in the NHS Boards. During 2000-2006, 34,983 *Campylobacter* cases were reported across the mainland health boards of Scotland. All of these were assigned to health board and 34,234 to postal sector.

5.1.1 National Campylobacter rates

Fig 5.1 shows both the number of cases and incidence of *Campylobacter* in Scotland during the years 2000-2006. The population of the mainland health boards (4,994,298 people) was obtained from the Scottish census (General Register Office for Scotland, http://www.gro-scotland.gov.uk/files/00sap5.xls) and this was used as denominator data for estimating incidence.

(a)



(b)

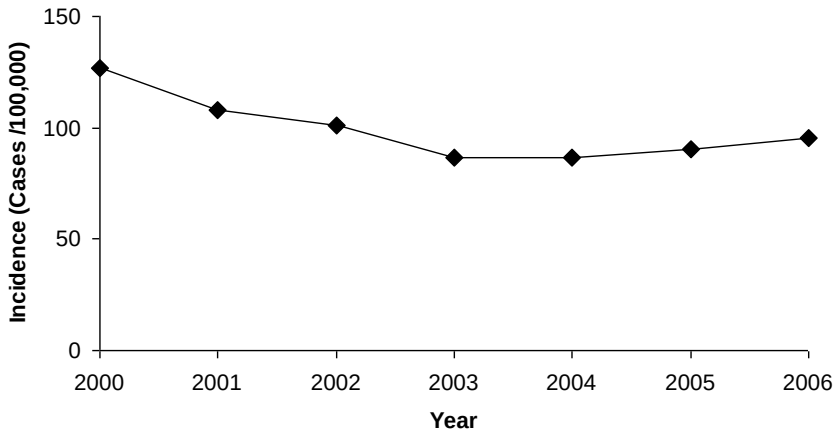


Figure 5.1 (a) Reported cases and (b) incidence of human campylobacteriosis in Scotland 2000-2006.

The average annual incidence is 99.2 (95% CI ± 13.5) cases per 100,000 per year. The

incidence was highest in 2000 and fell to a minimum in 2003 followed by a gradual increase.

5.1.2 Age distribution of cases

The average incidence over the seven year period was stratified by age group. Figure 5.2 shows *Campylobacter* incidence peaks in the 1-2 year olds and then falls and remains low between the ages of 5-15 years before it rises again in young adults. Over 65 the incidence of infection falls.

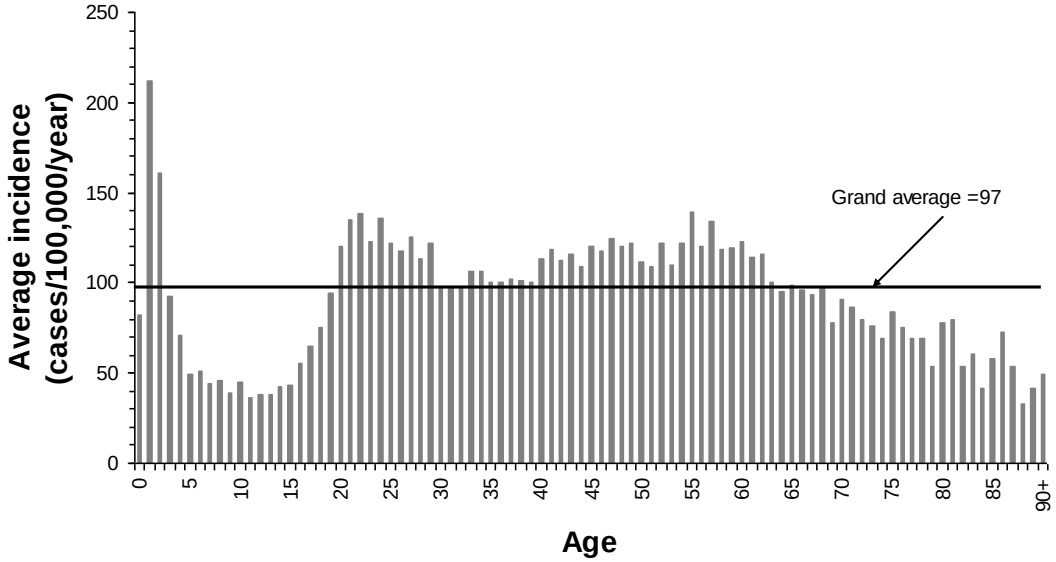


Figure 5.2 The average incidence of *Campylobacter* cases in Scotland (2000-2006) (Note: grand average is less than in the previous section because approximately 1.2% of cases had missing age information and this average is based on all seven years of cases not an average across years).

5.1.3 Gender distribution of cases

Figure 5.3 shows the yearly incidence of cases by gender. It can be observed over the seven year period that the number of male cases consistently exceeds that of female cases and that this is significant ($P<0.0001$) by one-sided paired t-test. Overall, there is an excess (approx 12 %) of male cases as has been reported previously in the literature (Gillespie et al, 2008).

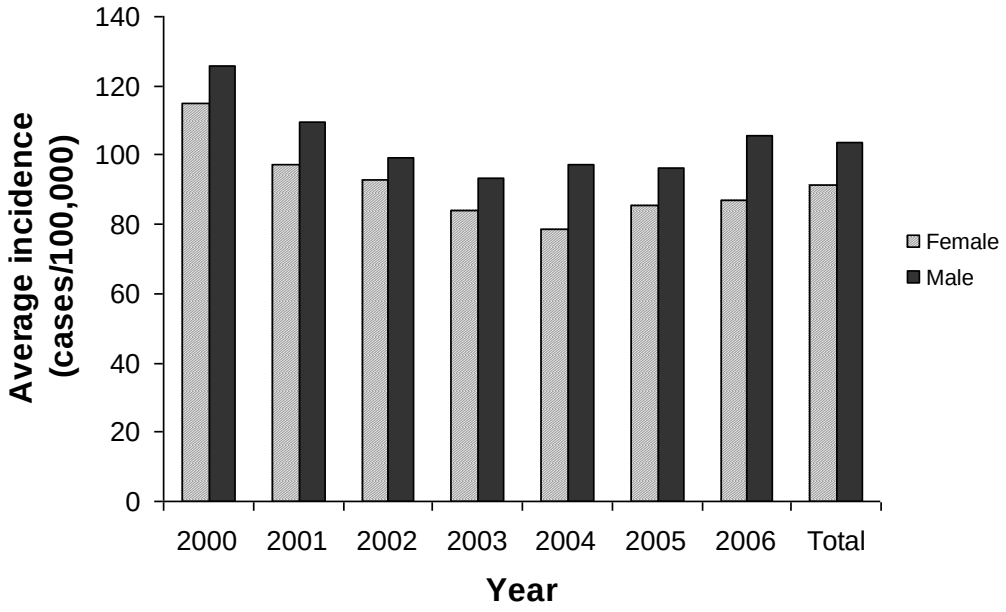
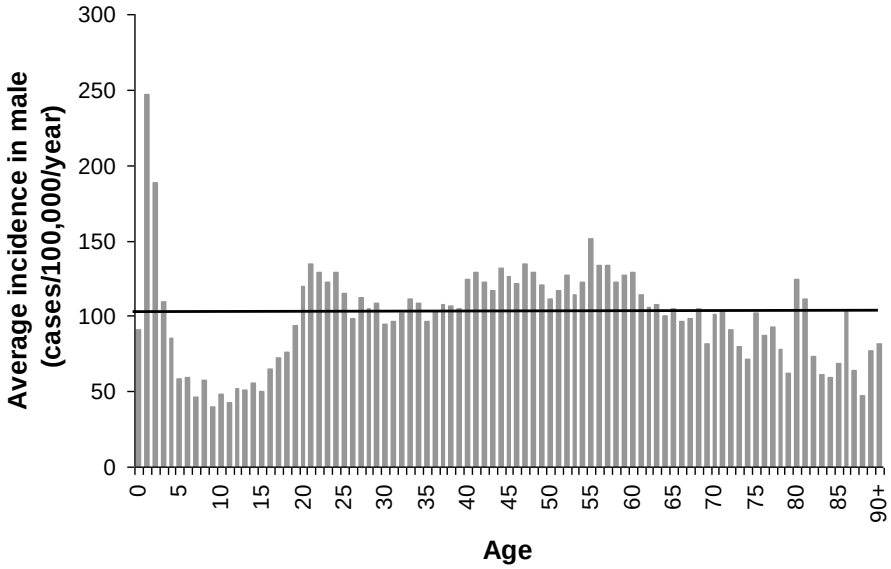


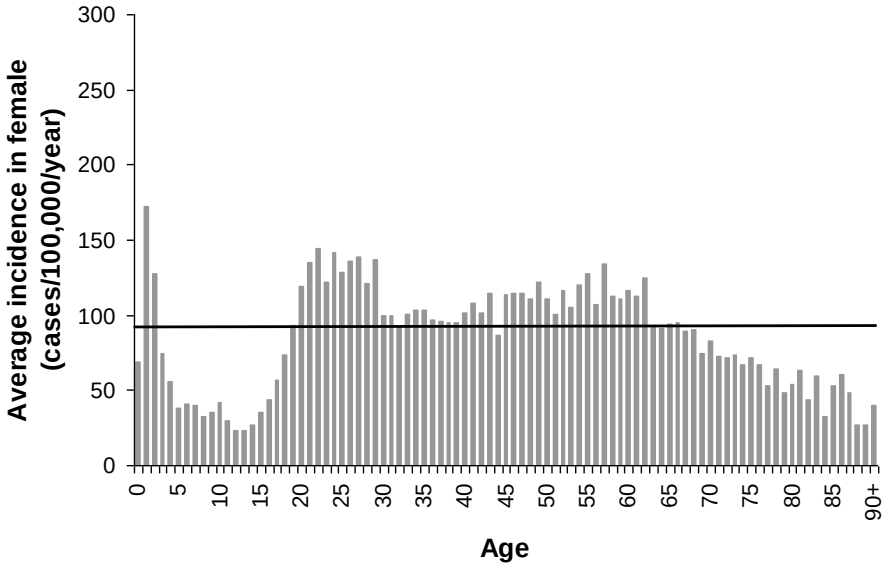
Figure 5.3 The average incidence of *Campylobacter* cases in Scotland, stratified by gender and year (2000-2006).

When the average incidence of cases over the seven year period was stratified by single year age groups (Fig. 5.4) for each gender it was shown that males had higher rates of infection from birth till approximately 18 years of age and also from retirement age onwards (>65 years).

(a)



(b)



(c)

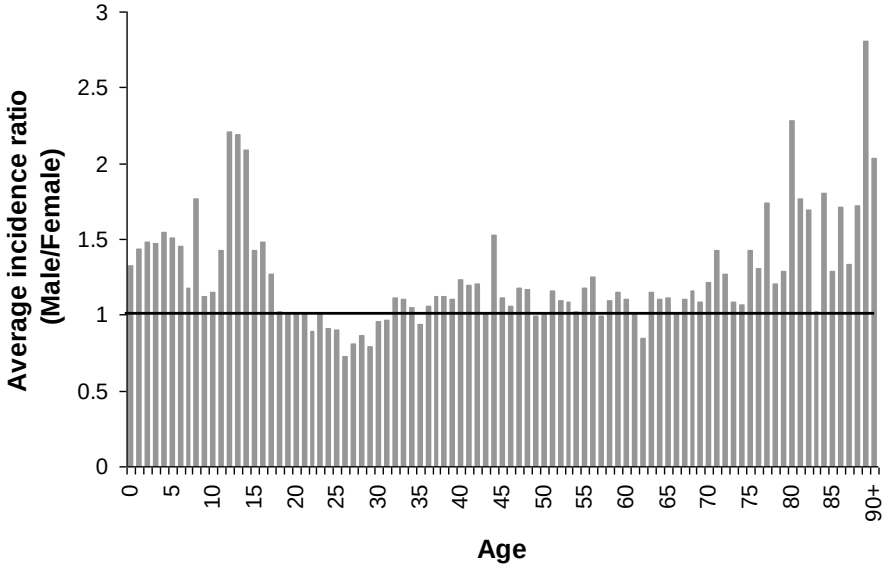


Figure 5.4 The average incidence in (a) males, (b) females and (c) male:female ratio by age.

5.1.4 The spatial description of *Campylobacter* cases across Scotland.

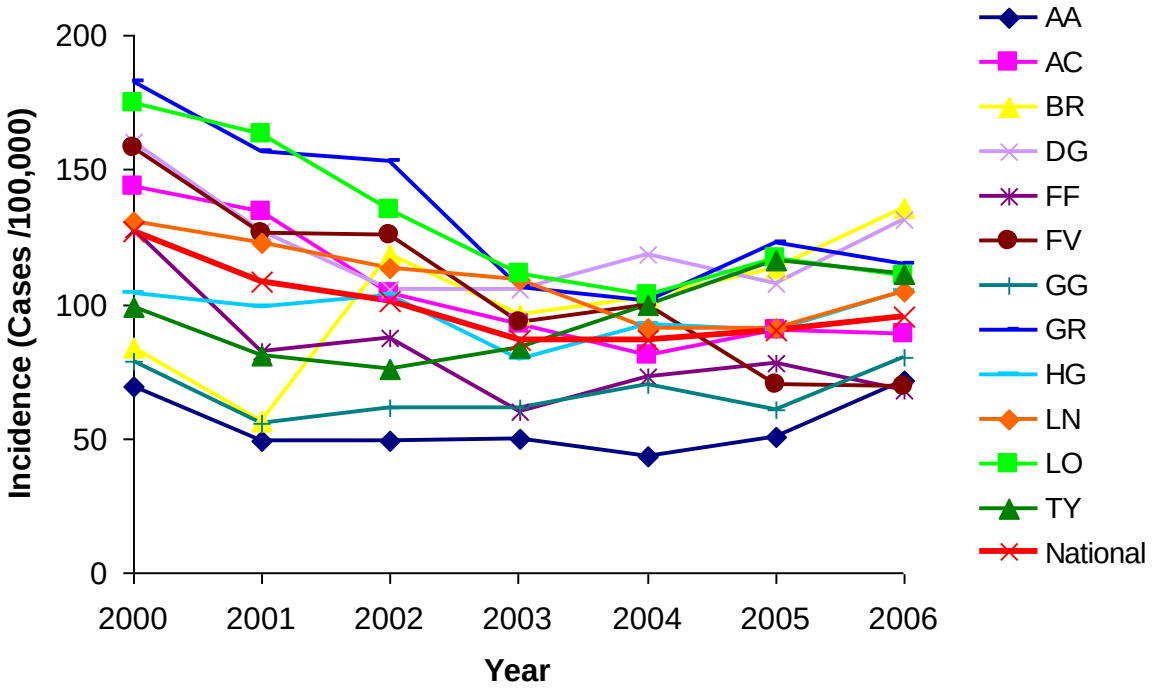


Figure 5.5 Yearly incidence of campylobacteriosis by mainland health board in Scotland (2000-2006).

The incidence fell throughout the majority of health boards from 2000-2003 and began to rise slowly in some health boards up till 2006 (Fig. 5.5). The figure also shows that the incidence in Ayrshire & Arran was low throughout the period whereas Grampian and Lothian had the highest reporting rates. The average incidence varied between health boards from 54.6 in Ayrshire and Arran to 134.0 in Grampian (Fig 5.6 and Table 5.1).

Table 5.1 Summary statistics for the health boards

Health board	Average incidence 2000-2006
Ayrshire & Arran (AA)	54.6
Argyll & Clyde (AC)	105.0
Borders (BR)	100.9
Dumfries & Galloway (DG)	122.1
Fife (FF)	82.2
Forth Valley (FV)	105.9
Greater Glasgow (GG)	66.9
Grampian (GR)	134.0
Highland (HI)	96.2
Lanarkshire (LA)	108.9
Lothian (LO)	130.7
Tayside (TY)	95.3

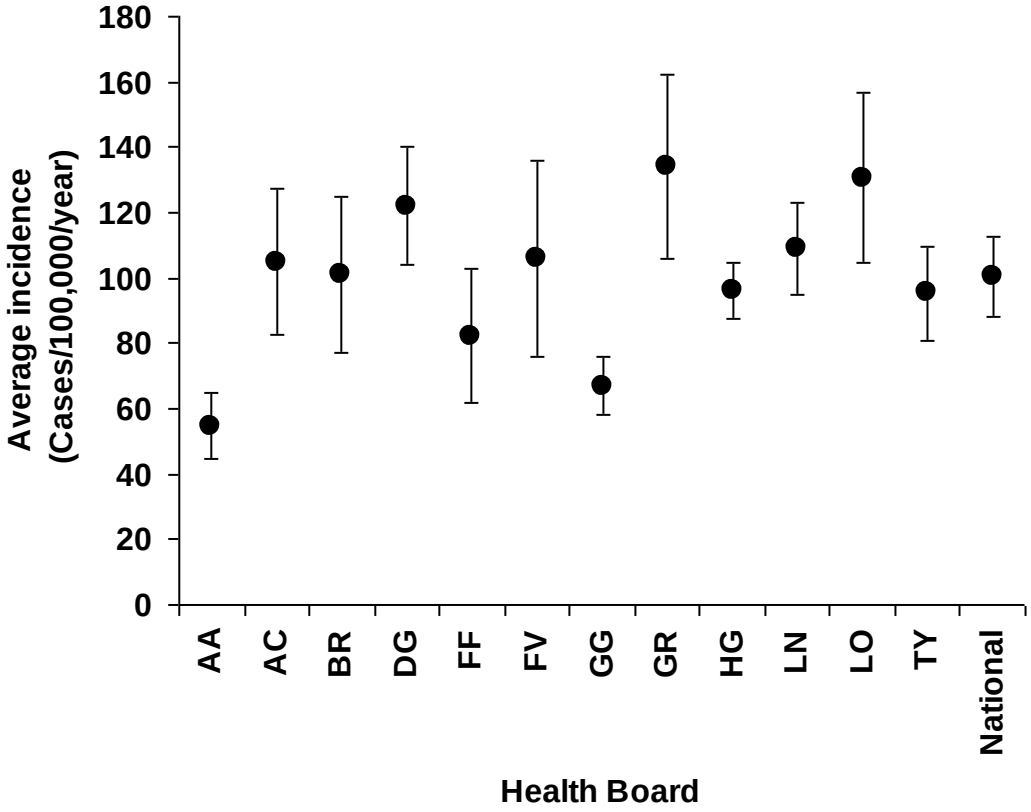


Figure 5.6 The average incidence of *Campylobacter* infections by health board (2000-06). (Error bars represent 95% CI's).

An analysis of variance was performed to determine whether reporting rates between each of the mainland health boards differed (Table 5.2). Ayrshire and Arran has a significantly lower reporting rate of *Campylobacter* than all of the other health boards except Fife and Greater Glasgow. Further to that Greater Glasgow is lower than Grampian, Lothian, Dumfries & Galloway and Lanarkshire, whilst Fife is lower than Grampian, Lothian and Dumfries & Galloway. These data provide evidence to demonstrate that there are differences in reporting rates at health board level. However, the analysis provides no explanation as to why this is the case.

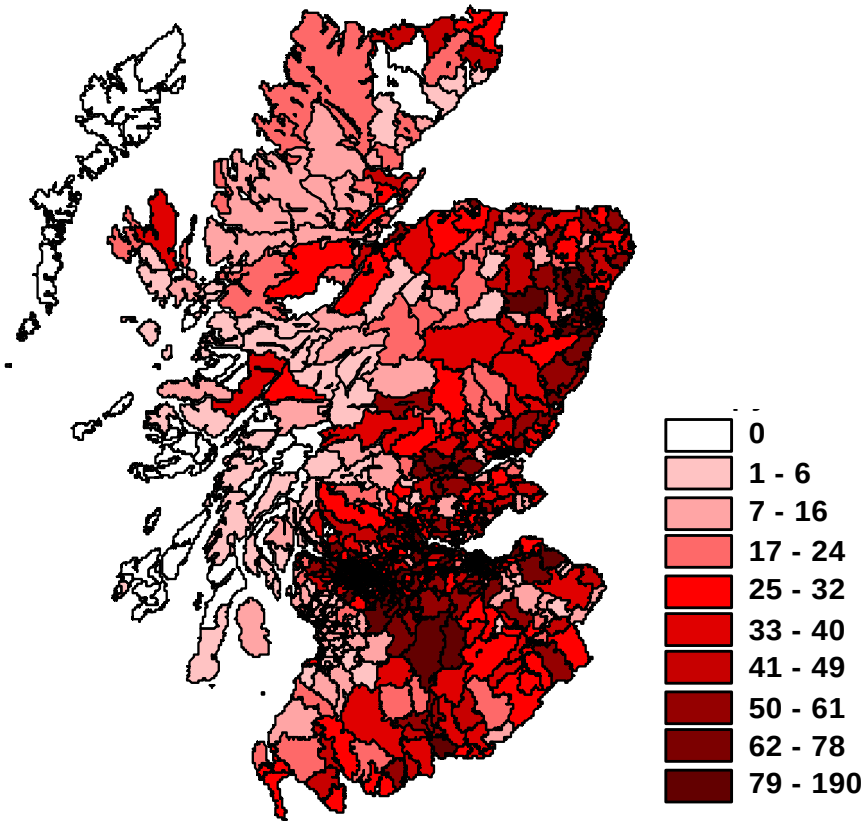
Table 5.2. Differences in reporting rates between health boards[a] by Analysis of Variance using Tukey's honest significant difference

Contrast	Difference in incidence between health boards	Probability of seeing observed difference or greater
GR vs AA	79.31	< 0.0001
LO vs AA	76.01	< 0.0001
DG vs AA	67.48	< 0.0001
GR vs GG	67.10	< 0.0001
LO vs GG	63.80	< 0.0001
DG vs GG	55.27	0.0006
LN vs AA	54.28	0.0008
GR vs FF	51.74	0.0017
FV vs AA	51.25	0.0020
AC vs AA	50.34	0.0026
LO vs FF	48.44	0.0046
BR vs AA	46.25	0.0085
LN vs GG	42.07	0.0258
HG vs AA	41.55	0.0294
TY vs AA	40.61	0.0370
DG vs FF	39.91	0.04382

