i-CaMPS
impact of interventions - *Campylobacter* MLST Project in Scotland

Division of Applied Medicine
School of Medicine and Dentistry
University of Aberdeen
Foresterhill
Aberdeen AB25 2ZD
United Kingdom

Tel: +44 (0)1224 554953
Email: CaMPS@abdn.ac.uk

Food Standards Agency- Scotland
Contract S14054

"Employing Source Attribution and Molecular Epidemiology to measure the impact of interventions on human campylobacteriosis in Scotland"

Final Report
24 November 2011

Contents

1. Introduction 1
1.1 Background 1
1.2 The molecular epidemiology of campylobacteriosis in Scotland can be modeled using Grampian data 3
1.3 Aims 5
2. Materials and Methods 6
2.1 Isolates Collections 6
2.2 MLST of isolates 6
2.3 Host reservoir isolate datasets 6
2.4 Molecular attribution methods 8
3. Results and Discussion 11
3.1 Has the prevalence of *Campylobacter* in food and animal reservoirs changed over time? 11
3.2 Do strain types change over time? 11
3.3 Does the host source dataset or the attribution method influence attribution rates? 17
3.4 Does attribution in Grampian mirror that in the rest of Scotland? 19
3.5 The sources of human campylobacteriosis in Grampian 20
3.6 Conclusions 23
4. Supplementary Data 24
5. References 27

Figures

Figure 1. Annual incidence of campylobacteriosis in UK..2
Figure 2. Incidence per 100,000 population of reports of *Campylobacter* infection 2010 (2009). 2
Figure 3. Incidence of campylobacteriosis in Scotland and Grampian from 1990 to 2010. 3
Figure 4. Regression of campylobacteriosis incidence in Grampian vs Scotland from 1990 to 2010. 4
Figure 5. Age structured incidence of campylobacteriosis in Grampian and Scotland for 2005- 06 and 2010- 11...4
Figure 6. Rarefaction (saturation) analysis. 13
Figure 7. Abundance of MLST Sequence Types of clinical isolates from Grampian. 14
Figure 8. Source attribution of 2010 -11 Grampian clinical isolates using (a) the 2010 -11 host datasets or (b) the combined 2005-06 + 2010 -11 host datasets. 18
Figure 9. Patient age vs attributed host source of Scottish (2005 -06) and Grampian (2005 -07) clinical isolates. 19
Figure 10. Attributed host sources of clinical isolates from Grampian in 2005-06 and from 2010 -11 partitioned by patient age. 20
Figure 11. Attribution to five potential host reservoirs of clinical *Campylobacter* cases in Grampian per month by (a) STRUCTURE with alleles Model, (b) Asymmetric Island Model. 21
Figure 12. Attribution to five potential host reservoirs of clinical *Campylobacter* cases in Grampian per month by (a) STRUCTURE with alleles Model, (b) Asymmetric Island Model. 22
Supplementary Figure 1. Patient age vs attributed host source of (a) Scottish clinical isolates (2005 -06) or (b) Grampian clinical isolates (2005 -07). 24
Supplementary Figure 2. Molecular attribution of clinical *Campylobacter* by STRUCTURE with alleles. 25
Supplementary Figure 3. Molecular attribution of clinical *Campylobacter* by Asymmetric Island. 26

Tables

Table 1. Number of specimens collected, number of presumptive *Campylobacter* spp. isolated, number of MLST-confirmed *Campylobacter* spp, number of MLST 7 locus isolates. ...6
Table 2. Isolate datasets. 7
Table 3. Molecular attribution methods used. 10
Table 4. *Campylobacter* prevalence in 2005 -06 and 2010 -11 in cattle, sheep and retail chicken.. 11
Table 5. Comparing the relative abundance of the most prevalent clinical MLST types over time. 15
Table 6. Genetic distances (Nei) between isolates in the two study periods for all sources. ... 16

1. Introduction

1.1 Background

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 (16), and has since become recognised as the commonest known cause of bacterial infectious intestinal disease (IID) worldwide. According to WHO estimates, *Campylobacter*-related illness affects around 1% of populations in developed countries every year.

Campylobacter infection causes almost half of all IID cases in the UK, with *Campylobacter jejuni* causing around 90% of cases and the closely-related *Campylobacter coli* causing almost all the rest. In 2010 a total of 6597 isolates of *Campylobacter* were reported in Scotland, which was an increase of 182 (2.8%) compared to the 6415 reports in 2009 though considerably less than in 2009 when reports had increased by 1537 (31.5%). The 6597 reported in 2010 is higher than the previous peak in the incidence in Scotland in 2000 when there had been 6482 reports (Figure 1). Because there is substantial under-reporting, the actual number of cases is likely to be closer to 500,000 (18). Further, about 10% of reported cases are hospitalised. This rise is all the more disappointing because rates of infection with *Campylobacter* spp had been falling between 2000 and 2005 (6). In Scotland the overall rate of *Campylobacter* infection in 2010 was 127.0 per 100,000 compared to 123.4 per 100,000 in 2009. Among the mainland NHS boards the lowest rate of 83.1 per 100,000 was in Fife, which is historically low. The highest rate of 162.9 per 100,000 was observed in Tayside, which also had the highest rate the previous year (Figure 2).