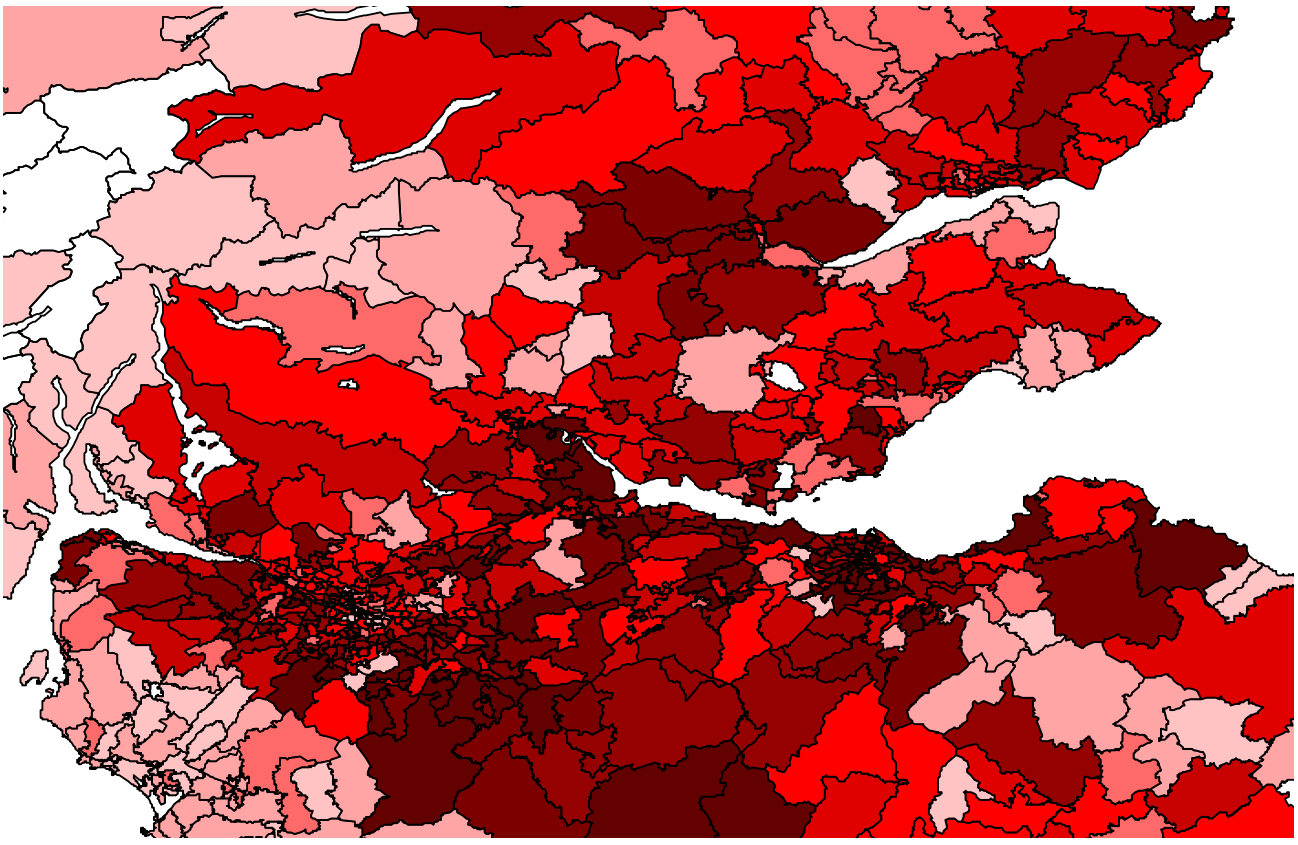
[a] Statistically significant comparisons displayed only

A greater number of reported cases (Fig. 5.7(a,b)) occur in areas with a denser population e.g. the central belt and along the east coast. However, when the incidence per 100,000 population is mapped (Fig. 5.7(c,d)) higher rates can be seen throughout some regions e.g. Grampian and lower rates in others e.g. Ayrshire & Arran and Argyll & Clyde.

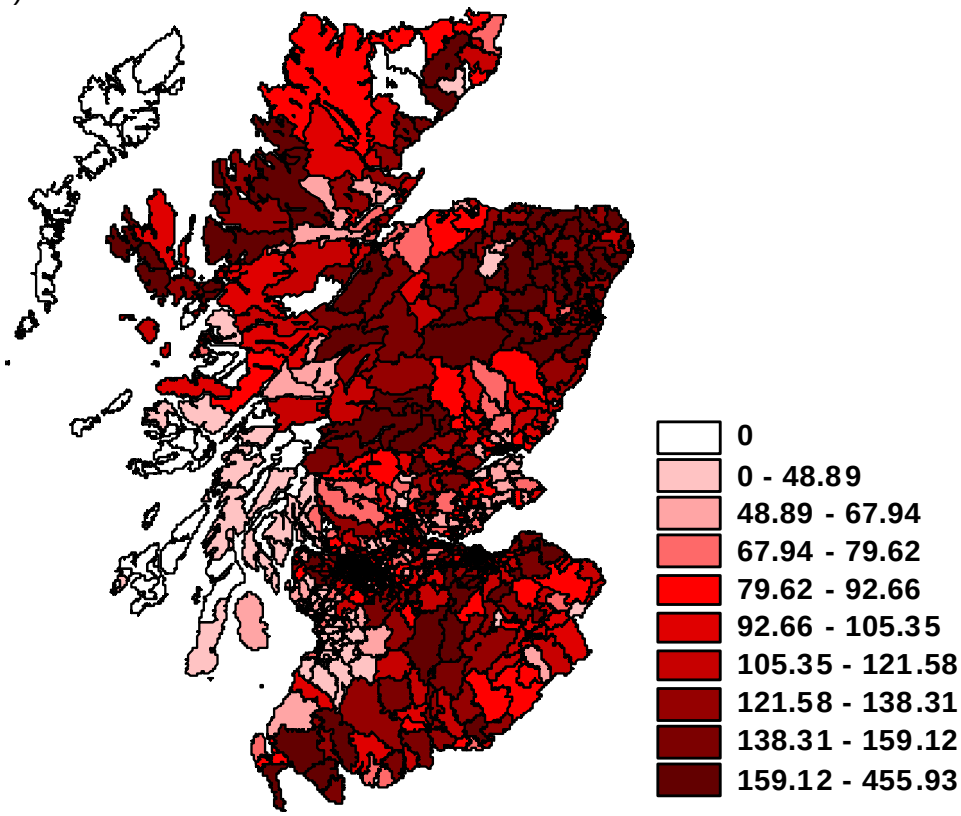
(a)



(b)



(c)



(d)

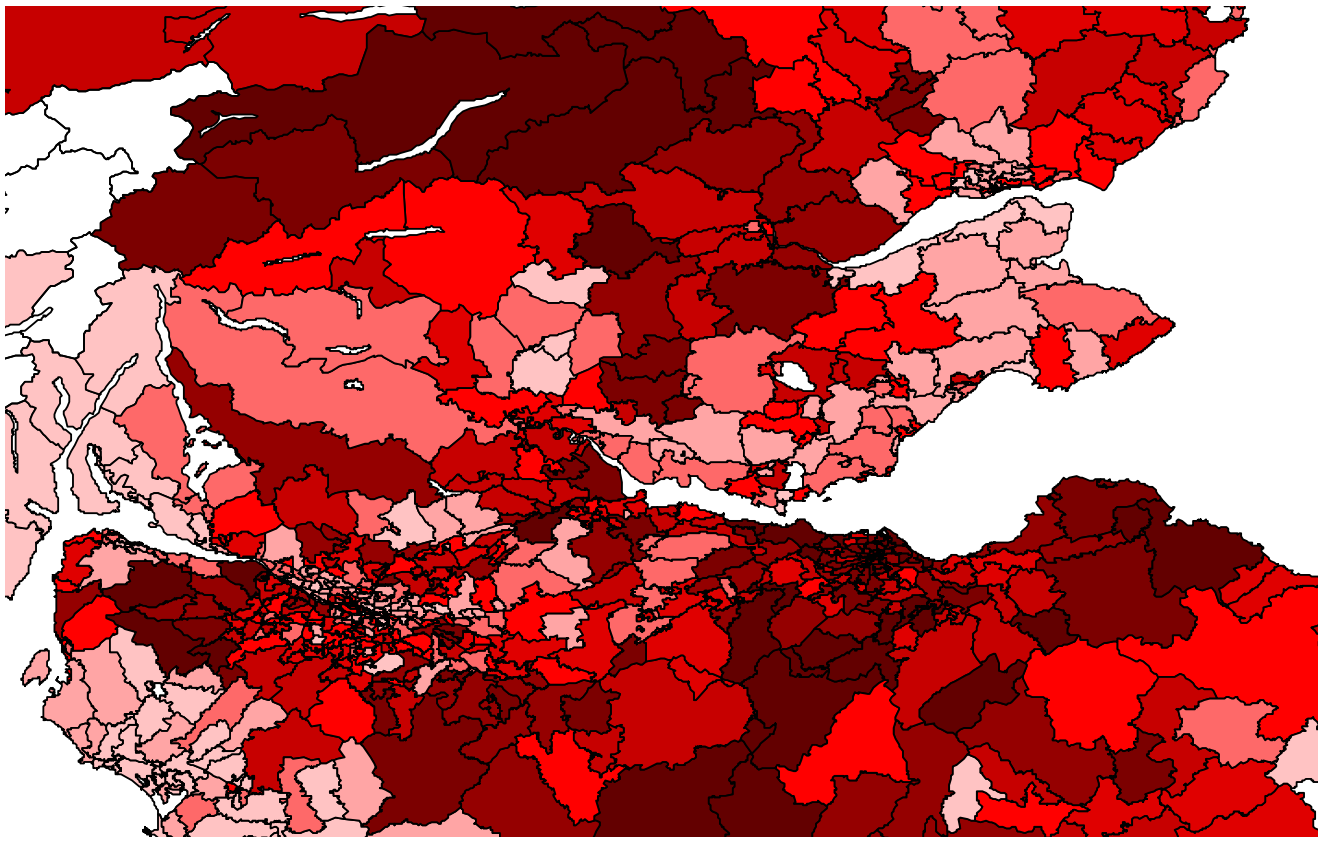


Figure 5.7 (a), (b) Reported cases and (c), (d) average incidence of human campylobacteriosis across Scotland and central belt at postcode sector level (2000-2006). (Note no data presented for Western Isles because of the very small numbers of cases).

5.2. Spatial distribution and clustering of human cases

SatScan analysis of the 2000-2006 case dataset

SatScan was undertaken to determine whether the human case data was clustered within space and time, An example of output obtained from SatScan is given in Figure A.2. The distribution of the clusters across Scotland is presented in Figure 5.8. There were 27 clusters detected, involving 315 cases, which represent 0.9 % of the total number of cases analysed. SatScan detected part of the Montrose outbreak. This outbreak, associated with chicken liver pate affected approximately half of 165 people at a farm dance of which there were 32 microbiologically confirmed cases (Forbes et al., 2009). SatScan detected a total of 18 cases in three different postcode sectors (DD96, DD97, DD109).

The average cluster size was 12 cases/cluster (min. 4 cases/cluster, max. 21 cases/cluster). The sensitivity of the method does not allow the detection of smaller clusters (< 4 cases/cluster – putative small household outbreaks) because the spatial resolution at postcode sector level is too coarse.

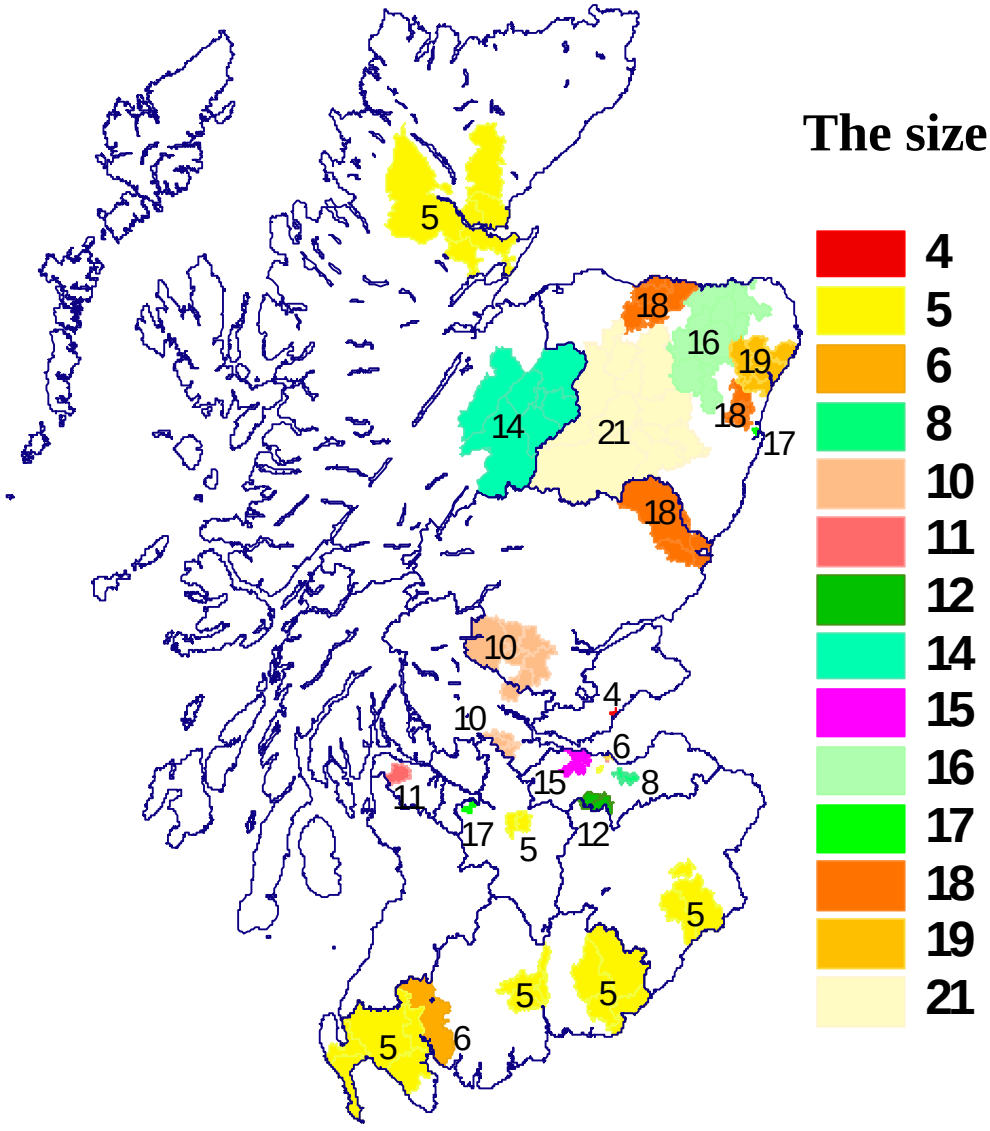


Figure 5.8. The geographical distribution of clusters detected by SatScan in space-time analysis mode. The numbers and colours represent the size of the clusters.

SatScan analysis of the 15 month MLST dataset

The spatiotemporal SatScan analysis was redone but for the 15 month MLST dataset using the typing information. This was run by each ST where there were 2 or more isolates within a 60 day period. SatScan space-time Poisson model detected 33 putative outbreaks containing a total of 85 cases (2.3%) and cluster sizes varied between 2 to 7 cases/cluster (Fig. 5.9) and Table 5.3. The spatial distribution of the clusters shows that the majority are distributed in the central belt and eastern and northeastern coasts of Scotland. However, the sequence types of the clusters are spread heterogeneously with no preferential location for specific STs. Also, common sequence types such as ST257 and ST61 were not detected in clusters which might be a consequence of the fact that the SatScan space-time routine does not detect diffuse spatial clusters.

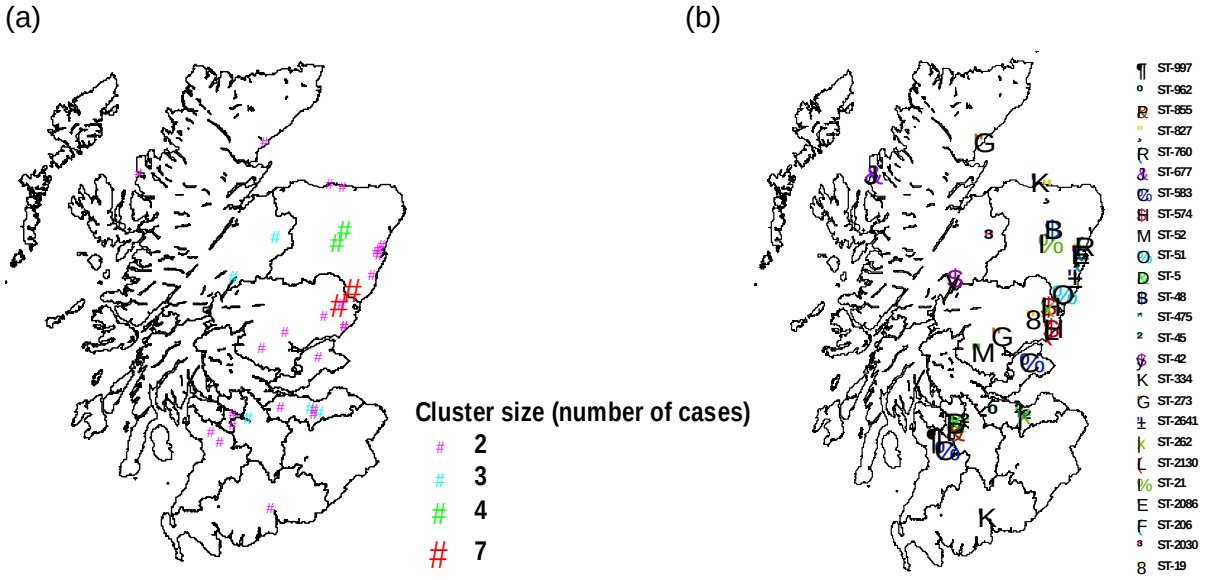


Figure 5.9 The distribution and position of the putative outbreaks detected by SatScan: (a) cluster size and (b) sequence types in clusters.

Table 5.3 The distribution of the clusters by MLST sequence type. The coordinates represent the centroid of the postcode sector in which the cluster belongs.

Sequence Type	Easting	Northing	Number of cases in cluster
ST-5	262133	661035	2
ST-19	344131	749465	2
ST-21	356674	815646	4
ST-42	263390	784704	3
ST-45	331880	668581	3
ST-45	276457	659365	3
ST-45	339736	663736	3
ST-48	363490	827764	4
ST-51	370614	772220	7
ST-52	287740	721559	2
ST-206	392091	803794	2
ST-262	334465	667151	2
ST-262	359227	760378	2
ST-273	309057	735349	2
ST-273	290146	904277	2
ST-334	348726	867460	2
ST-334	295182	576904	2
ST-475	334280	661549	2
ST-574	359227	760378	7
ST-574	361835	740995	2
ST-583	249975	636478	2
ST-583	337772	713105	2
ST-677	178206	877252	2
ST-760	394740	813379	2
ST-827	360775	864453	2
ST-827	392243	808433	2
ST-855	262071	652383	2
ST-962	305029	667330	2
ST-997	242359	646109	2
ST-2030	300700	820049	3
ST-2086	390265	806291	2
ST-2130	361835	740995	2
ST-2641	386457	785729	2

Part of the Montrose outbreak was detected, but only 16 out of 32 cases were contained within clusters (Table 5.4).

Table 5.4 ST types sequenced and clusters detected by SatScan from the Montrose outbreak

ST	Isolates in outbreak	Isolates clustered b y SatScan
574	14	7
51	9	7
262	5	2
257	4	0

The distribution of the clusters during the MLST study period, together with a 3 week moving average of the actual number of cases is presented in Figure 5.10. The majority of the clusters appear to be concentrated during the summer peaks, but there is an unexplained excess of clusters at the beginning of the 2006 peak.

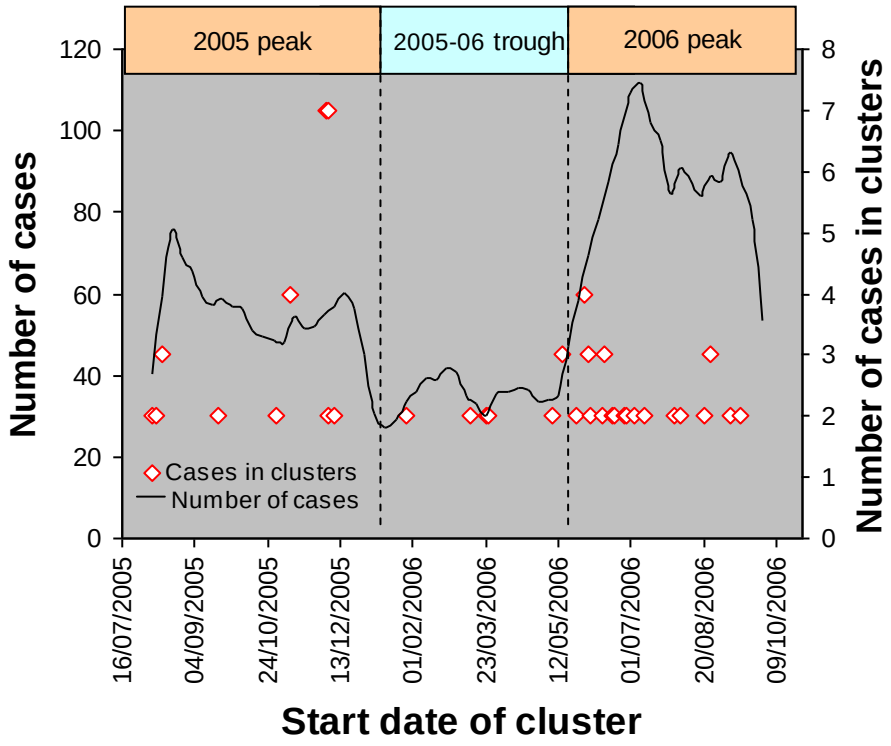


Figure 5.10 Temporal distribution of SatScan space-time clusters (July 05-September 06).

5.3 Statistical modelling

5.3.1 Reporting bias at health board level

Summary statistics for the rates of cases per year reported in each health board are given above in Table 5.1.