Most cases of *Campylobacter* are apparently sporadic with few identified outbreaks. There was one outbreak of *Campylobacter* in 2010 reported to ObSurv (the surveillance system for all general outbreaks of IID in Scotland), where 18 persons were reported to be ill and three of whom were laboratory confirmed with *Campylobacter*. In each of the previous two years there had also been one general outbreak of *Campylobacter*.

High rates of *Campylobacter* incidence translate into substantial annual economic costs, estimated at £503M in the UK (all likely cases) (7), EUR9M in the Netherlands (reported cases in 1999) (20), and $4.3bn in the USA (all likely cases) (1). *Campylobacter* infection can also lead to serious longer-term illness. Approximately one case for every 1000 reported cases leads to Guillain-Barré syndrome: a serious condition of reversible or permanent loss of limb motor function that is the commonest cause of acute flaccid paralysis. *Campylobacter* infection is also associated with the non-paralytic version of GBS, Miller-Fisher syndrome, and with reactive arthritis.

The main source of human campylobacteriosis in Scotland (3,15) and elsewhere in the developed world is retail chicken with a significant proportion of the remainder attributable to ruminants{EFSA, 2010 3327 /id;Strachan, 2010 3453 /id;Little, 2010 3507 /id;NELSON, 2010 3568 /id;Tustin, 2011 3589 /id}. Both the UK and Scottish governments have a responsibility to promote health and minimise logistic burden on the health care sector, and therefore want the incidence of human *Campylobacter* infection substantially reduced. Human *Campylobacter* infection is viewed as having a significant food-borne component, and therefore food safety regulation bodies and organisations in the food production sector are best-placed to identify and implement effective interventions. The 'Joint Working Group on *Campylobacter*' was established in August 2009 as a joint industry and government group (www.food.gov.uk/safereating/microbiology/campylobacterevidenceprogramme/wgcampy). It aims to identify interventions that would reduce *Campylobacter* in chicken. The membership includes the British Poultry Council (BPC), the National Farmers' Union (NFU) the British Retail Consortium (BRC), the FSA and Defra. They are identifying and

putting in place interventions that will reduce *Campylobacter* through a Joint Action Plan. The key activities of the action plan relate to on-farm, transport, processing, retail, consumer and catering sector trials and interventions, as well as surveillance and monitoring. The present study will address aspects of this last heading of surveillance and monitoring by seeking to clarify the sources of human campylobacteriosis in Scotland in 2010-11 which in 2005-07 (3) were determined to be principally retail chicken with a significant proportion of the remainder attributable to ruminants. It will establish base line data of campylobacteriosis and the molecular attribution of source of these clinical isolates which can be used to monitor the success of the other elements of the Joint Action Plan.

Figure 1. Annual incidence of campylobacteriosis in UK.

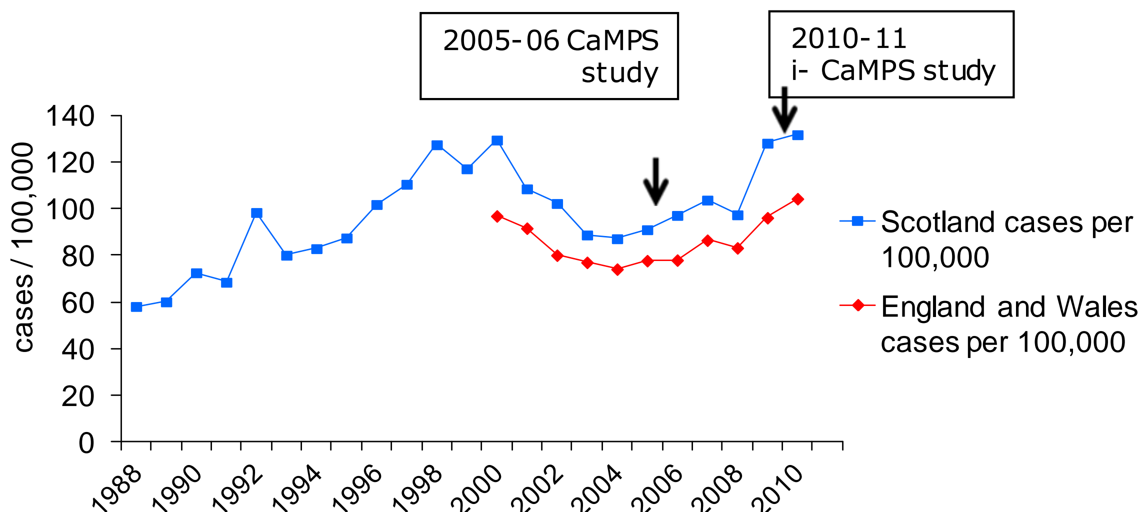
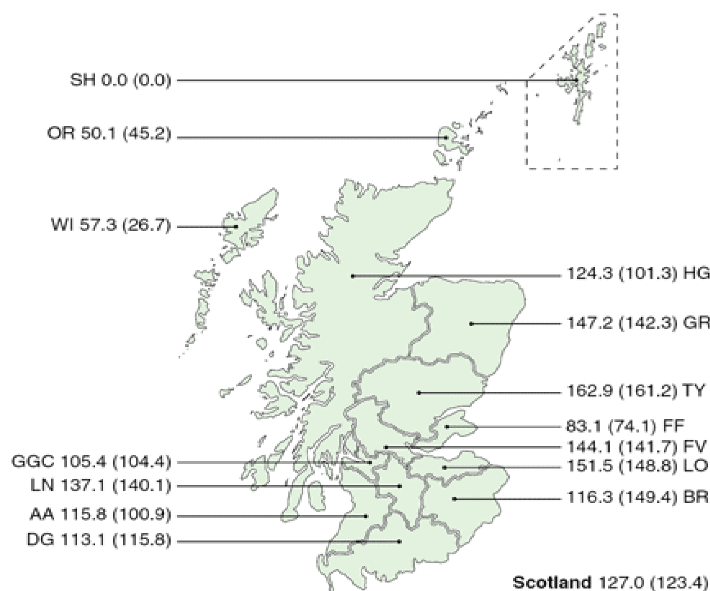


Figure 2. Incidence per 100,000 population of reports of *Campylobacter* infection 2010 (2009).



Data from Health Protection Scotland

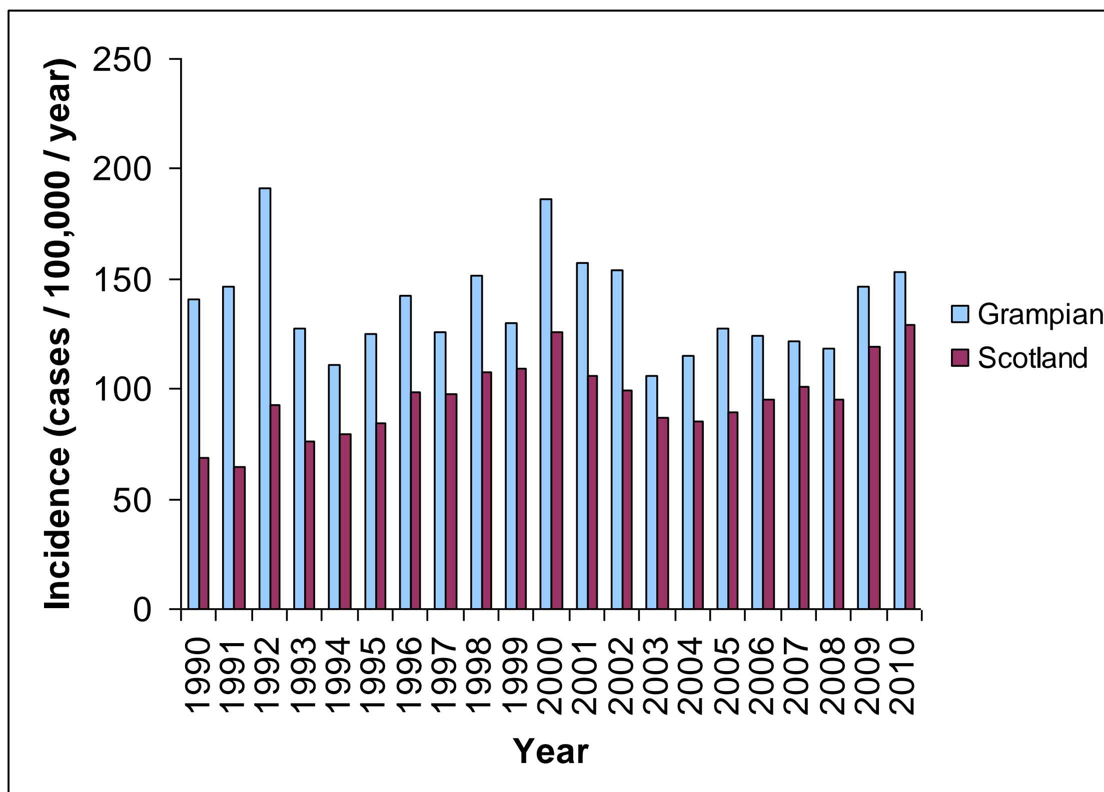
1.2 The molecular epidemiology of campylobacteriosis in Scotland can be modeled using Grampian data