5.3.1.1 Statistical epidemiology

Preliminary descriptive analyses showed the Argyll and Clyde and Ayrshire and Arran health boards to have a substantially lower case rate than the other health boards, and that in the former case this low reporting rate was found in the Argyll region of the health board (Fig. 5.11)

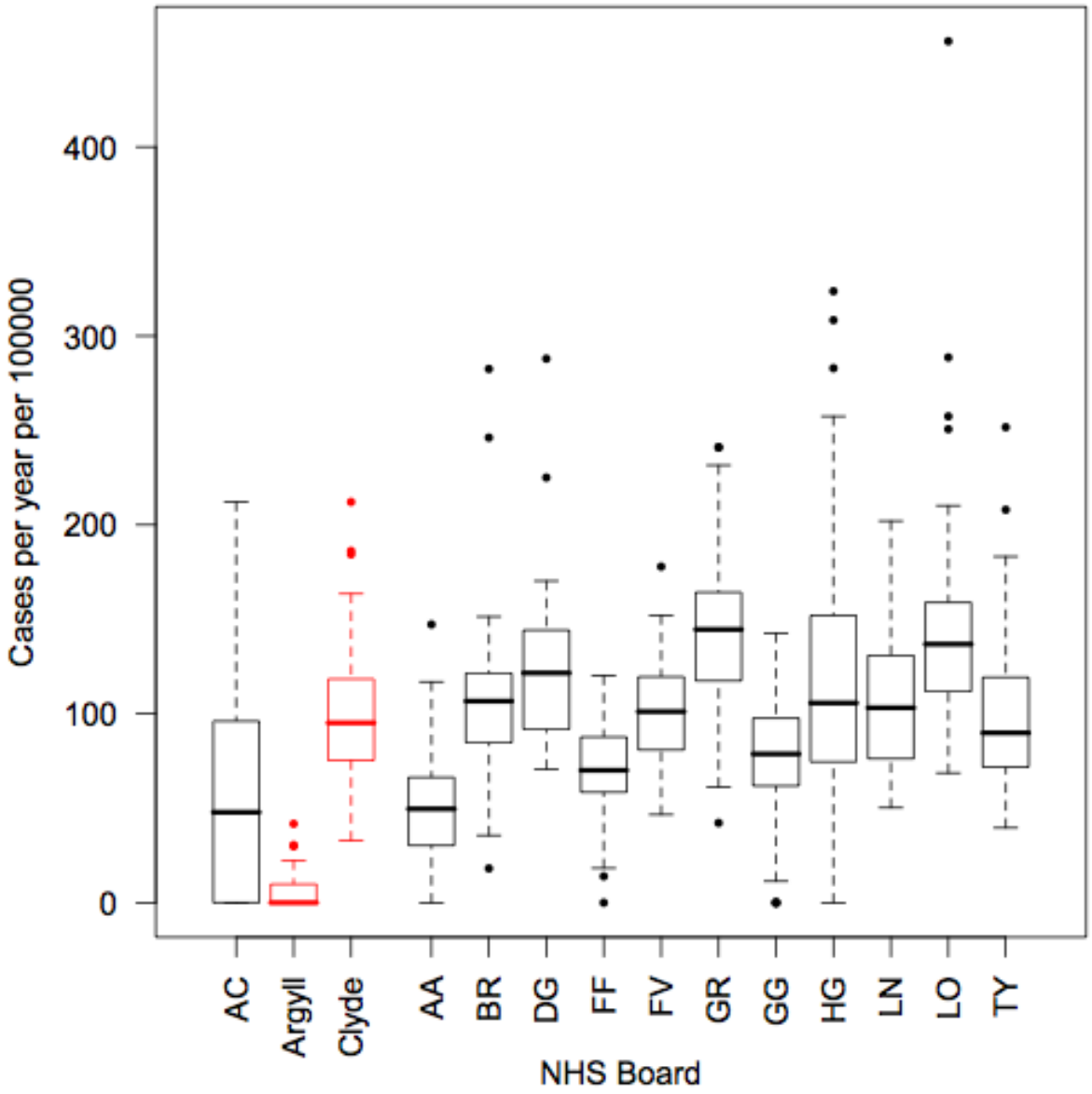


Fig 5.11 Box plot of cases per 100000 for the 12 health boards with Argyll and Clyde broken down into two separate units (shown in red).

Whereas the descriptive analyses (section 5.14) investigated potential differences in reporting rates across whole health boards, a statistical analysis of cross-border differences can help identify whether real differences in reporting rate exist. A postal sector map (Fig. 5.12) showing the incidence of campylobacteriosis in Scotland illustrates low rates in the west comprising the Ayrshire & Arran and Argyll & Clyde health boards. Figure 5.13 demonstrates that for Ayrshire & Arran there is an abrupt change at the border whereas for Argyll & Clyde the change occurs at the Argyll part of the health board. Argyll has a low incidence compared with Clyde and the surrounding health board.

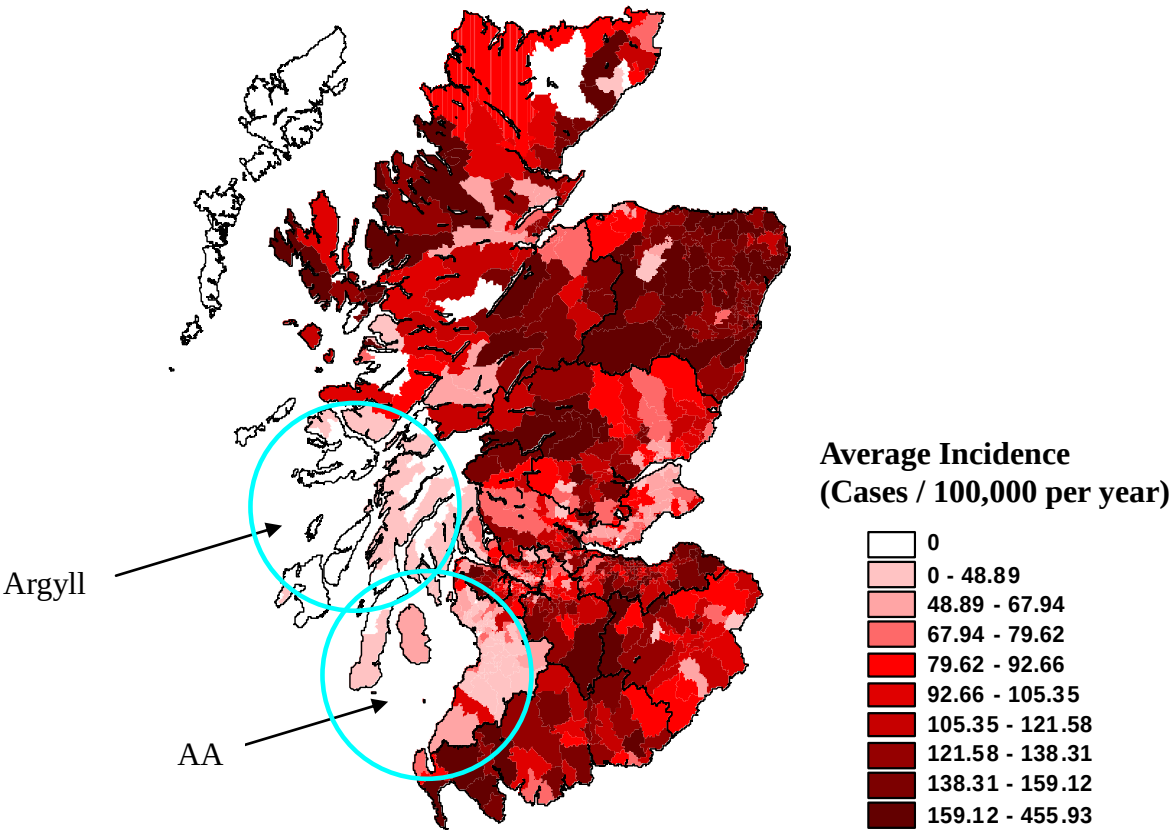


Figure 5.12 Average Campylobacter incidence (cases per 100,000 per year, 2000-2006) across Scotland at the level of postal sector.

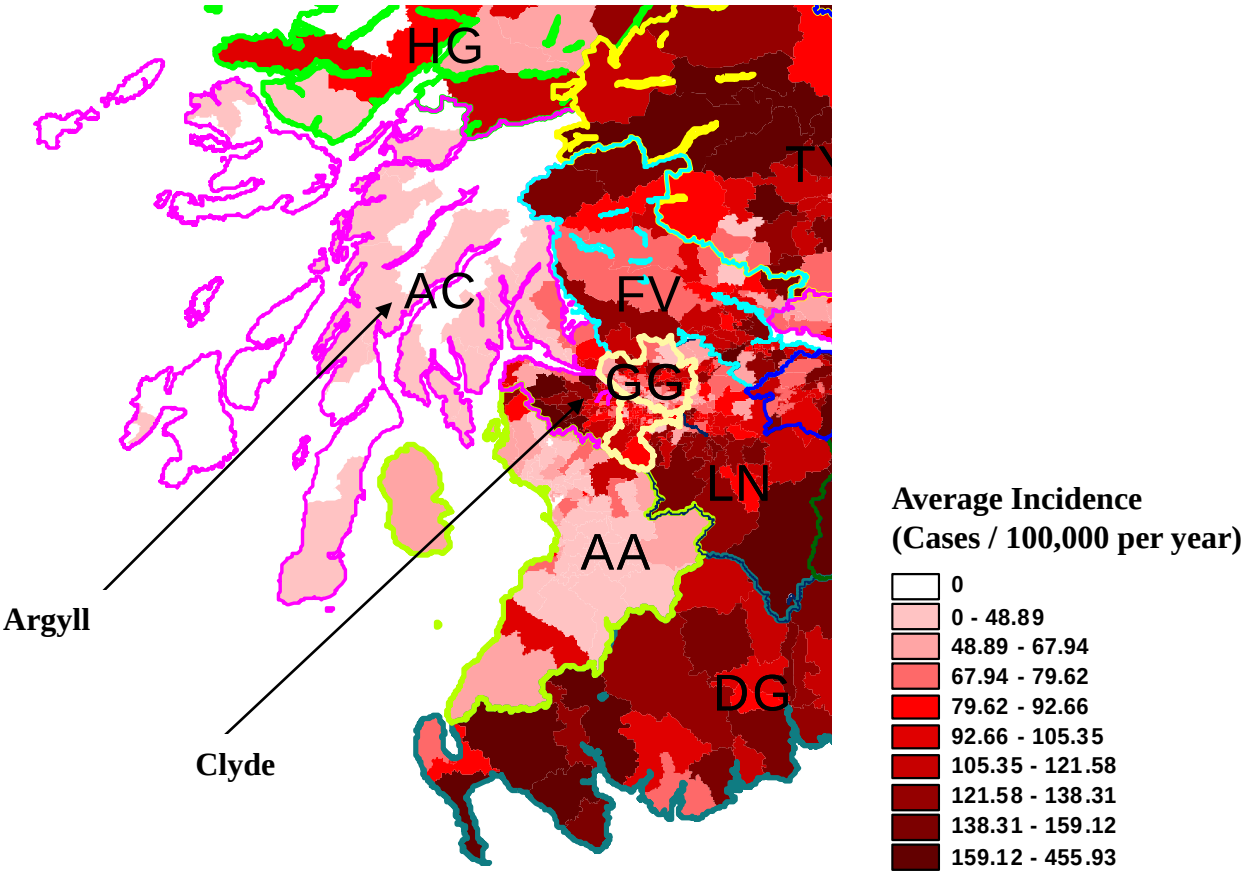


Figure 5.13 Average *Campylobacter* incidence (cases per 100,000 per year, 2000-2006) highlighting Ayrshire & Arran and Argyll and Clyde independently.

The difference in incidence across the borders of each health board is given in Table 5.5 (The Argyll part of Argyll & Clyde is also included). The results show that Ayrshire & Arran and the Argyll part of Argyll & Clyde both have significantly lower incidence than their surrounding health boards. For Ayrshire & Arran, figure 5.14 shows the position of the observed difference across the border relative to the frequency distribution of expected differences under the null hypothesis of no difference across the border. This clearly illustrates a significant difference between Ayrshire & Arran and neighbouring health boards.

Table 5.5 Differences in incidence (cases per 100,000 per year, 2000-2006) across health board borders together with statistical significance by randomisation test.

Health Board	Average incidence at border within health board	Average incidence at border outwith health board	Difference in incidence across border	P-value[a]
AA	62.59	138.60	-76.01	**<0.0001**
AC	87.65	87.59	0.06	0.5017
Argyll	17.55	98.81	-79.26	**<0.0001**
BR	140.98	152.01	-11.03	0.6013
DG	140.99	89.74	51.24	0.0409*
FF	74.79	94.19	-19.40	0.1716
FV	103.77	78.57	25.20	0.0161
GG	89.82	103.37	-13.55	0.0951
GR	159.79	115.38	44.41	0.0256**
HI	106.34	91.95	14.39	0.6617
LA	108.33	105.41	2.92	0.8052
LO	131.82	119.82	12.00	0.4721
TY	100.25	104.95	-4.70	0.7116

[a]Critical P-value following Bonferroni correction is 0.004
* The difference is caused by the low incidence in AA.
** The difference is caused by the low incidence in TY

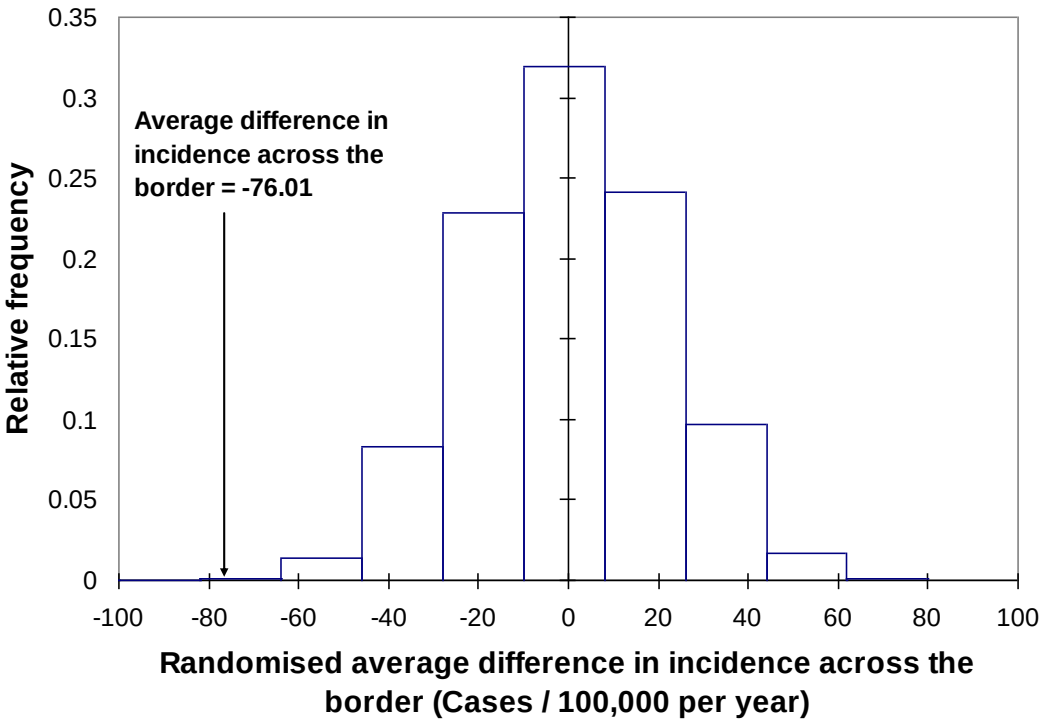


Figure 5.14 Frequency distribution of randomised average differences in incidence across Ayrshire & Arran's border compared with the actual difference in incidence (-76.01). The actual difference was less than the difference obtained from all of the 10,000 random shuffles of the incidence in each of the postcode sectors along the border. Hence the P value is < 0.0001.

5.3.1.2 Analysis of the microbiological methods

Within the confines of this study we were unable to audit the whole reporting pyramid but obtained laboratory protocols from each *Campylobacter* testing hospital. Initial contact to individual laboratories was made by email and supplementary information was sought by repeated mails and clarification made by telephone contact. Table 5.6 is a summary of protocols used and a summary of the hospitals and their respective *Campylobacter* isolation protocols (i.e. standard operating procedures). All laboratories that responded to the question stated that all faeces submitted were tested for the presence of *Campylobacter* spp. Not unexpectedly we found several differences in isolation protocols. The differences however do not help explain variation in the reporting rates between health boards.

Isolation medium

Charcoal Cefoperazone Deoxycholate agar (CCDA) and Skirrow's medium were used by six and five laboratories respectively. Those using CCDA included health boards with high reporting rates (Lothian and Lanarkshire) and low reporting rates (Fife and Greater Glasgow). Similarly, those using Skirrow's medium included health boards with high (Dumfries & Galloway and Borders) and low (Ayrshire & Arran) reporting rates. We conclude that the primary isolation medium is not consistent with reporting rates. Grampian use a Preston formula that contains one additional antimicrobial agent (to inhibit fungal growth) compared to Skirrow's medium but with other health boards (Lothian, Borders and Dumfries & Galloway) having similar isolation rates it is unlikely that this is in some way responsible for the continued high reporting rate in Grampian.

Isolation incubation temperature

Isolation temperatures were reported by most health boards as 42/43ºC with two Lanarkshire laboratories (Hairmyres and Wishaw) using 37ºC. health boards with both high and low reporting rates used the higher temperature while the two laboratories using 37ºC had mid-range reporting rates suggesting that incubation temperature was not correlated with reporting rates. We noted that length of incubation time (48h) was consistent throughout all laboratories.

Supplementary laboratory test information (Table 5.7) including biochemical tests, the use of control strains and type of microaerophilic incubation also failed to identify microbiological practices as being responsible for *Campylobacter* isolation rate differences in health boards. Oxidase and Gram tests were performed routinely by all laboratories apart from Raigmore where feedback data were equivocal. The majority of laboratories did not respond to the question of whether they routinely performed the catalase test but the benefits of this biochemical reaction are unclear and Aberdeen University microbiologists did not include this reaction in the CaMPS FSA Scotland project. The use of control organisms in diagnostic microbiology is now performed routinely and is an essential component of accreditation schemes, UK hospital laboratories take part in the Clinical Pathology Accreditation. Seven laboratories here failed to report to us their use of control organisms. It was interesting to note that only three laboratories used microaerophilic cabinets, the rest using gas jars. Large throughput of samples is facilitated by cabinet use but their locations in this study did not correlate with laboratory size e.g. Lothian, with the largest throughput used gas jars while Gilbert Bain hospital in Shetland have installed a cabinet. Both systems used correctly allow the effective isolation of target pathogens but it is possible that slow growing *Campylobacter* strains might be more easily recognised in a cabinet which could result in higher isolation rates.

From the data supplied, regional variations in *Campylobacter* isolation rates cannot be attributed to differences in media and isolation procedures. Unfortunately it was not possible to assess laboratory staff expertise and ability in this study. However, our own experience with staff in Aberdeen University's microbiology laboratories and in particular those involved in *Campylobacter* isolation does suggest that familiarity with the technique and a dedication

to meticulous procedure increases isolation rates. We cannot therefore rule out laboratory personnel in having an effect on isolation rates but the fact that current health board variations are consistent over time suggests that this does not significantly contribute.

Table 5.6 Health Board laboratory isolation protocols for *Campylobacter* from stool samples.

Hospital (HB)	What is Tested	Temperature of incubation (ºC)	Time of incubation (h)	Isolation medium	Isolation rate (/100,000) 2002-06
AA - Crosshouse	All faeces	43	40-48	*Campylobacter* selective agar - based on Skirrow's medium	52.9
BR -Borders	NR	NR	NR	Skirrow's medium	113.3
DG -D&G RI	All faeces	42	48	Skirrow's medium[1]	113.6
FF -Fife	All faeces	42	48	CCDA	73.0
FF -Victoria	Routine	42	48	*Campylobacter* selective agar (CMP)	73.0
FV –Stirling	NR	42	48	Oxoid (PB0295A) *Campylobacter* agar (Preston)	91.4
GG -Glasgow RI	All faeces	43	48	commercial campy agar	66.8
GG -Royal Alex	all faeces	42	48	CCDA	66.8
GG -Vale of Leven	NR	NR	NR	Skirrow's *Campylobacter* selective agar	66.8
GR -Aberdeen	All faeces	42	48	Selective *Campylobacter* (Preston like)	119.5
HG -Raigmore	All faeces	42	48	CCDA	94.0
LN -Hairmyres	All faeces	37	48	Oxoid *Campylobacter* agar[2]	101.7
LN - Monklands	All faeces	NR	48	CCDA	101.7
LN -Wishaw	All faeces	37	48	Oxoid *Campylobacter* agar[2]	101.7
LO -St John's	All faeces	42	48	Oxoid selective *Campylobacter* plates (Skirrow)	115.3
LO -Western General	NR	43	48	CCDA	115.3
OR -Balfour	All faeces	42	48	Oxoid selective *Campylobacter* agar	
SH -Gilbert Bain	All faeces	42	48	Selective *Campylobacter* (Preston like)	
TY -Ninewells	All faeces	42	48	CCDA	97.3
TY -Perth	All faeces	42	48	CCDA	97.3

NR, no response to the request for information
[1], not the medium specified in the HPA National Standard BSOPID 23

[2], Campylobacter blood-free selective medium contains charcoal in place of blood and antibiotic supplements cefoperazone (as the selective agent) and amphotericin-B

Table 5.7 Supplementary microbiological methods for analysis of stool samples for *Campylobacter*

Hospital	Oxidase test	Gram test	Catalase test	Use of controls	Use of gas jar or cabinet
AA - Crosshouse	Y	Y	Y	Y	NR
BR –Borders	Y	Y	Y	NR	Gas jar
DG -D&G RI	Y	Y	NR	Y	Gas jar [2]
FF –Fife	Y	Y	NR	Y	NR
FF –Victoria	Y	Y	Y	Negative control	Gas jar
FV –Stirling	Y	Y	NR	NR	NR
GG - Glasgow RI	Y	Y	NR	Y	Gas jar
GG -Royal Alex	Y	Y	NR	Y	Gas jar
GG -Vale of Leven	Y	Y	NR	NR	Gas jar
GR - Aberdeen	Y	Y	NR	Y	Cabinet
HG - Raigmore	NR	Sometimes	NR	Y	Gas jar
LN - Hairmyres	Y	Y	NR	Y	Gas jar
LN - Monklands	Y	Y	NR	NR	Gas jar
LN -Wishaw	Y	Y	NR	Y	Gas jar
LO -St John's	Y	Y	NR	Y [1]	Gas jar
LO -Western General	Y	Y	NR	Y	Gas jar
OR -Balfour	Y	Y	NR	NR	Microaerophilic
SH -Gilbert Bain	Y	Y	NR	Y	Cabinet
TY - Ninewells	Y	Y	NR	NR	Microaerophilic
TY -Perth	Y	Y	NR	NR	NR

NR, no response to the request for information
[1]wild clinical strain positive control used, [2]Oxoid Campygen gas kits

From the data supplied, regional variations in *Campylobacter* rates can not be attributed to differences in media and isolation protocols.

As no obvious explanation for the regional variations could be drawn form the laboratory protocols for *Campylobacter*, the infection rates for four gastrointestinal pathogens were ranked by health board across Scotland (Table 5.8). This was performed to identify whether

some health boards had high reporting rates for all of the GI pathogens or whether *Campylobacter* was the exception and therefore more likely be due to differences in laboratory protocol. It was noticeable that Grampian and Dumfries & Galloway are ranked consistently high for all four pathogens whilst apart from *E. coli* O157 Ayrshire & Arran are ranked bottom for the other three pathogens. Overall, Argyll & Clyde is ranked low (8-9) but the data solely for Argyll was unavailable. This strengthens the view that *Campylobacter* media and protocols are probably not responsible for different HB *Campylobacter* rates.

Table 5.8 Ranking of Health Board infection rates 2002-2006 for four gastrointestinal pathogens (highest infection rate in Scotland ranked 1, lowest rate ranked 12).

HB	***Campylobacter***	***E. coli* O157**	***Salmonella***	***Cryptosporidium***
HG	7	11	10	4
GR	1	1	1	5
TY	6	6	6	1
FF	10	5	11	7
FV	8=	7	3	9
LO	2	9=	2	6
BR	4	3	7	2
DG	3	2	5	3
AA	12	4	12	12
LN	5	9=	4	10
GG	11	12	8	11
AC	8=	8	9	8

5.3.1.3 Analysis of health board

Inclusion of health board as a random effect in the spatial regression models provides the opportunity to examine the effect of health board once the risk factors have been accounted for. These analyses revealed lower than expected rates in the Argyll sector and Ayrshire and Arran Health boards (Fig. 5.15 below). The Grampian, Highland and Tayside Health boards had higher than expected *Campylobacter* incidence. In spite of having the second highest case rate the Lothian Health Board did not have significantly different from expected incidence rates, showing that the model has accounted for the risk in this area.

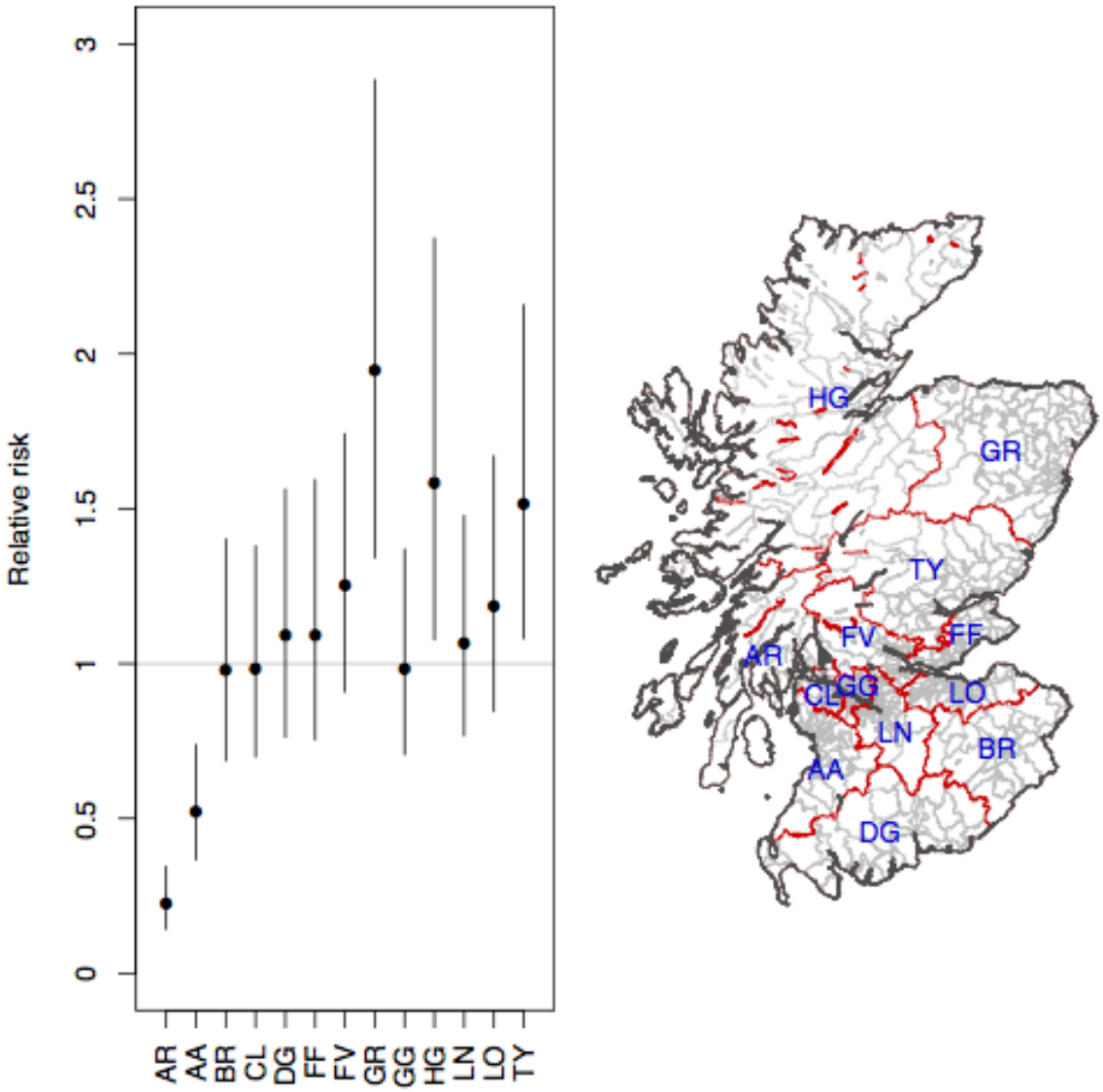


Figure 5.15 Relative risks and 95% Confidence intervals attached to each NHS Health Board from the multivariate model and the geographical distributions of the health boards (red borders) relative to postcode sectors (grey borders).

5.3.2 Urban – rural

5.3.2.1 Population density as a risk factor

In the risk factor model, population density was used a continuous measure reflecting highly urban through to highly rural populations and was found to be a significant protective effect in univariate analysis (Table 5.9) and the final model (Table 5.10). The relative risk (Table 5.7) of 0.945 shows that the expected number of cases in postcode sector i will be 94.5% that of sector j if the population density in i is 1 $\log_{10}$ unit higher than j and all other factors are the same. In univariate screening all predictors, except the livestock density predictors were significant at $p<0.05$, however in the multivariate model, only Carstairs and population density remained.

The association with lower population densities is likely to be a proxy for some qualities of rural environments, such as exposure to all environmental reservoirs, including private water supplies and ruminant sources of infection. Indeed, risk factor analyses conducted on subsets of the data (results for poultry and ruminant attributed strains shown in Tables 5.13 and 5.14 respectively), showed population density to be a significant protective factor for ruminant acquired strains, but not for poultry strains. This suggests that population density is indeed providing a measure of rurality, but that this effect is dominated by other risk factors in the analysis of the full dataset.

Table 5.9 Univariate poisson GLM analysis of risk factors. The p-values indicate whether each of the putative risk factors is significant of *Campylobacter* incidence in a univariate model. The estimates for each predictor give a relative measure of the expected change in outcome per unit change in the predictor.

Predictor	Unit	Estimate	Std. error	p-value
Easting	m	$3.533 * 10^{-6}$	$2.457 * 10^{-7}$	<0.001
Northing	m	$9.412 * 10^{-7}$	$1.835 * 10^{-7}$	<0.001
Carstairs score		-0.050	0.004	<0.001
Private water density[1]	supply/person	4.754	0.709	<0.001
Human density[1]	people/ km^2	-0.033	0.015	0.031
Cattle density	cattle/ km^2	$-6.401 * 10^{-4}$	$3.854 * 10^{-4}$	0.097
Sheep density	sheep/ km^2	$-4.655 * 10^{-6}$	$1.631 * 10^{-4}$	0.977
Poultry density	poultry/ km^2	$-2.460 * 10^{-6}$	$3.654 * 10^{-6}$	0.501

[1] $\log_{10}$ transformed

Table 5.10 Reduced multivariate spatial regression model. See section 4.2.4 for the definition of the Carstairs index. As a higher Carstairs index is associated with greater deprivation, a negative coefficient means that the risk of reporting infection diminishes with greater deprivation.

Predictor	Unit	Relative risk (95% CIs)
Carstairs		0.965 (0.959, 0.971)
Population density	$\log_{10}$(people/Km^2)	0.945 (0.916, 0.974)

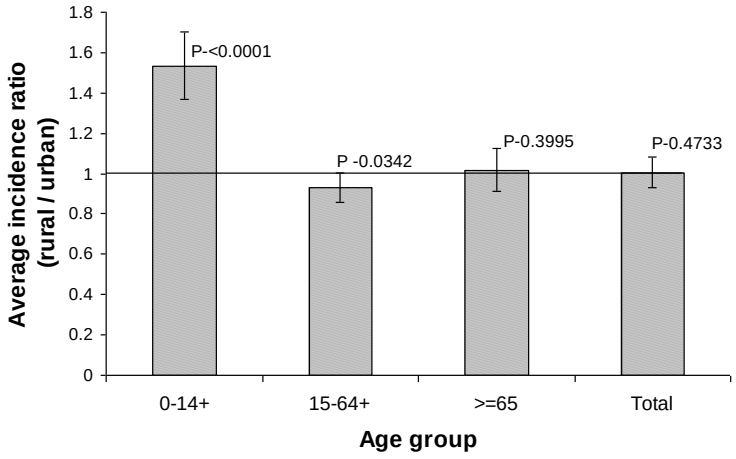
5.3.2.2 The difference between rural and urban reporting rates

Using population density as a proxy for rural/urban, the ratio of reporting between rural and urban age groups was calculated. The results (Fig. 5.16(a)) demonstrate that there is no overall rural-urban reporting difference of *Campylobacter cases* during the period 2000-06, but for 0-14 year olds there are more rural cases (P<0.0001) whilst for 15-64 year olds there are more urban cases (P=0.0342)

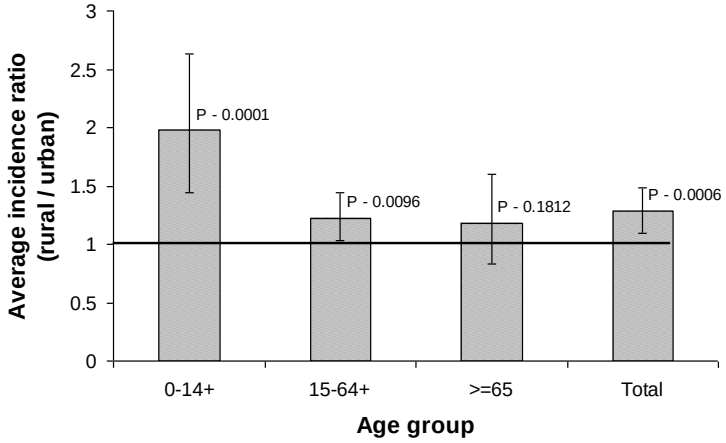
When studying the MLST dataset the overall incidence was higher in rural areas for both *C. jejuni* and *C. coli* (Fig. 5.16(b) & (c)).

The biggest differences in incidence were in rural children who present with approximately 1.5 (2000-06 dataset), 2.0 (MLST *C. jejuni*) and 2.5 (MLST *C. coli*) fold higher incidence compared with their urban counterparts.

(a)



(b)



(c)

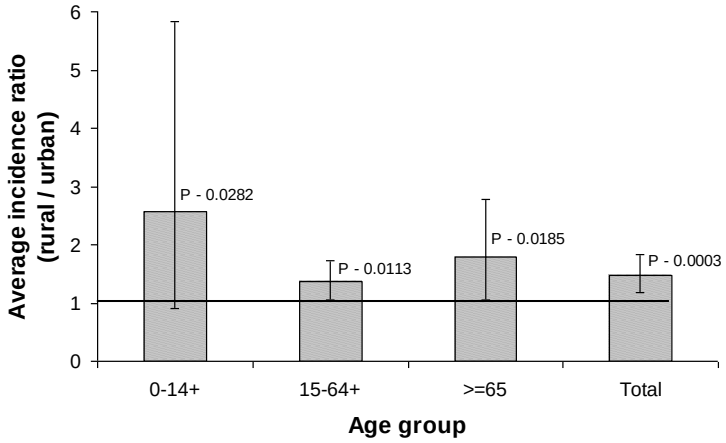
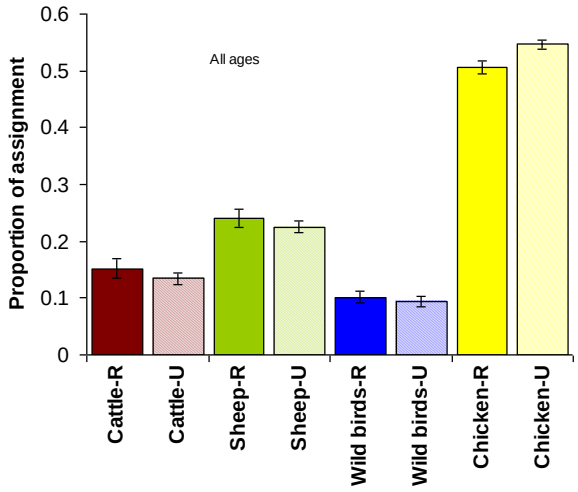


Figure 5.16 Average incidence ratio (rural/urban) by age groups for (a) *Campylobacter* 2000-06, (b) *C. jejuni* and (c) *C. coli* for the 15 months of the MLST study (P-values indicate whether rural-urban ratio is significantly different from 1). Error bars are 95% CI's.

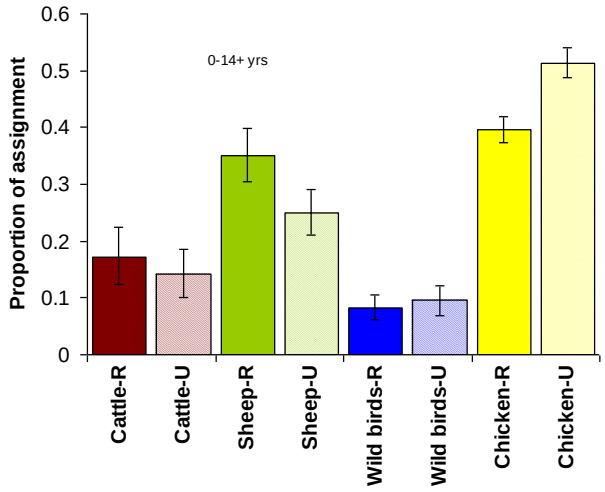
Source attribution was applied using STRUCTURE to identify the source of human strains from the animal reservoirs. This was carried out independently for both *C. jejuni* and *C. coli.* Further, this was stratified by urban and rural population to determine whether the relative importance of the sources were different.

The source attribution results (Fig. 5.17) for both the urban and rural human populations demonstrate that retail chicken is the most important source. However the data suggests, particularly for children, that retail chicken has a reduced importance in rural areas whereas this is contrasted with an increase in the importance of ruminant sources. Conversely urban children may be less exposed to the environmental route via ruminants and more likely to acquire infection via retail chicken consumption.

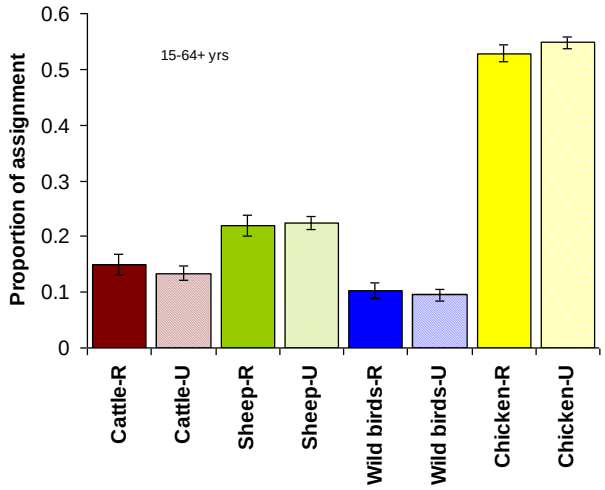
(a)



(b)



(c)



(d)

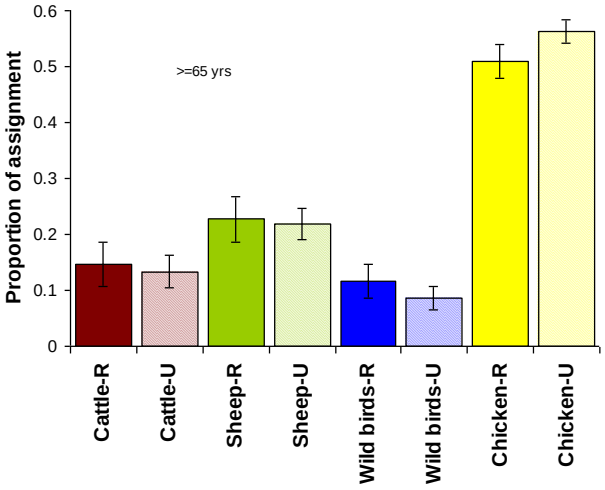
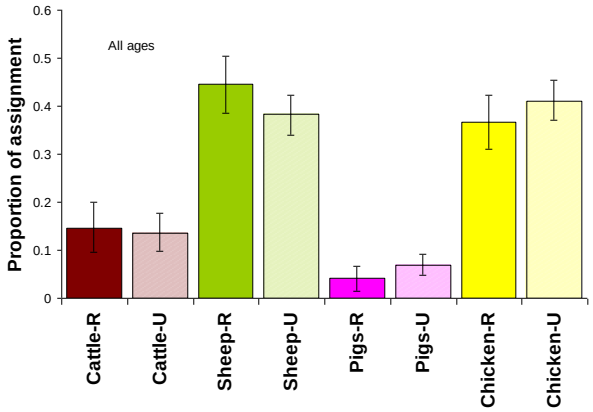


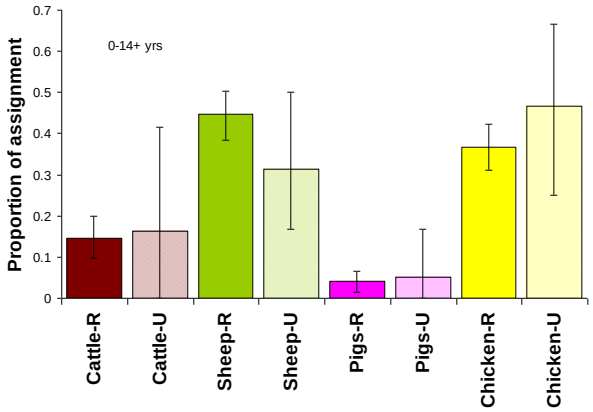
Figure 5.17 Rural and urban attribution of human *C. jejuni* isolates to source reservoirs: (a) all ages, (b) 0-14+ years, (c) 15-64+ years and (d) >=65 years. Error bars are 95% CI's.

C. coli: The source attribution results (Fig. 5.18) show that poultry and sheep are the most important source reservoirs and that pigs are of little importance. The trend for all of the age groups shows that the attribution to retail chicken is less compared to *C. jejuni* (rural retail chicken attribution 51% *C. jejuni* and 37% *C. coli*, urban attribution 55% *C. jejuni* and 41% *C. coli*) and that attribution to sheep is more important in rural areas. However, the confidence intervals are large due to the small numbers of data and this may be causing these results not to be statistically significant (e.g. attribution 45±6% *C. coli* in rural children and 31±16% for their urban counterparts).

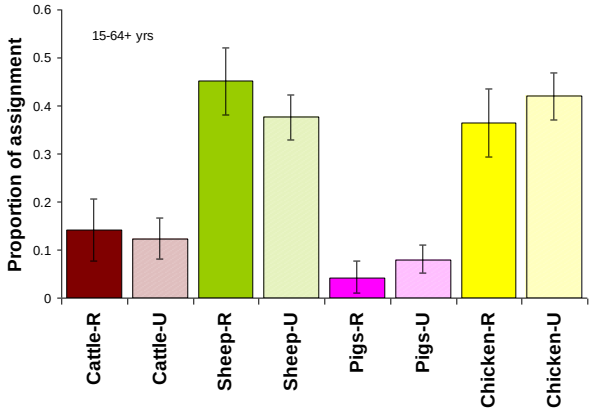
(a)



(b)



(c)



(d)

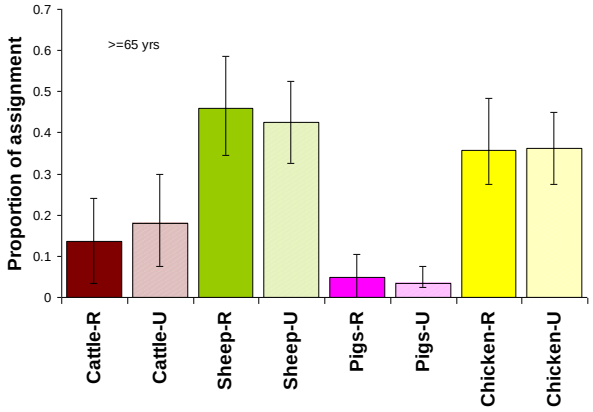


Figure 5.18. Rural and urban attribution of human *C. coli* isolates to source reservoirs: (a) all ages, (b) 0-14+ years, (c) 15-64+ years and (d) >=65 years.

5.3.3 Reservoirs

5.3.3.1 Similarity between human and animal MLST types

The proportional similarity index was determined (Table 5.11) between human isolates and source reservoirs for *C. jejuni* and *C. coli* independently. The most similar source to humans was chicken for *C. jejuni* and sheep for *C. coli*. Chicken was second most important source for *C. coli* and pigs were the least important.

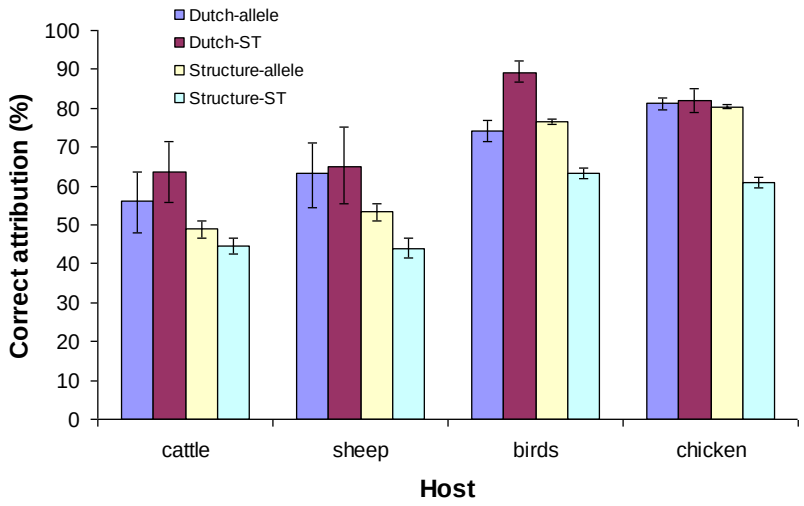
Table 5.11 Proportional similarity index comparing human and source reservoir isolates.

Campylobacter species	Source reservoir	Proportional Similarity Index	95% CI
C. jejuni	Cattle	0.23	0.22-0.24
	Sheep	0.25	0.24-0.27
	Wild birds	0.14	0.13-0.15
	Chicken	0.49	0.47-0.50
C. coli	Cattle	0.30	0.27-0.31
	Sheep	0.53	0.48-0.57
	Pigs	0.06	0.04-0.07
	Chicken	0.45	0.39-0.46

5.3.3.2 Self attribution

Self attribution was used to determine the accuracy of STRUCTURE and the DUTCH models in predicting whether an isolate found from a particular animal source actually came from that source. It was found that self-attribution ranged from 44-89% for *C. jejuni* and 52-88% for *C. coli*. All the methods were significantly better than would be expected by chance (25%). STRUCTURE at the ST level appeared to have the poorest performance (Fig 5.19). For *C. jejuni*, cattle and sheep tend to have a lower score because it is known that a number of the ST's are common in these two host species (McCarthy et al. 2007).

(a)



(b)

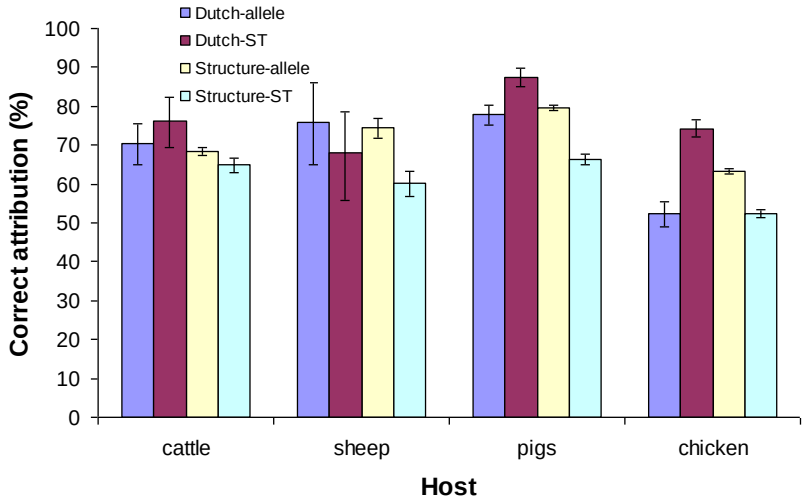


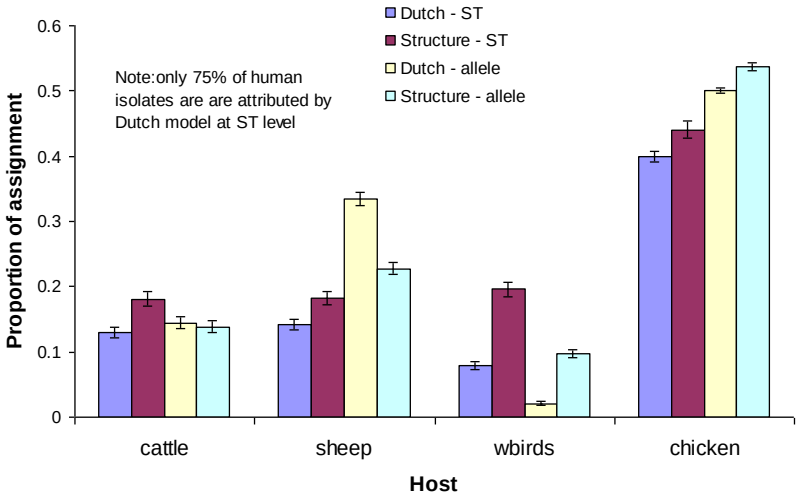
Figure 5.19 Self-attribution of animal source reservoirs for (a) *C jejuni*, and (b) *C. coli* for the DUTCH and STRUCTURE models at the level of ST and allele. Error bars are 95% CI's.

5.3.3.3 Attribution of human isolates to reservoirs

C. jejuni: Both models at both ST and allele level assigned most (40%-53%) human isolates to chicken (Fig. 5.20(a)). Ruminant strains were then the next most important (27-47%) followed by birds (8-20%) although STRUCTURE at ST level ranked birds 2nd. It should be noted that the DUTCH MODEL at ST level could only attribute 75% of isolates.

C. coli: Sheep (32-49%) and chicken (32-40%) were the most important sources (Fig. 5.20(b)). Cattle were less important (14-17%) and pigs were least important (4-20%) for all models except for STRUCTURE at ST level.

(a)



(b)

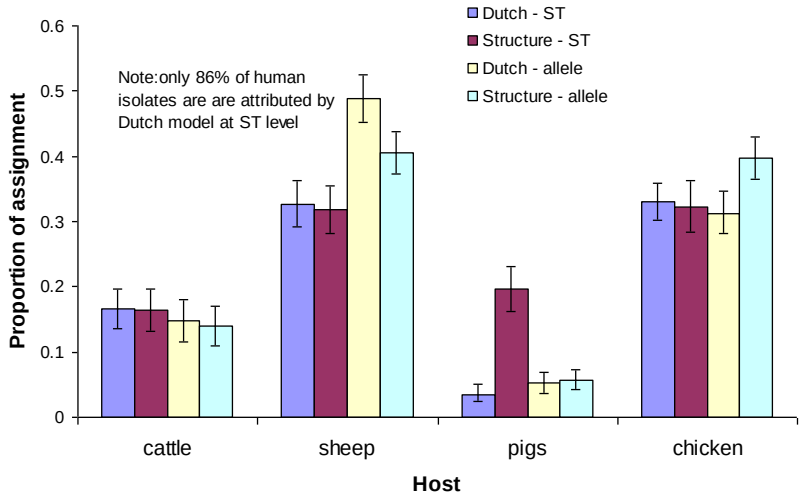


Figure 5.20 Assignment of human isolates to source reservoirs using STRUCTURE and DUTCH models at ST and allele for (a) *C. jejuni* and (b) *C. coli.*

5.3.3.4 Strain-specific risk factors

Sheep, cattle and poultry densities were considered important putative risk factors, and were included in the univariate analysis of the risk factor modelling. However, none of these factors were sufficiently significant, and were not included in the final model. Thus, for the full *Campylobacter* dataset, we conclude that we cannot discern potential risks associated with high cattle, sheep or poultry densities.

However, conducting the risk factor modelling at the resolution of poultry attributed or ruminant attributed strains allows the potential identification of strain specific risk factors. The majority of STs could be assigned to either a poultry or a ruminant source (Table 5.12). The risk factor modelling was repeated for the chicken assigned STs and for the ruminant assigned STs. Overall, the risk factor models for these subsets identified similar types and magnitudes of risk factor to those identified in the analysis of the full dataset (Tables 5.10 and 5.11), with deprivation in particular remaining a significant protective factor. The subset models have resulted in some large change in the distribution of the health board levels, suggesting that there are significant variations between health boards in the species specific *Campylobacter* strains.

Table 5.12 Attribution of human case sequence types to a poultry, ruminant, wild bird or unassigned source.

Attribution	Number of cases	Number of STs
Poultry	1531	99
Ruminant	1038	49
Wild bird	64	20
Unassigned	1046	348

Table 5.13 Multivariate GLM of the risk of being infected with a poultry attributed strain. See section 4.2.4 for the definition of the Carstairs index. As a higher Carstairs index is associated with greater deprivation, a negative coefficient means that the risk of reporting infection diminishes with greater deprivation.

Predictor	Unit	Estimate	Std. error	t-value	p-value
Health Board	Argyll	0.915	1.149	0.796	0.426
	Ayrshire & Arran	1.434	0.973	1.472	0.141
	Borders	1.818	0.995	1.872	0.068
	Clyde	1.880	1.025	1.834	0.067
	Dumfries & Galloway	2.699	0.874	3.090	0.002
	Fife	2.592	1.062	2.442	0.015
	Forth Valley	1.920	1.055	1.821	0.069
	Grampian	3.828	1.243	3.081	0.002
	Greater Glasgow	2.177	1.011	2.153	0.031
	Highland	3.672	1.293	2.840	0.005
	Lanarkshire	2.742	0.999	2.746	0.006
	Lothian	2.969	1.018	2.916	0.004
	Tayside	3.316	1.113	2.980	0.003
Northing		$-4.013 * 10^{-6}$	$1.514 * 10^{-6}$	-2.652	0.008
Carstairs		-0.037	0.011	-3.244	0.001

Table 5.14 Multivariate GLM of the risk of being infected with a ruminant attributed strain.

Predictor	Unit	Estimate	Std. error	t-value	p-value
Health Board	Argyll	-1.475	0.501	-2.946	0.003
	Ayrshire & Arran	-0.695	0.250	-2.777	0.005
	Borders	-3.182	0.232	-0.137	0.891
	Clyde	-0.600	0.197	-3.048	0.002
	Dumfries & Galloway	0.758	0.188	4.046	<0.001
	Fife	-0.386	0.176	-2.199	0.028
	Forth Valley	-0.900	0.227	-3.966	<0.001
	Grampian	0.760	0.173	6.476	<0.001
	Greater Glasgow	-0.282	0.153	-1.846	0.065
	Highland	0.356	0.134	2.659	0.008
	Lanarkshire	0.348	0.361	2.560	0.010
	Lothian	0.317	0.111	2.855	0.004
	Tayside	0.429	0.110	3.891	<0.001
Carstairs		-0.069	0.015	-4.510	<0.001
Human density	People/km^2	$-6.722 * 10^{-5}$	$3.129 * 10^{-5}$	-2.148	0.032
Cattle density	Cattle/km^2	-0.003	-0.002	-1.979	0.048

An additional logistic regression analysis was conducted to identify risk factors for infection with a ruminant strain rather than a poultry strain (Table 5.12). The analysis demonstrated that relative to the risk of acquiring a poultry strain, adults were significantly less likely than children (defined as 15 years or under) to be infected with a ruminant strain (OR=0.654, CIs 0.529-0.81); infection was significantly less likely to be acquired in the winter (OR=0.804, CIs 0.676-0.96); and infection rates were higher for ruminant relative to poultry strains in areas of greater sheep density (OR=1.002 CIs=1.001-1.003).

Table 5.15 Results of univariate logistic regression of risks associated with being infected by a ruminant attributed strain (a case) relative to being infected with a poultry attributed strain (a control).

Predictor	Unit	OR (95% CIs)	z-value	p-value
Age	Child	1	-	-
	Adult	0.654 (0.529, 0.810)	-3.888	<0.001
Season	Summer	1	-	-
	Winter	0.804 (0.673, 0.960)	-2.440	0.015
Sheep density	Sheep/km^2	1.002 (1.001, 1.003)	3.103	0.002

A risk factor analysis which separately examined the infection rates in over and under 15 year olds, showed no difference in the estimates for the Carstairs deprivation score; population density however changes from being a non-significant predictor in over 15s to a highly significant risk factor in under 15s (Fig. 5.21). This result supports our interpretation (in Section 5.3.2) that population density is providing a measure of rurality, a risk factor which results in an increased risk of acquiring ruminant associated strains in children.

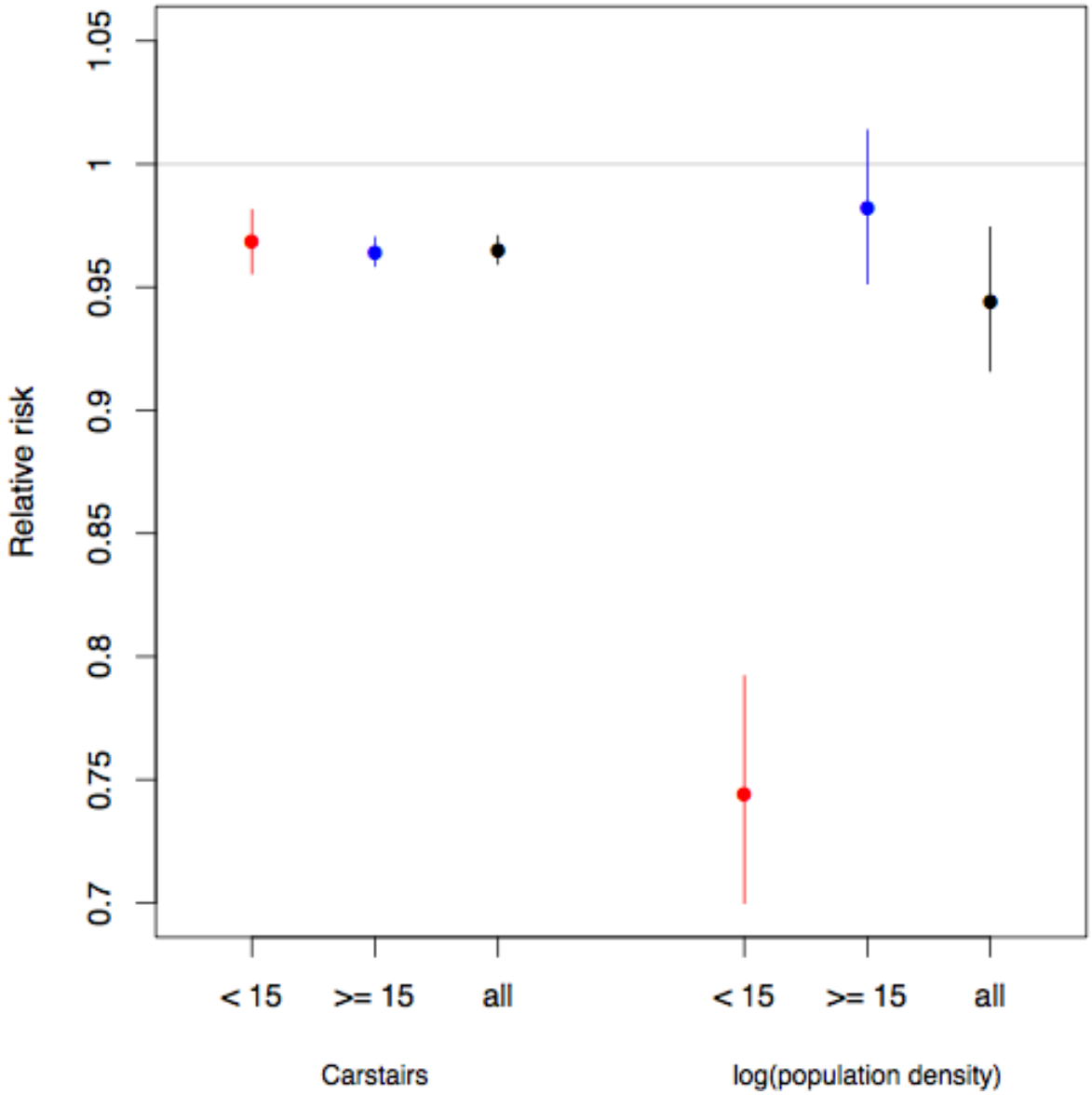


Figure 5.21 Relative risks and 95% confidence intervals (CIs) from the multivariate models (Tables 5.10 and 5.11). Separate models for cases under 15 (red), 15 and over (blue) and all data (black).

The K function analysis is also able to shed light on potentially differing epidemiologies between ruminant and poultry strains. K function analysis of the full dataset showed significant clustering in both space and time (Fig 5.22). With the data broken down into chicken and ruminant strains, the effects were stronger with regard to time, indicating differing temporal seasonal patterns between the chicken and ruminant strains with a stronger seasonal association in ruminant strains compared to chicken strains, with both chicken and ruminant strains associated over a period of around 12 weeks, but a weaker association in chicken strains (Fig 5.23 and 5.24). Viewed using a spatial K function, the chicken strains demonstrated a clear peak at short distances (Fig 5.25) indicating clustering of cases. In contrast, the spatial K function for the ruminant strains was less than 1 across almost all distances (Fig 5.26), indicating that cases are dispersed with respect to the expectation (the Scottish population). This is consistent with the view that ruminant acquired strains are associated with rural environments with sparse populations. The space-time plots confirm these results (Fig 5.27 and Fig 5.28), with both showing a peak (in red) at small

space and time scales, but with the chicken associated strains appearing to be more clustered in space.

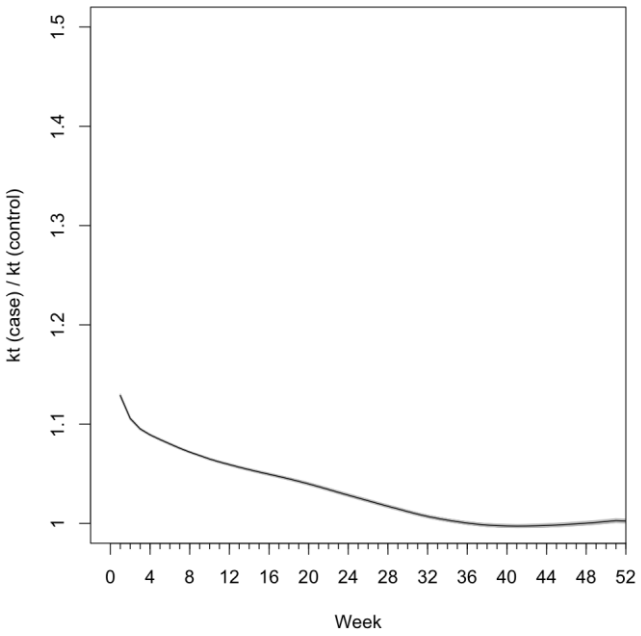


Figure 5.22 Temporal K function analysis of the full dataset. The y-axis represents the k-function with values above one indicating association and values below one representing dispersal.

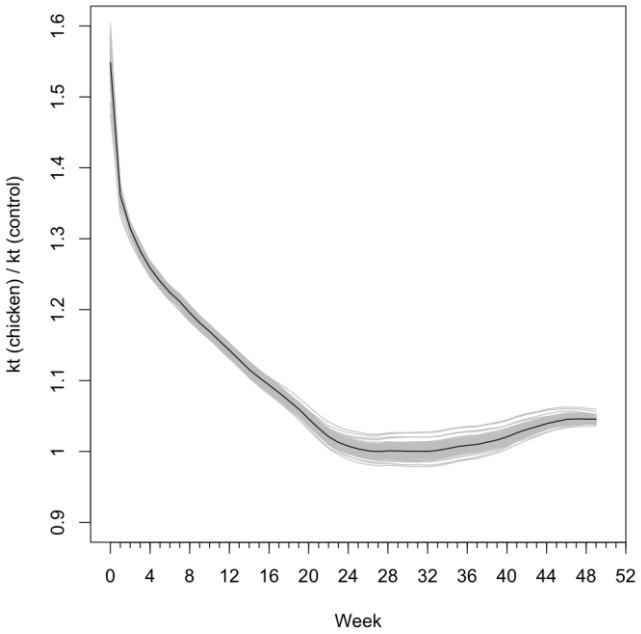


Figure 5.23 Temporal K function analysis of the chicken attributed strains. The y-axis represents the k-function with values above one indicating association and values below one representing dispersal.

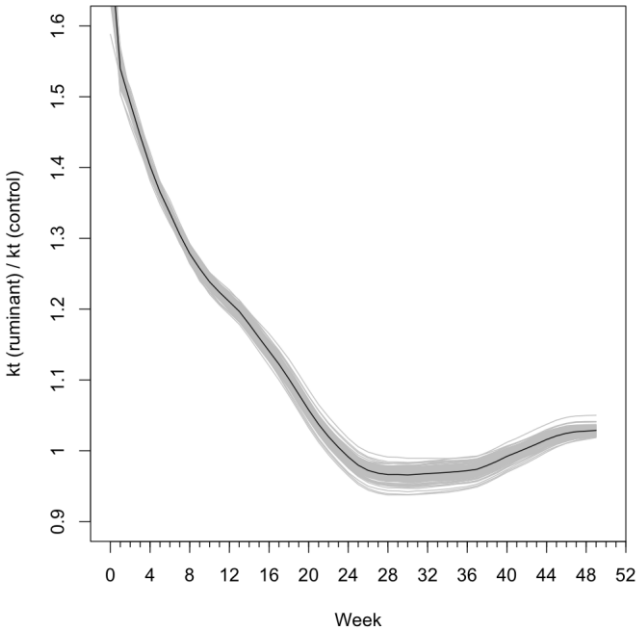


Figure 5.24 Temporal K function analysis of the ruminant attributed strains. The y-axis represents the k-function with values above one indicating association and values below one representing dispersal.

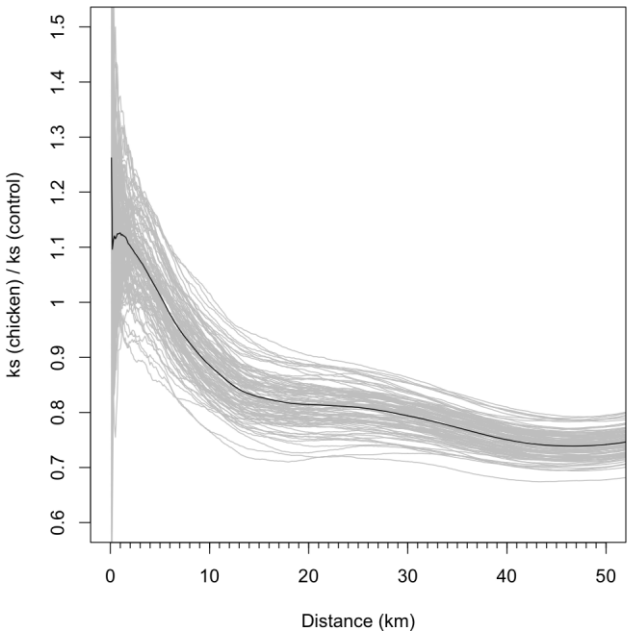


Figure 5.25 Spatial K function analysis of the chicken associated strains. The y-axis represents the k-function with values above one indicating association and values below one representing dispersal.

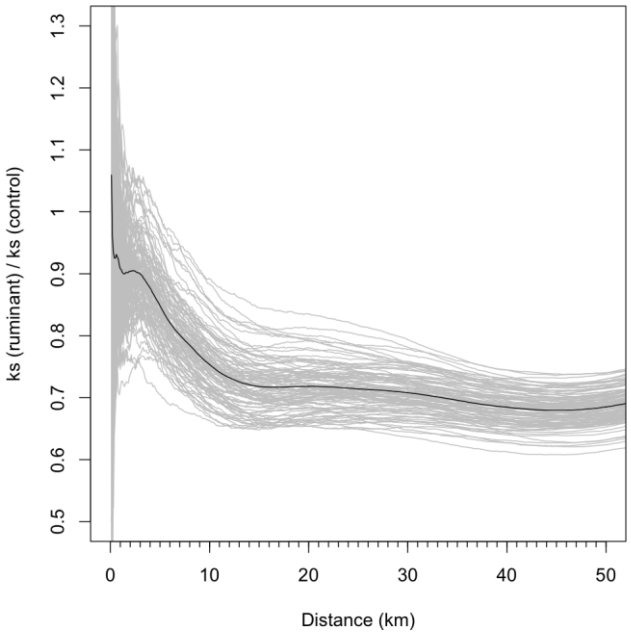


Figure 5.26 Spatial K function analysis of the ruminant associated strains. The y-axis represents the k-function with values above one indicating association and values below one representing dispersal.

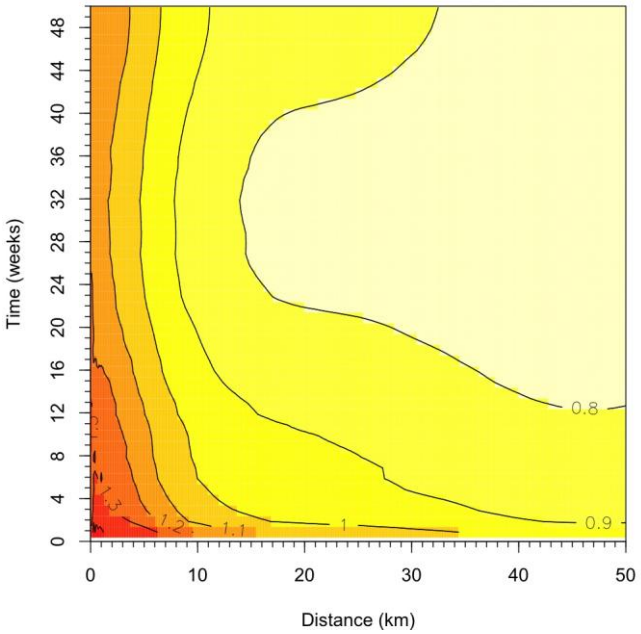


Figure 5.27 spatio-temporal analysis of the chicken associated strains. The contours indicate the k-function levels of association.

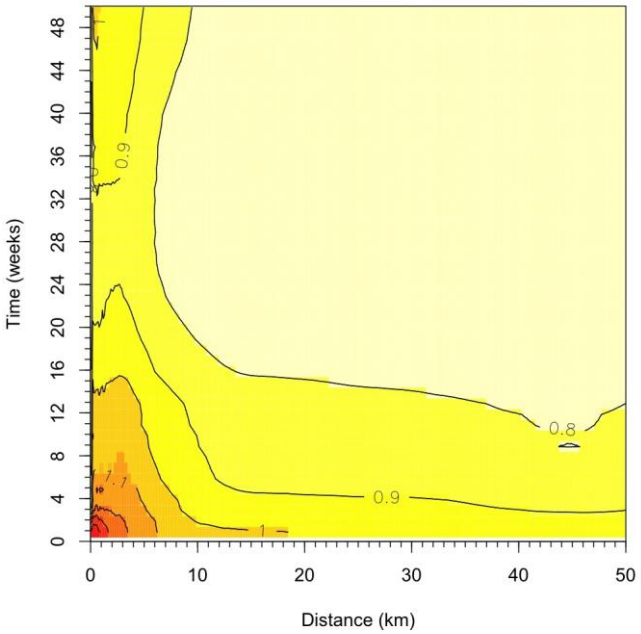


Figure 5.28 spatio-temporal analysis of the ruminant associated strains. The contours indicate the k-function levels of association.

K function analysis was also used to examine the spatio-temporal distributions of the three dominant sequence types in human cases (ST 21 a ruminant associated strain, and ST 45 and ST 257 which are both poultry associated strains). ST 21 and ST 257 generated similar K function results in both space and time, both showing a peak in time consistent seasonal variation in occurrence (Fig 5.29 and Fig 5.30). ST 45 however showed a much stronger peak in time (Fig 5.31), indicative of its strong seasonal pattern (Fig. 5.33). There was little evidence of spatial clustering in any of the three sequence types, with ST 45 in particular appearing dispersed relative to the control population (Fig 5.32). As ST 45 has also been linked with water-borne outbreaks, cases might be disproportionately expected in locations with private water supplies ie predominantly rural areas with a sparse population.

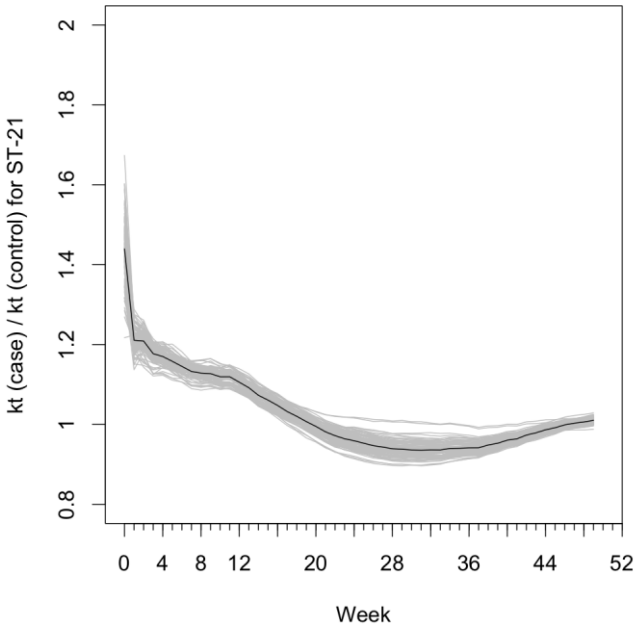


Figure 5.29 Temporal K function analysis of ST 21. The y-axis represents the k-function with values above one indicating association and values below one representing dispersal.

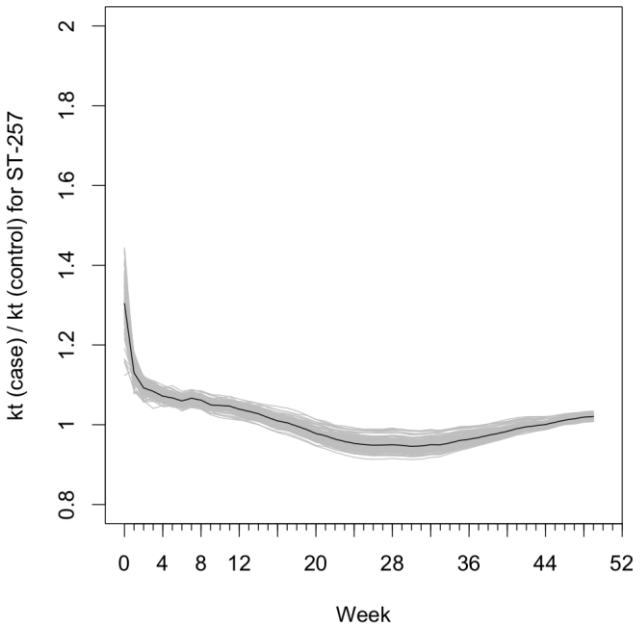


Figure 5.30 Temporal K function analysis of ST 257. The y-axis represents the k-function with values above one indicating association and values below one representing dispersal.

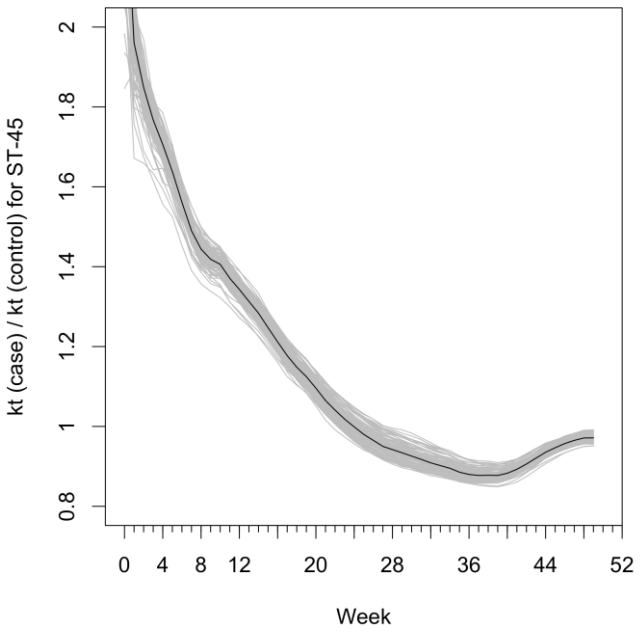


Fig 5.31 Temporal K function analysis of ST 45. The y-axis represents the k-function with values above one indicating association and values below one representing dispersal.

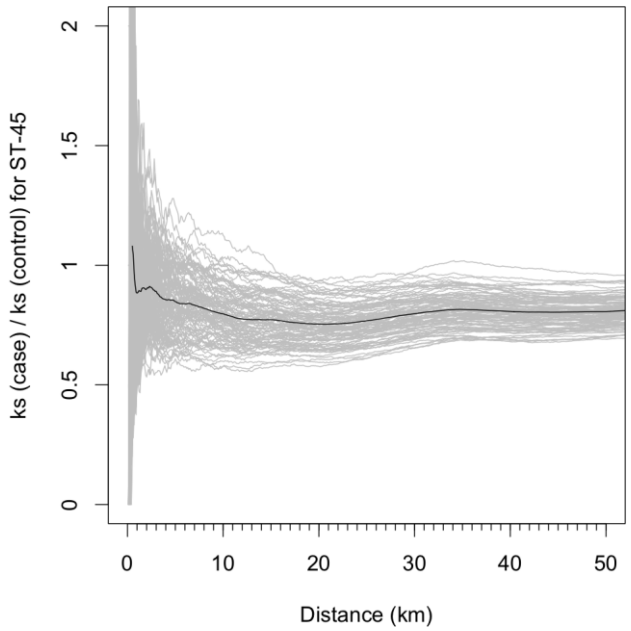


Figure 5.32 Spatial K function analysis of ST 45. The y-axis represents the k-function with values above one indicating association and values below one representing dispersal.

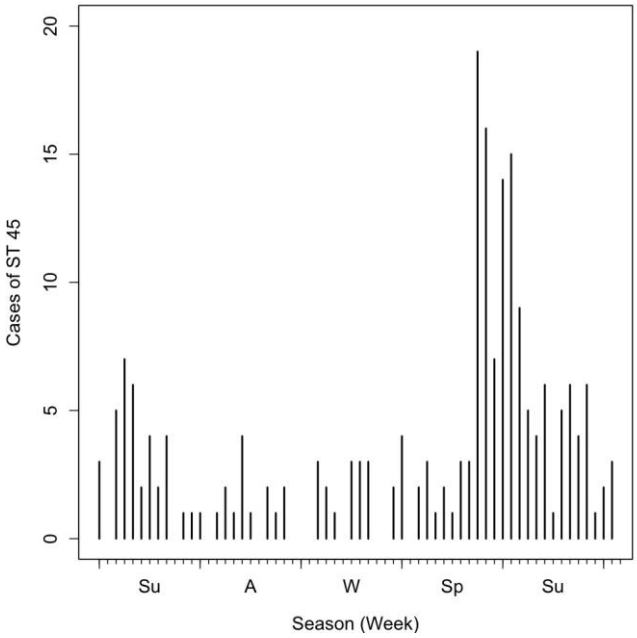


Figure 5.33 Seasonal incidence of ST 45.

5.3.4 Deprivation

Straightforward examination of the dataset shows an apparent negative trend in case rate with increasing deprivation (Fig 5.34 (d)-(f)). This result is confirmed by the risk factor model (see Table 5.10), which revealed deprivation to be a significant protective factor (relative risk = 0.965, 95% CIs = 0.959, 0.971). This result was robust to the choice of deprivation score (Carstairs, depcat or SIMD), as would be expected from the strong evident correlation between the measures (Fig 5.34 (a)-(c)).

The result was also robust to the breakdown of the data into sheep and ruminant associated sequence types, and into individuals sequence types (namely 21, 45 and 257).

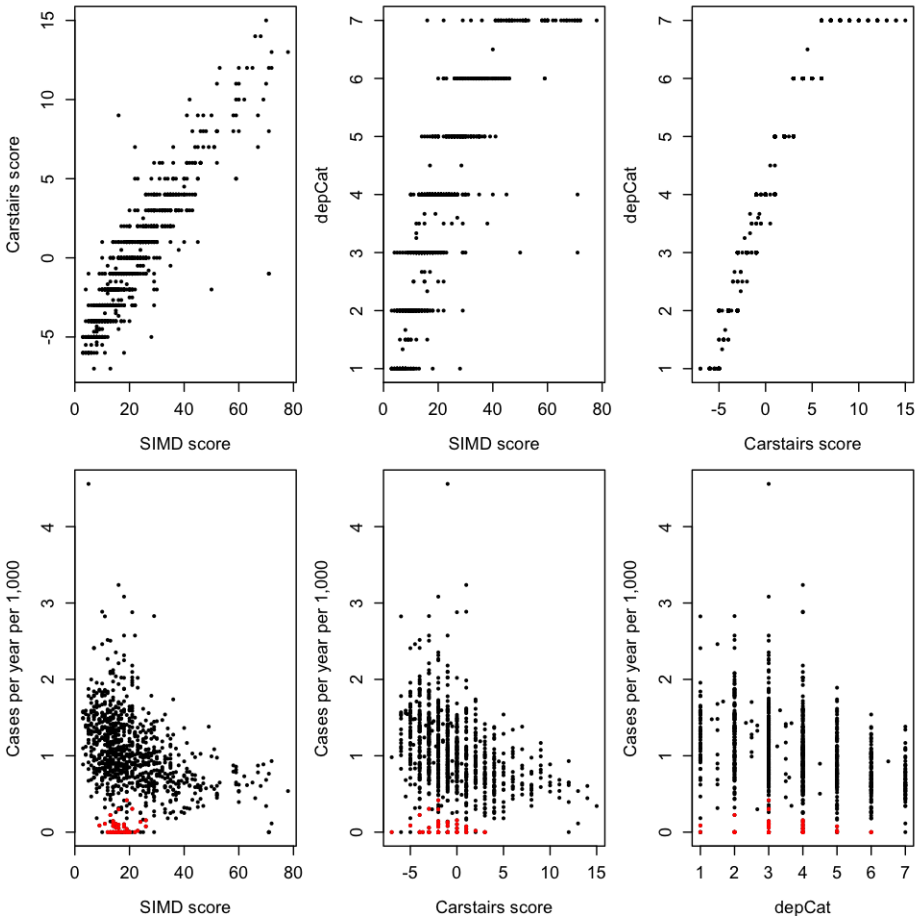


Figure 5.34 (a)-(c) Pairwise plots of the three deprivation scores, Carstairs, depCat and SIMD. (d)-(f) Case rate per thousand plotted against deprivation score. The lowest scores represent the most affluent regions and the highest the most deprived ones.

5.3.5 Foreign travel

5.3.5.1 Travel/deprivation as a risk factor

Travel abroad is a potential risk factor for *Campylobacter* infection and may also act as a proxy for deprivation. As only a proportion of the health boards supplied travel data, the data were not sufficient to quantify the potential risk of infection due to travel. However, we were able to assess whether our conclusions about the impact of deprivation were being biased by a potential higher rate of travel abroad in less deprived areas. The risk factor modelling was repeated for the Grampian, Lothian and Forth Valley health boards, with and without the cases reporting recent travel abroad. The two models did not differ in the role of deprivation as a protective factor. We conclude therefore that deprivation is not appearing as a protective factor because of reduced travel abroad amongst individuals in deprived areas.

5.3.5.2 Diversity of MLST types

The FSA funded Grampian case-control study (S01023/S14024) identified 21% (64/309) of cases associated with foreign travel and for which MLST type had been linked.

Simpson's index of diversity was 0.969 (0.958 – 0.981) for foreign travel and 0.852 (95% CI 0.793 – 0.894) for indigenous cases respectively. This higher diversity in foreign travel cases was found to be statistically significant using a randomisation test ($P < 0.0001$).

Odds ratios were calculated to determine whether particular MLST types were more common (or rare) abroad compared with indigenous cases. Odds ratios > 1.0 indicate that the type is more common in foreign cases (Table 5.13). The only type found more commonly abroad is ST 572. These isolates were associated predominantly with countries around the Mediterranean (Spain (3), Libya, Turkey and Morocco). The PubMLST database has reports of ST 572 originating from chicken (Spain, Belgium) and environmental water (Spain) as well as human isolates from the Netherlands (2002-03) and England (2004).

Types over-represented as being indigenous were ST 21, ST 45 and ST 257 (these are the three most common types in human clinical cases in Scotland). ST 257 has a strong association with retail poultry. ST 21 is associated with cattle and sheep but can also be found in retail poultry. Finally, ST 45 is the 3rd most common type found in retail poultry and wild birds and also in cattle and sheep.

Table 5.13 Odds ratios for common STs for foreign travel compared with indigenous cases.

Sequence Type (ST)	Foreign cases	Indigenous Cases	Odds Ratio[a]	95% CI	*P* value
19	2	10	0.97	0.21-4.51	0.59
21	1	34	0.13*	0.02-0.96	0.004
45	0	20	0*	-	0.01
48	1	17	0.27	0.04-2.09	0.10
51	2	15	0.63	0.14-2.83	0.34
53	1	9	0.53	0.07-4.25	0.40
257	1	35	0.12*	0.02-0.92	0.003
572	6	3	10.6*	2.57-43.4	0.002
574	2	11	0.87	0.19-4.04	0.98
827	0	12	0	-	0.08
2030	0	8	0	-	0.18

[a] OR are defined as the ratio of the odds of having the proportion of a particular type from a foreign travel cases to the odds of that type occurring in an indigenous case. 95% CI, 95% confidence interval; *, statistically significant by Fisher's exact test.

5.3.6 Household outbreaks

Household outbreaks were identified from epidemiological data where two or more cases were reported within 28 days at the same address with the same family surname. It was found that there were 94 such outbreaks (89 containing 2, 3 containing 3 and 2 containing 4 cases respectively) contained within the 15 month MLST dataset. These outbreaks comprise 3.3% of all reported *Campylobacter* cases. The randomisation test showed that 6.4 household outbreaks would be expected by chance and that the 94 obtained was significantly higher ($P<0.00001$) than would be expected by chance.

Of the 94 household outbreaks, there was MLST genotyping available in 28 of them for at least two of the isolates (Table 5.14) and of these 28 outbreaks, 23 (82%) comprised a single MLST type with the remainder comprising more than one MLST type. This is significantly higher ($P<0.00001$) than would be found by chance using a randomisation test, which would predict 0.13 of the 28 outbreaks having the same type. Further, 20 of the outbreaks had cases reporting within the one week which indicates the same source of infection.

The reporting date for these household outbreaks showed that 71% were within 1 week, 11% between 1 and 2 weeks and 18% three weeks or more. This provides some evidence that there may be some secondary infections which have been reported to be very rare (Blaser, 1997) in *Campylobacter*. However, care should be taken in interpreting these results as dates of onset, rather than reporting date as used here, would be a more accurate means of identifying secondary infections.

It has been noted that human campylobacteriosis infections are more common in males than females up to approximately 20 years of age (Strachan et al, 2008). This is not the case for women between the ages of approximately 20-30. This change has been attributed to cross infection between mother and child. (Strachan et al., 2008). In the dataset there were 11 cases which involved both an adult and a child <16 years where the reporting dates were >1 week apart. On 7 out of 11 occasions infection in the child was reported first. Of these 7 cases 6 of the adults were female and although these data are not statistically significant (binomial distribution $P=0.06$), further work could demonstrate to this being part of the reason why there is a gender ratio of approximately one for adults in their twenties.

The putative source of the types obtained in the household outbreaks as determined by source attribution is given in the final column of Table 5.14. For those 5 outbreaks with different MLST types it was found that three had the same source attribution (outbreaks 8, 24, 28) whilst 2 did not (outbreaks 13, 15).

Table 5.14. Epidemiological data for putative household outbreaks that have been genotyped by MLST.

Putative household outbreak	Age	Sex	MLST Type	Reported time difference between cases[a]	Association by source attribution (Fraction association)[b]
1	8	F	583	< 1 week	Chicken (0.99)
	6	M	583		"
2	57	F	5	< 1 week	Chicken (0.96)
	60	M	5		"
3	7	F	464	3 - 4 weeks	Chicken (0.86)
	4	M	464		"
4	5	M	21	2 - 3 weeks	Sheep (0.70)
	2	F	21		"
5	45	M	21	2 - 3 weeks	"
	16	F	21		"
6	77	M	206	< 1 week	Sheep (0.90)
	74	F	206		"
7	7	M	53	2 - 3 weeks	Sheep (0.67)
	5	M	53		"
8	42	F	1365	< 1 week	Chicken (0.99)
	1	M	2131		Chicken (0.89)
9	3	F	273	3 - 4 weeks	Sheep (0.99)
	1	M	273		"
10	16	F	21	< 1 week	Sheep (0.70)
	15	F	21		"
11	2	F	2030	< 1 week	Chicken (0.99)
	4	M	2030		"
12	12	M	2030	< 1 week	Chicken (0.99)
	10	M	2030		"
13	36	F	400	1 - 2 weeks	Chicken (0.98)
	7	M	42		Cattle (0.88)
14	39	M	75	< 1 week	Sheep (0.80)
	14	F	75		"
15	42	M	267	< 1 week	Chicken (0.99)
	41	F	962		Sheep (0.95)
16	44	F	51	< 1 week	Chicken (0.99)
	46	M	51		"
17	5	M	53	< 1 week	Sheep (0.67)
	3	M	53		"
18	21	M	61	1 - 2 weeks	Cattle (0.63)
	50	M	61		"
19	3	M	262	< 1 week	Ca (0.51)/Sh (0.48)
	38	F	262		"
	0	M	262		"
20	72	F	257	< 1 week	Chicken (0.99)
	72	M	257		"
21	41	M	257	< 1 week	"
	40	F	257		"
22	33	F	45	< 1 week	Chicken (0.96)

	6	M	45		“
23	31	M	262	< 1 week	Ca (0.51)/Sh (0.48)
	28	F	262		“
24	38	M	574	< 1 week	Chicken (0.99)
	37	F	51		Chicken (0.99)
25	82	F	52	< 1 week	Chicken (0.99)
	88	M	52		“
26	8	F	19	< 1 week	Sheep (0.73)
	4	F	19		“
27	7	M	475	1 - 2 weeks	Sheep (0.77)
	4	F	475		“
28	35	F	574	< 1 week	Chicken (0.99)
	36	M	51		“

[a]Isolates are listed for each outbreak in chronological order
[b]The source attribution method that was used is the Dutch allele model.

6. CONCLUSIONS

Our first objective was to provide a descriptive analysis of the geographic and demographic distribution of cases. This analysis shows that the incidence in males is greater than in females particularly in children <18 years and the elderly >65 (section 5.1.2), a phenomenon which has been identified in previous analyses of *Campylobacter* incidence (Gillespie et al, 2008; Strachan et al, 2008). The descriptive analysis also showed there to be apparent differences in reporting between health boards (section 5.1.4), which due to significant drops in incidence across boundaries can not plausibly be attributed to differing underlying risks of acquisition (5.3.1.1).

Formal statistical analyses, which accounted for differences in underlying risk factor between health boards, confirmed this finding, showing there to be significant under-reporting of *Campylobacter* incidence in Ayrshire and Arran and the Argyll part of Argyll and Clyde section (5.3.1.3). The reasons for the difference in reporting rates between health board are unclear but it does not appear to be due to the microbiological methods (5.3.1.2). It seems unlikely that the differences are real (ie that these regions actually have lower *Campylobacter* rates) because the changes are abrupt across the regional boundaries (5.3.1.1). The under-reporting can be attributable to any part or combination of the steps in the reporting pyramid of disease (Wheeler et al., 1999), for example the likelihood of an individual presenting to general practice, the likelihood of submission of a sample to diagnostic lab and the likelihood of a laboratory correctly identifying a positive case. Our analyses suggest that the differences are not due to variation in laboratory protocols, but further study would be needed to identify the reasons for the under reporting observed here.

The second objective was to link epidemiological and typing data to develop statistical models that best describe the spatial distribution of cases and expose the underlying epidemiology of the disease. There is an elevated incidence in campylobacteriosis in rural children (section 5.3.2.2), which coroborates the findings of previous studies (Miller et al, 2004b; O'Neill et al, 2004). The availability of MLST data in this project allowed attribution of human infections to potential sources or reservoirs of infection. These source attribution methods suggest that relative to poultry, ruminant sources are more common in children, more frequently acquired in summer, are associated with the presence of sheep (section 5.3.3.4).

Analyses of the spatial distribution of ruminant and poultry attributed strains showed ruminant strains to be relatively dispersed in space reflecting occurrence in sparse rural populations, whereas poultry strains are more spatially clustered suggesting that they occur as outbreaks (section 5.3.3.4). Though there are broad overall differences between poultry and ruminant strains, there are also differing epidemiologies at the resolution of individual STs: ST45 shows much more seasonal peak than ST21 (both poultry attributed), with ST45 being more spatially dispersed than ST21, which is consistent with the expected epidemiology of a water-borne strain (Sopwith et al, 2008).

Our analyses suggest that though these types may dominant in Scotland, this is unlikely to be the case in other countries. The higher diversity of MLST type associated with foreign travel cases is suggestive of heterogeneity of MLST type between countries (section 5.3.5.2). This is in contrast to the homogeneity of types found between the health boards of Scotland (Sheppard et al, 2009). Analysis of the common types found from both foreign and indigenous cases shows that there are differences in the proportion of MLST types found (Table 5.13). This provides evidence to suggest that particular MLST types (e.g. ST 257, ST21 and ST45) are more common in Scotland (and perhaps elsewhere in the UK) than abroad. The reason why this is the case is not known and is worthy of investigation.

SatScan can be used to identify clusters and was able to detect the Montrose outbreak

(section 5.2). However, to identify household outbreaks a resolution higher than postal sector is required. SatScan appears not to be suitable for detecting diffuse spatial outbreaks with MLST data. However, inclusion of additional loci such as fla may improve the power of analysis.

It is interesting to note that the results here, showing that 3.3% of cases are in household outbreaks (section 5.3.6), are in agreement with those found in a region of Wales (5%) (Ribeiro and Frost 2000) and in Denmark (3.3%) (Ethelberg et al, 2004). Further, in the Welsh family outbreak study, which employed a combination of sero- and phage-typing, it was found that 65% of outbreaks contained identical isolates compared with 82% using the MLST method reported here. It is possible that some of the 5 putative family outbreaks found in the current study that did not have the same MLST type were caused by co-infection with different strains from the same source (e.g. food vehicle) as has previously been reported elsewhere (Forbes et al, 2009).

The effect of private water supplies could not be ascertained by this study, as the number of private water supplies per person was closely correlated with population density. Higher population density was identified as a significant protective factor, but may be acting as proxy for for some qualities of rural environments, such as exposure to all environmental reservoirs, including private water supplies and ruminant sources of infection.

The availability of MLST data in this project allowed attribution of human infections to potential sources or reservoirs of infection (section 5.3.3). Both the proportional similarity index and attribution models showed that for *C. jejuni* human isolates were most similar to retail chicken, though the signal for ruminants is also important. For *C. coli*, sheep and retail chicken isolates were most similar to humans whilst the contribution from pigs was negligible in comparison. These source attribution results for both the urban and rural human populations demonstrate that retail chicken is the most important source. However the data suggests, particularly for children, that retail chicken has a reduced importance in rural areas whereas this is contrasted with an increase in the importance of ruminant sources. Conversely urban children may be less exposed to the environmental route via ruminants and more likely to acquire infection via retail chicken consumption.

Deprivation was an important risk factors for all *Campylobacter* strains whether measured using Carstairs, Depcat or SIMD, with increased deprivation corresponding to reduced rate of *Campylobacter* reports (section 5.3.4). The observed protective effect of deprivation could be due to real differences in rates of infection or due to differences in ascertainment. These results are in line with findings from other countries for both *Campylobacter* and other gastrointestinal diseases (Simonsen et al., 2008; Snel et al., 2009a). A number of potential explanations have been offered for the relationship with deprivation

1. Acquired immunity through exposure amongst the more deprived. However, the age-stratified analysis (section 5.3.3.4) demonstrates that there is no significant difference in Carstairs deprivation score in different age-groups. If acquired immunity were the explanation then the younger groups would be more commonly infected in more deprived areas whilst older age groups would be more commonly infected in less deprived areas, but our analysis shows no difference. These findings are supported by other studies that suggest that there is no difference between age and deprivation (Simonsen et al., 2008; Snel et al., 2009a).
2. Deprivation may be associated with differences in dietary habits (Simonsen et al., 2008); differences in the quality of the available fresh food have been observed elsewhere (Cummins et al., 2009). If there is greater consumption of processed rather than fresh meat among the more deprived there will be less *Campylobacter* because the process of freezing reduces the number of *Campylobacter* organisms (Ritz et al., 2007). Furthermore, the less deprived may also eat at restaurants more

frequently, which has been demonstrated as a risk factor in other studies (Danis et al., 2009).
3. Differences in environmental exposure associated with different leisure activities and differences in access to rural areas.
4. Differences in reporting. Lower reporting rates for gastrointestinal disease among the more deprived have been noted in the UK (Olowokure et al., 1999; Snel et al., 2009c, b), Denmark (Simonsen et al., 2008) and New Zealand (Snel et al., 2009c, b).

We also examined the possibility that lower rates of foreign travel in deprived areas might be responsible for the protective effect (section 5.3.5.1). However, analyses on those health boards for which available data allowed us to exclude recent travel cases still showed deprivation to be a significant protective effect. Further research is necessary to fully understand the process operating, however, it is likely that some combination of these factors is responsible for the relationship with deprivation.

Whilst one of the greatest sources of *Campylobacter* in rural areas is thought likely to be livestock (Brown et al., 2004; French et al., 2005; Horrocks et al., 2009), our analysis did not show density of livestock alone to be associated with *Campylobacter* infections. Rather, this model suggests that rurality in general is a risk for infection. These findings suggest that environmental exposures, whilst these may ultimately be the result of contamination from livestock sources, are best characterised by low population densities (see Table 5.10).

This study has demonstrated that there are real differences in the geographic distribution of *Campylobacter* infections within Scotland caused by differences in exposure to infection. Those at greatest risk are children in rural environments. As argued above, we suggest that the relationship with deprivation is unlikely to result from differences in acquired immunity. However large differences remain in ascertainment between the deprived and the less deprived as well as differences between the authorities administering the health care. Accurate quantification of the level of under reporting would aid further studies.

7. RECOMMENDATIONS

Recommendations "for bodies to act"

1. A public health education campaign should be targeted at young rural children and their parents to focus on the hazards of *Campylobacter* as well as other relevant GI pathogens.
2. With the recent sharp rise in *Campylobacter* infections and the potential of intervention strategies being put in place by the FSA in the broiler industry it is important that routine baseline data on clinical isolates and representative animal/food species are collected. These data should take the form of sequence type as well as basic epidemiological information on human isolates. From this, it would then be possible to infer the efficacy of intervention strategies and investigate reasons for the changes in incidence across the human population.
3. The results from this project together with the CAMPS FSA project (S14006) clearly show that retail chicken as well as ruminants are important sources of human campylobacteriosis. The FSA is already communicating with the poultry industry about this problem but this needs to be accelerated. It is also important that meetings with sheep and cattle farmers and their representing organisations are held to discuss the problem.
4. The results suggest that pigs are not an important source of human *Campylobacter* infection and hence effort should be placed in other areas.
5. A very small percentage of *Campylobacter* infections are attributable to community and household outbreaks. There is no point spending extra effort in this area over and above what is currently done routinely by the health boards.
6. A meeting of health board officials should be convened to put a plan of action in place to resolve the reporting bias between health boards.

Future Work

- Further work needs to be carried out to try and explain why different ST's are more prominent in particular countries. Is it the distribution driven by chance, or do particular environmental and management conditions select for different ST's.
- Further work is required to determine whether the fewer cases reported in the deprived population is actually a true reflection of the disease incidence, an artefact of reporting, or a signature of differential health care uptake by deprived communities. Investigating whether the types associated with these populations are different would give an indication if exposure may be different. Further, quantifying whether the rate of disease severity is similar or otherwise would also shed light on this issue.
- The role played by acquired immunity in determining patterns of incidence remains unresolved. Could some types of increased exposure act to increase immunity and reduce incidence. Immunity and exposure studies would shed light on this issue.

8. REFERENCES

ACMSF (2004) Second Report on *Campylobacter*. Available at http://www.food.gov.uk/multimedia/pdfs/acmsfcampyloreport.pdf accessed 8th June 2004.

Adak G K, Cowden J M, Nicholas S, Evans H S. The Public Health Laboratory Service national case-control study of primary indigenous sporadic cases of *Campylobacter* infection. Epidemiol Infect 1995; 115(1) : 15-22.

Bivand, R. (2009) spdep: Spatial dependence: weighting schemes, statistics and models

Blaser MJ. Epidemiologic and clinical features of *Campylobacter jejuni* infections. J Infect Dis 1997; 176: S103-5.

Brown, P.E., Christensen, O.F., Clough, H.E., Diggle, P.J., Hart, C.A., Hazel, S., Kemp, R., Leatherbarrow, A.J., Moore, A., Sutherst, J., Turner, J., Williams, N.J., Wright, E.J., French, N.P., 2004. Frequency and spatial distribution of environmental Campylobacter spp. Appl Environ Microbiol 70, 6501-6511.

Carstairs, V., Morris, R., 1990. Deprivation and health in Scotland. Health Bull (Edinb) 48, 162-175.

Cowden, J (1992) Campylobacter: epidemiological paradoxes – The Vehicles for most cases of infection remain unknown. British Medical Journal, 305, 132-133.

Diggle, P.J., Chetwynd, A.G., Haggkvist, R., Morris, S.E., 1995. Second-order analysis of space-time clustering. Statistical Methods and Methodological Research 4, 124-136.

Ethelberg, S., Olsen, K. E. P., Gerner-Smidt, P. & Molbak, K. 2004 Household outbreaks among culture-confirmed cases of bacterial gastrointestinal disease. American Journal of Epidemiology, 159, 406-412.

Food Standards Agency. UK-wide survey of *Salmonella* and *Campylobacter* contamination of fresh and frozen chicken on retail sale. 2003.
http://www.food.gov.uk/multimedia/pdfs/campsalmsurvey.pdf (accessed 7th June 2004).

Forbes, K. J., Gormley, F. J., Dallas, J. F., Labovitiadi, O., MacRae, M., Owen, R. J., Richardson, J., Strachan, N. J., Cowden, J. M., Ogden, I. D. & McGuigan, C. C. 2009 Campylobacter immunity and coinfection following a large outbreak in a farming community. J Clin Microbiol, 47, 111-6.

French, N., Barrigas, M., Brown, P., Ribiero, P., Williams, N., Leatherbarrow, H., Birtles, R., Bolton, E., Fearnhead, P., Fox, A., 2005. Spatial epidemiology and natural population structure of Campylobacter jejuni colonizing a farmland ecosystem. Environ Microbiol 7, 1116-1126.

French, N.P. 2008 Enhancing Surveillance of Potentially Foodborne Enteric Diseases in New Zealand: Human Campylobacteriosis in the Manawatu, 1-56

Friedman C, Reddy S, Samual M, Marcus R, Bender J, Desai S, Shiferaw B, Helfrick D, Carter M, Anderson B, Hoekstra M and the EIP Working Group 2000 Risk factors for sporadic *Campylobacter* infections in the United States: A case Control Study on the FoodNet Sites. 2nd International Conference on Emerging Infectious Diseases, Atlanta, GA, July. Available at http://www.cdc.gov/foodnet/pub/iceid/2000/friedman_c.htm accessed 7th June 2004.

Gillespie, I. A., O'Brien, S. J., Penman, C., Tompkins, D., Cowden, J., Humphrey, T. J. & Campylobacter Sentinel, S. 2008 Demographic determinants for Campylobacter infection in England and Wales: implications for future epidemiological studies. Epidemiology and Infection, 136, 1717-1725.

Harris N V, Weiss N S, Nolan C M. The role of poultry and meats in the etiology of *Campylobacter jejuni/coli* enteritis. Am J Public Health 1986; 76(4) : 407-411.

Horrocks, S.M., Anderson, R.C., Nisbet, D.J., Ricke, S.C., 2009. Incidence and ecology of Campylobacter jejuni and coli in animals. Anaerobe 15, 18-25.

Ikram R, Chambers S, Mitchell P, Brieseman M A, Ikam O H. A case control study to determine risk factors for *Campylobacter* infection in Christchurch in the summer of 1992-3. N Z Med J 1994; 107(988) : 430-432.

Kapperud G, Skjerve E, Bean N H, Ostroff S M, Lassen J. Risk factors for sporadic *Campylobacter* infections: results of a case-control study in southeastern Norway. J Clin Microbiol 1992; 30(12) : 3117-3121.

Kulldorff, M., Heffernan, R., Hartman, J., Assuncao, R. & Mostashari, F. 2005 A space-time permutation scan statistic for disease outbreak detection. Plos Medicine, 2, 216-224.

Lewin-Koh, N. J. & Bivand, R. 2009 maptools: Tools for reading and handling spatial objects.

Lighton L L, Kaczmarski E B, Jones D M. A study of risk factors for *Campylobacter* infection in late spring. Public Health 1991; 105(3) : 199-203.

Manly, B. F. J. 2007 Comments on a note on permutation tests of significance for multiple regression coefficients by long, et al. Psychological Reports, 101, 1041-1042.

Martino S, Rue H. 2009 INLA: Functions which allow to perform a full Bayesian analysis of structured additive models using Integrated Nested Laplace Approximaxion.

McCarthy, N. D., F. M. Colles, K. E. Dingle, M. C. Bagnall, G. Manning, M. C. J. Maiden, and D. Falush. 2007. Host-associated genetic import in Campylobacter jejuni. Emerg.Infect.Dis. 13:267-272.

Miller G, Dunn GM, Smith-Palmer A, Ogden ID, and Strachan NJC (2004a) Human Campylobacteriosis in Scotland: seasonality, regional trends and bursts of infection. Epidemiology and Infection; 132 : 185-593.

Miller G, Strachan NJC, Dunn GM, Ogden ID, Howie H and Reid TMS (2004b) Is the high morbidity of gastrointestinal zoonotic disease in young children associated with living in rural areas? Health Protection Agency, Annual Conference 13-15 September 2004, Warwick, UK.

Miller G, Dunn GM, Reid TMS, Ogden ID, and Strachan NJC (2005) Does age acquired immunity confer selective protection to common serotypes of *Campylobacter jejuni*? BMC Infectious Diseases; 5(66)

Neal K R, Slack R C. Diabetes mellitus, anti-secretory drugs and other risk factors for *Campylobacter* gastro-enteritis in adults : a case-control study. Epidemiol Infect 1997;119(3) : 307-311.

O'Brien S J, Elson R, Gillespie I A, Adak G K, Cowden J M. Surveillance of foodborne

outbreaks of infectious intestinal disease in England and Wales 1992-1999: contributing to evidence-based food policy ? Public Health 2002; 116(2) : 75-80.

O'Neill B, Ejidokun OO and Evans MR (2004) Campylobacter infection in Gloucestershire: an ecological study. Health Protection Agency, Annual Conference 13-15 September 2004, Warwick, UK.

Pires, S. M., Evers, E. G., van Pelt, W., Ayers, T., Scallan, E., Angulo, F. J., Havelaar, A., Hald, T. & Med-Vet-Net Workpackage, W. (2009) Attributing the Human Disease Burden of Foodborne Infections to Specific Sources. Foodborne Pathogens and Disease, 6, 417-424.

Pollock K, Locking M, Browning L, Smith-Palmer A, Brownlie S. Gastro-intestinal and foodborne infections: laboratory reports for common bacterial, protozoal and viral infections 2009. *HPS Wkly Rep* 2010; **44:** 38–41.

Pritchard,J.K., Stephens,M. and Donnelly,P. Inference of population structure using multilocus genotype data Genetics, 2000, 155, 2, 945-959

R Development Core Team, 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ribeiro,C.D.; Frost,J.A. 2000 Family clusters of campylobacter infection. Commun. Dis. Public Health, 2000, 3, 4, 274-276

Ripley, B.D., 1976. The second-order analysis of stationary point processes. Journal of Applied Probability 13, 255-266.

Rodrigues L C, Cowden J M, Wheeler J G, Sethi D, Wall P G, Cumberland P *et al*. The study of infectious intestinal disease in England: risk factors for cases of infectious intestinal disease with *Campylobacter jejuni* infection. Epidemiol Infect 2001; 127(2) :185-193.

Rosef, O., Kapperud, G., Lauwers, S. & Gondrosen, B. 1985 Serotyping of Campylobacter-jejuni, campylobacter-coli and campylobacter-laridis from domestic and wild animals. Applied and Environmental Microbiology, 49, 1507-1510.

Rowlingson, B., Diggle, P.J., 2008. splancs: Spatial and Space-Time Point Pattern Analysis.

Rue H, Martino S, Chopin N. 2009 Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J R Stat Soc Ser B-Stat Methodol*, 71:319-392.

Said B, Wright F, Nichols GL, Reache M and Rutter M (2003) Outbreaks of infectious disease associated with private drinking water supplies in England and Wales 1970–2000. Epidemiology and Infection, 130, 469-479.

Sheppard, S.K., Dallas, J.F., Strachan, N.J., MacRae, M., McCarthy, N.D., Wilson, D.J., Gormley, F.J., Falush, D., Ogden, I.D., Maiden, M.C., Forbes, K.J., 2009. Campylobacter genotyping to determine the source of human infection. Clin Infect Dis 48, 1072-1078.

Simonsen, J., Frisch, M., Ethelberg, S., 2008. Socioeconomic risk factors for bacterial gastrointestinal infections. Epidemiology 19, 282-290.

Skirrow MB 1977 *Campylobacter* enteritis: a "new" disease. Br Med J 2, 9-11.

Sopwith, W., A. Birtles, M. Matthews, A. Fox, S. Gee, M. Painter, M. Regan, Q. Syed, and E.

Bolton. 2008. Identification of potential environmentally adapted *Campylobacter jejuni* strain, United Kingdom. Emerg. Infect. Dis. 14:1769-1773

Snel, S.J., Baker, M.G., Kamalesh, V., French, N., Learmonth, J., 2009a. A tale of two parasites: the comparative epidemiology of cryptosporidiosis and giardiasis. Epidemiol Infect, 1-10.

Snel, S.J., Baker, M.G., Venugopal, K., 2009b. The epidemiology of cryptosporidiosis in New Zealand, 1997-2006. N Z Med J 122, 47-61.

Snel, S.J., Baker, M.G., Venugopal, K., 2009c. The epidemiology of giardiasis in New Zealand, 1997-2006. N Z Med J 122, 62-75.

Strachan, N. J., Watson, R. O., Novik, V., Hofreuter, D., Ogden, I. D. & Galan, J. E. 2008 Sexual dimorphism in campylobacteriosis. Epidemiol Infect, 136, 1492-5.

Studahl A, Andersson Y. Risk factors for indigenous *Campylobacter* infection : a Swedish case-control study. Epidemiol Infect 2000; 125(2) : 269-275.

The Campylobacter Sentinel Surveillance Scheme Collaborators (2003) Foreign and Domestic Travel and the Risk of Campylobacter Infection: Results from a Population Based Sentinel Surveillance Scheme. Journal of Travel Medicine, 10(2), 136-139.

Wheeler JG, Sethi D, Cowden JM, Wall PG, Tompkins DS, Hudson MJ, Roderick PJ on behalf of the Infectious Intestinal Disease Executive (1999) British Medical Journal, 318, 1046-1050.

9. PUBLICATIONS

Bessell, P. R., Matthews, L., Smith-Palmer, A., Rotariu, O., Strachan, N.J.C., Forbes, K.J., Cowden, J.M., Reid, S.W.J and Innocent, G.T. Geographic determinants of reported human Campylobacter infections in Scotland. To appear in BMC Public Health.

MLST typing data facilitate comparative epidemiological analysis of different strains, enabling different transmission routes and their risks to be identified. Bessell, P. R., Matthews, L., Innocent, G.T., Smith-Palmer, A., Rotariu, O., Strachan, N.J.C., Forbes, K.J., Cowden, J.M. and Reid, S.W.J. In preparation.

Rotariu, O., J. F. Dallas, I. D. Ogden, M. MacRae, S. K. Sheppard, M. C. J. Maiden, F. J. Gormley, K. J. Forbes, and N. J. C. Strachan. 2009. Spatiotemporal homogeneity of *Campylobacter* subtypes from cattle and sheep across NE and SW Scotland. Appl.Environ.Microbiol. 75:6275-6281.

Strachan, N. J. C., F. J. Gormley, O. Rotariu, I. D. Ogden, G. Miller, G. M. Dunn, S. K. Sheppard, J. F. Dallas, T. M. S. Reid, H. Howie, M. C. J. Maiden, and K. J. Forbes. 2009. Attribution of *Campylobacter* infections in northeast Scotland to specific sources using multi-locus sequence typing (MLST). Journal of Infectious Diseases 199:1205-1208.

O. Rotariu, A. Smith-Palmer, J. Cowden, P. R. Bessell, G. T. Innocent, S. W. J. Reid, L. Matthews, J. Dallas, I. D. Ogden, K. J Forbes, N. J. C. Strachan Putative household outbreaks of campylobacteriosis typically comprise single MLST genotypes.Submitted to Epidemiology and Infection

Conference Presentations and Invited Talks

Dallas, J. F., Rotariu, O., Ogden, I. D, Gormley, F. J., Reay, K. W., MacRae, M., Sheppard, S. K., Strachan, N. J. C., and Forbes, K. J. Temporal clustering of clinical *Campylobacter* strains across Scotland during 2005-06. 15th Into Workshop on Campylobacter, Helicobacter and Related Organisms (CHRO-09). 15th Into Workshop on Campylobacter, Helicobacter and Related Organisms (CHRO-09) , 49. 2009. Niigata, Japan. 2009.

Strachan NJC et al Burden of Infectious Diseases Meeting, Tokyo, Japan, August 2009. Invited presentation on Campylobacter source attribution. A Scottish perspective.

Strachan NJC et al European Food Safety Authority, VLA, UK June 2009. Invited presentation to Broiler working group on attribution of *Campylobacter*.

Strachan NJC et al Food Standards Agency, Edinburgh, UK, June 2009. Invited presentation on the Geography of Campylobacter infections in Scotland.

Strachan NJC et al New Zealand Food Safety Authority, October 2009. Invited presentation on Microbial subtyping and sentinel sites, Scotland.

Posters

FoodMicro 2010, Copenhagen, Denmark August 2010, The changing *Campylobacter* epidemic. Poster/oral submission. Norval J. C. Strachan, Ovidiu Rotariu, Paul R. Bessell, Louise Matthews, Alison Smith-Palmer, John Cowden, Iain Ogden and Ken J. Forbes.

CHRO 2007, Rotterdam, The Netherlands, August 2007. Applying host attribution to quantify the role of chicken and the environment as a source of human campylobacteriosis. NJC

Strachan, O Rotariu, JF Dallas, ID Ogden, PR Bessell, GT Innocent, A Smith-Palmer, JM Cowden, SWJ Reid, L Matthews, SK Sheppard, MCJ Maiden, and KJ Forbes.

10. APPENDIX A

Carstairs and DEPCAT indices

Carstairs and DEPCAT indices were obtained from MRC Social & Public Health Sciences Unit (Carstairs scores for Scottish postcode sectors from the 2001 Census, Philip McLoone's report, University of Glasgow, http://www.msoc-mrc.gla.ac.uk/). They represent socio-economical indices for the Scottish population that were derived by combining selected variables (e.g. number of cars owned per household, male unemployment, overcrowding etc.) to generate indices at postcode sector level. The Carstairs and DEPCAT scores range between -7 to +15 and 1 to 7 respectively. The lowest scores represent the most affluent regions of Scotland and the highest the most deprived ones. The distribution of these two descriptors are presented in Figure A.1 .

Scottish Index of Multiple Deprivation (SIMD)

The Scottish Index of Multiple Deprivation (SIMD) (Fig A.1(c)) was obtained from the Scottish Government (http://www.scotland.gov.uk/Topics/Statistics/SIMD) and is an equivalent to the Carstairs and DEPCAT indices. However, it was derived at a smaller spatial resolution (6,505 data zones) to allow the identification of small area concentrations of multiple deprivation across Scotland. The calculation of SIMD index is based on 37 indicators from seven socio-economic domains: Current Income, Employment, Health, Education Skills and Training, Geographic Access to Services (including public transport travel), Housing and Crime. The index ranges from 0 to 89 and its average at postcode sector level results in values varying between 0 to 78 (0 - most affluent, 78 - most deprived).

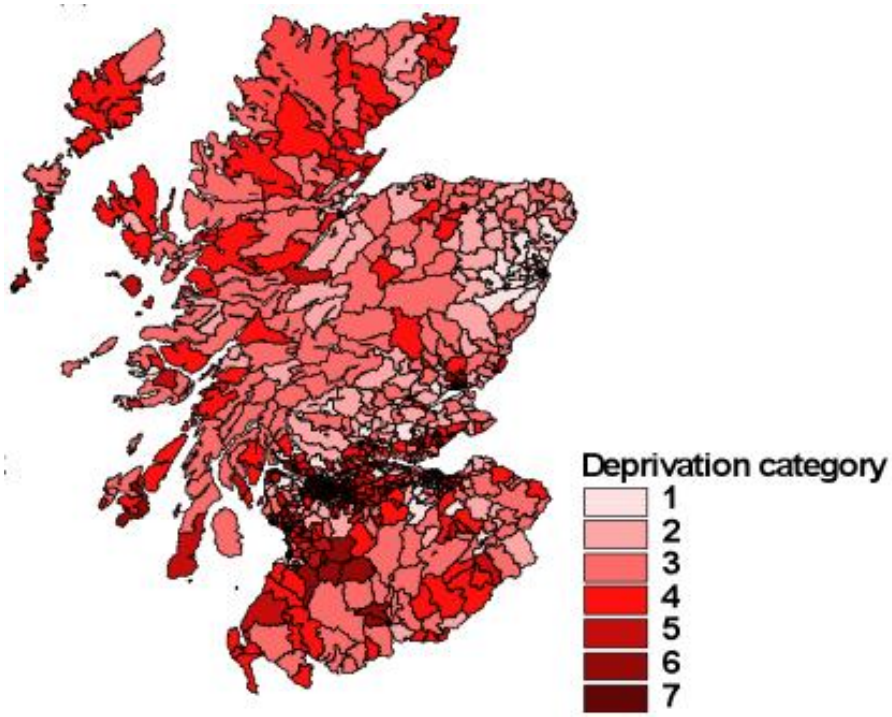
Mosaic Scotland Groups and Types

Mosaic Scotland (Experian Ltd, UK, http://strategies.experian.co.uk) provided a detailed socio-economic statistic about people living in Scotland, at postcode level, covering 2.2 million households. It classifies each Scottish postcode into one of 10 groups and 44 types (Table A.1). The percentage of people belonging to a specific group or type was integrated up to postcode sector level. As an example Figure A.1(d) presents the distribution of the "Country Lifestyles" group across Scotland.

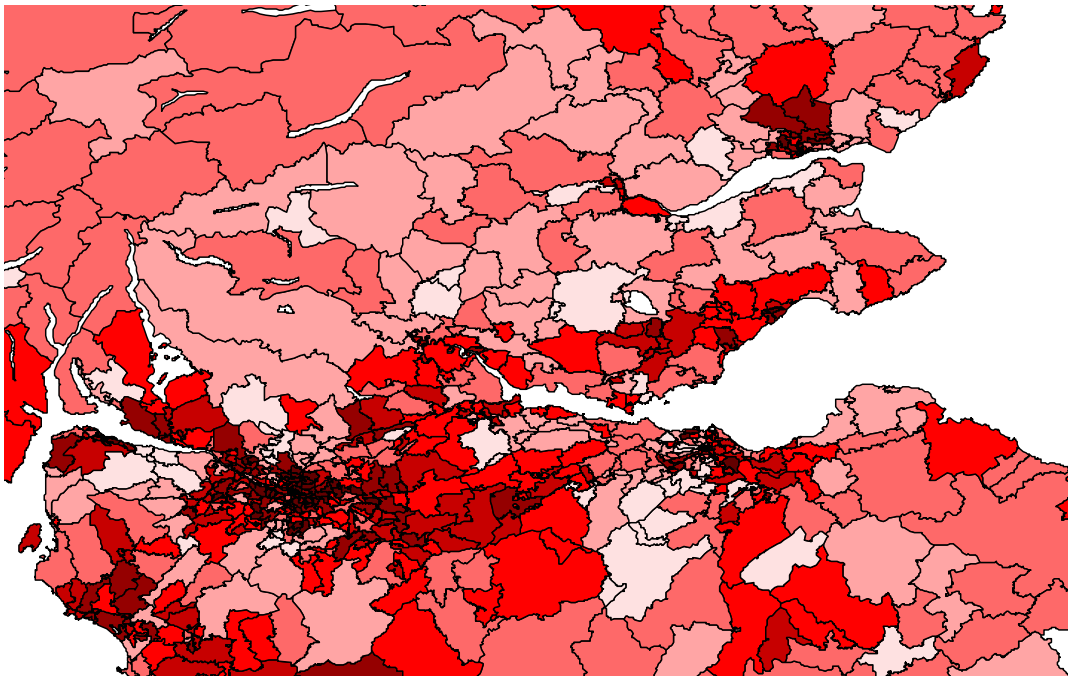
Table A.1 Mosaic Scotland Type, Group codes and names

Group		Type	
01	A Upper Echelons	**01**	A01 Captains of Industry
		02	A02 Wealth of Experience
		03	A03 New Influentials
02	B Families on the Move	**04**	B04 Successful Managers
		05	B05 White Collar Owners
		06	B06 Emerging High Status
		07	B07 New Suburbanites
		08	B08 Settling In
		09	B09 Military Might
03	C Small Town Propriety	**10**	C10 Songs of Praise
		11	C11 Ageing in Suburbia
		12	C12 Blue Collar Owners
04	D Country Lifestyles	**13**	D13 Towns in Miniature
		14	D14 Rural Playgrounds
		15	D15 Agrarian Heartlands
		16	D16 Isolated Farmsteads
		17	D17 Scenic Wonderland
		18	D18 Far Away Islanders
05	E Urban Sophisticates	**19**	E19 Prestige Tenements
		20	E20 Studio Singles
		21	E21 Rucksack and Bicycle
		22	E22 College and Campus
		23	E23 Inner City Transience
		24	E24 Cosmopolitan Chic
06	F Town Centre Singles	**25**	F25 Tenement Lifestyles
		26	F26 Downtown Flatlets
		27	F27 30 Something Singles
		28	F28 Small Town Pride
		29	F29 Dignified Seniors
07	G Renters Now Owning	**30**	G30 Sought after Schemes
		31	G31 Rustbelt Renaissance
		32	G32 Planners Paradise
08	H Low Income Families	**33**	H33 Smokestack Survivors
		34	H34 Quality City Schemes
		35	H35 Lathe and Loom
		36	H36 Indebted Families
09	I State Beneficiaries	**37**	I37 Pockets of Poverty
		38	I38 Mid Rise Breadline
		39	I39 Room and Kitchen
		40	I40 Families in the Sky
10	J Shades of Grey	**41**	J41 Elders 4-in-a-Block
		42	J42 Greys in Small Flats
		43	J43 Skyline Seniors
		44	J44 Twilight Infirmity
99	Unclassified	**99**	Unclassified
00	Not in Mosaic Scotland Area	**00**	Not in Mosaic Scotland Area

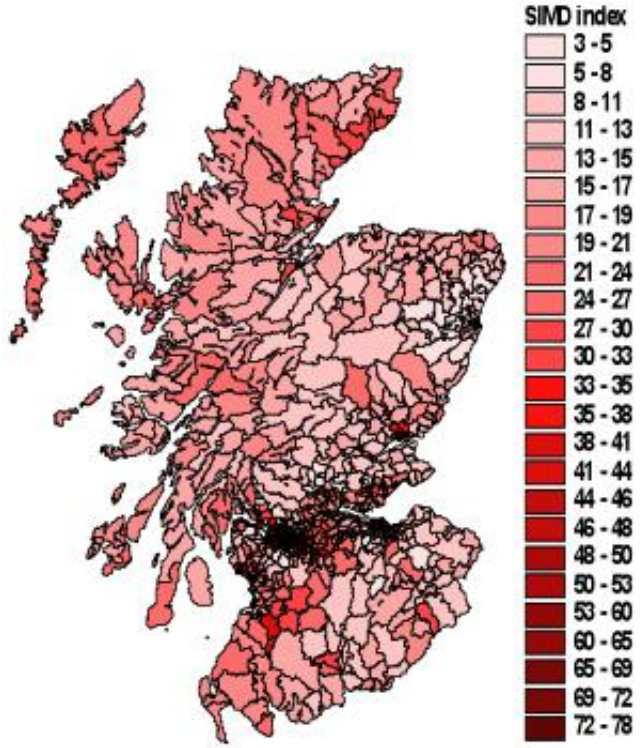
(a)



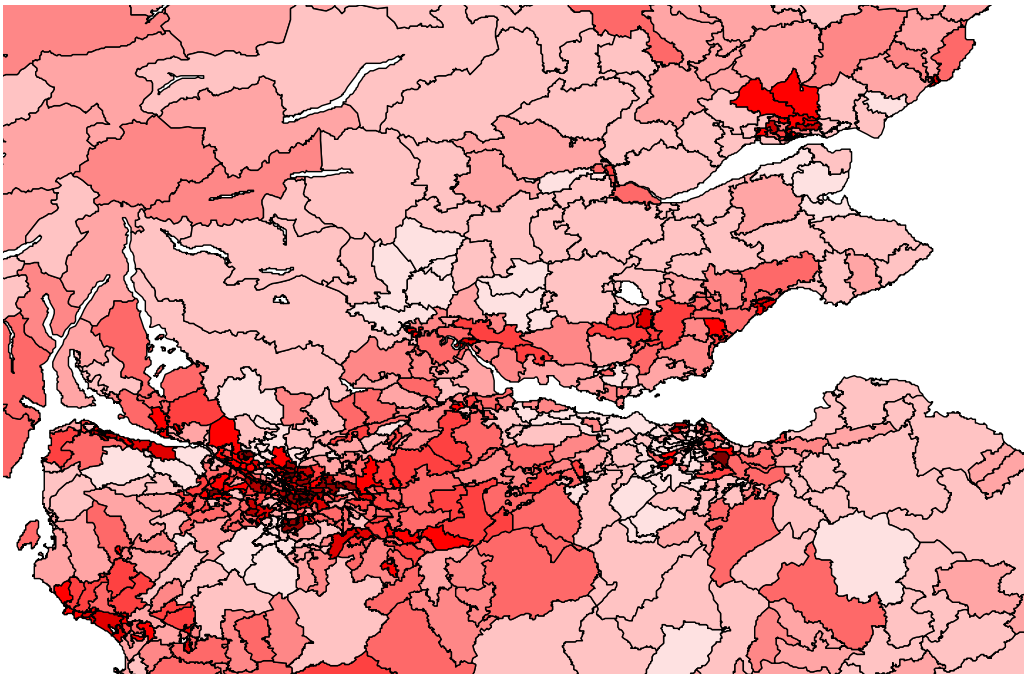
(b)



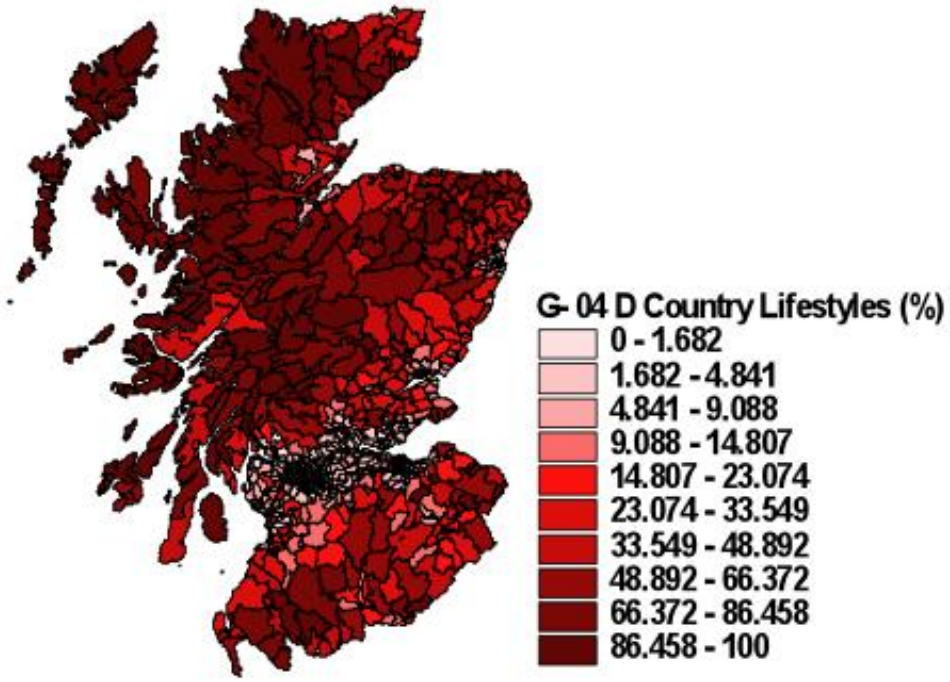
(c)


SIMD index
3 - 5
5 - 8
8 - 11
11 - 13
13 - 15
15 - 17
17 - 19
19 - 21
21 - 24
24 - 27
27 - 30
30 - 33
33 - 35
35 - 38
38 - 41
41 - 44
44 - 46
46 - 48
48 - 50
50 - 53
53 - 60
60 - 65
65 - 69
69 - 72
72 - 78

(d)



(e)



(f)

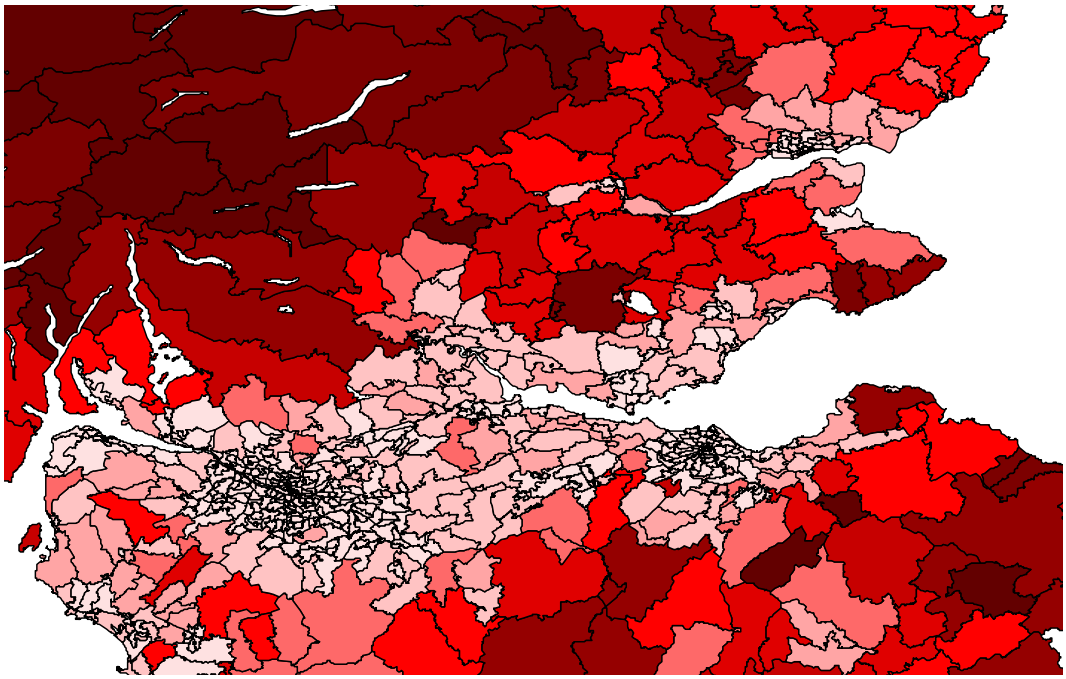


Figure A.1 Distribution of socioeconomic indices across mainland Scotland and central belt at postcode sector level: (a-b) DEPCAT, (c-d) SIMD index and (e-f) Mosaic "Country Lifestyles" group.

Figure A.2 Example output from space time analysis by SatScan (2000-06 case dataset) [a]

SatScan v8.0.1 Results

Program run on: Thu Dec 04 16:56:15 2008

Retrospective Space-Time analysis
scanning for clusters with high rates
using the Poisson model.

SUMMARY OF DATA

Study period.....: 2000/1/5 - 2006/12/29
Number of locations.....: 936
Total population.....: 5062049
Total number of cases....: 34173
Annual cases / 100000....: 96.7

MOST LIKELY CLUSTER
1.Location IDs included.: DD96, DD97, DD109 (Montrose outbreak part)
  Coordinates / radius..: (359227,760378) / 11598.32
  Time frame.....: 2005/12/1 - 2005/12/7
  Population.....: 18505
  Number of cases.....: 18
  Expected cases.....: 0.34
  Annual cases / 100000.: 5075.4
  Observed / expected...: 52.510
  Relative risk.....: 52.537
  Log likelihood ratio..: 53.645300
  Monte Carlo rank.....: 1/1000
  P-value.....: 0.001

SECONDARY CLUSTERS
2.Location IDs included.: PA134
  Coordinates / radius..: (235471,669855) / 0.00
  Time frame.....: 2000/7/20 - 2000/7/25
  Population.....: 4456
  Number of cases.....: 11
  Expected cases.....: 0.07
  Annual cases / 100000.: 15027.2
  Observed / expected...: 155.471
  Relative risk.....: 155.521
  Log likelihood ratio..: 44.583566
  Monte Carlo rank.....: 1/1000
  P-value.....: 0.001

---

[a] Only the two most likely clusters are presented