The trends in incidence of Scottish campylobacteriosis have been broadly mirrored by those in Grampian (Figure 3). The incidence in Grampian is greater than in Scotland (P-value <0.0001, bootstrap pairwise comparison), however there is a decreasing trend in this difference (p=0.00048) (Figure 4). Looking more specifically at the two periods for which *Campylobacter* typing data is available (Scotland 2005- 06; Grampian 2010- 11 (Figure 5), there was not a significant change in incidence between Grampian and Scotland for either time period. As noted elsewhere (5), for older age groups the incidence of campylobacteriosis is increasing.

The FSA funded CaMPS study (3) of 2005/06 identified that the 12 mainland Health Boards all had similar proportions of the 25 most common STs. When all pairwise comparisons of ST composition among Health Boards were tested, only one pair (Dumfries & Galloway versus Forth Valley) yielded a significant difference (ARLEQUIN, exact test of population differentiation, details as above, P=0.0097). The differentiation was quantified as the index FST, which was found to be very low (ARLEQUIN, FST=0.003) and to be due to minor cumulative differences involving several rare STs.

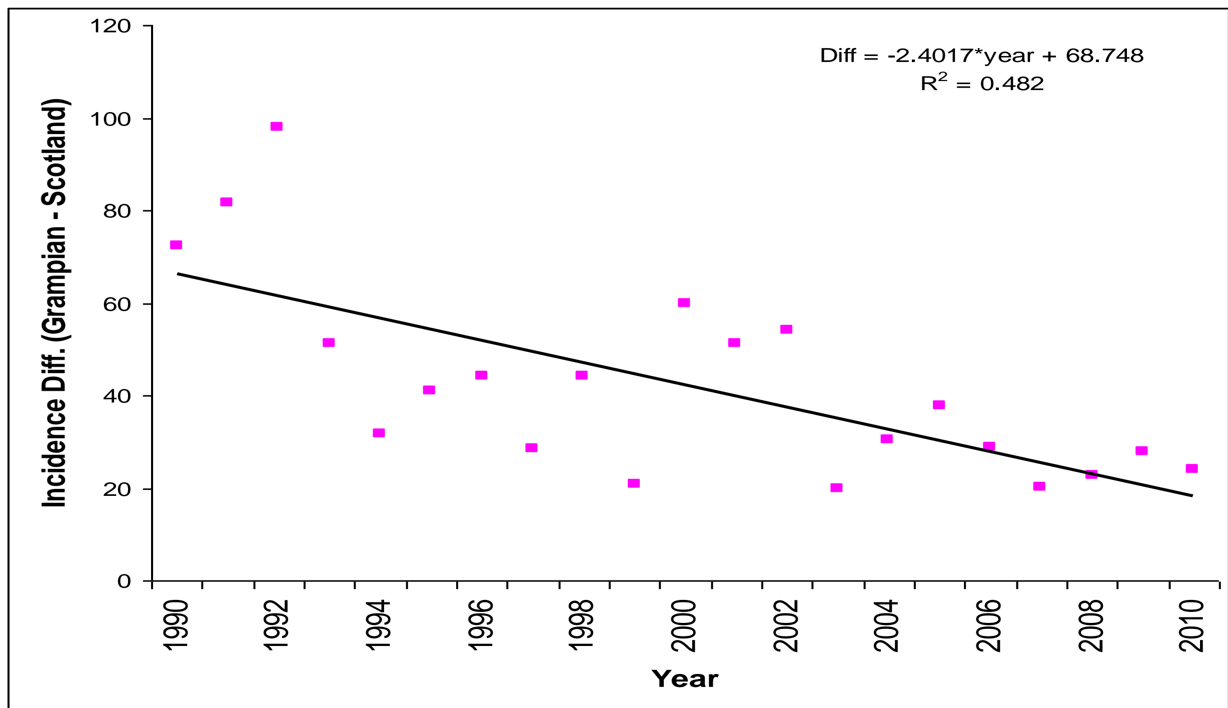
From the perspective of this FSAS Research Requirement, Scotland and Grampian have different, but quantifiable differences in the incidence of campylobacteriosis. However, the molecular attribution of sources of Grampian clinical isolates was virtually identical to that of Scotland as a whole. Hence, source attributions based on a region of Scotland can be scaled up to Scotland as a whole.

Figure 3. Incidence of campylobacteriosis in Scotland and Grampian from 1990 to 2010.



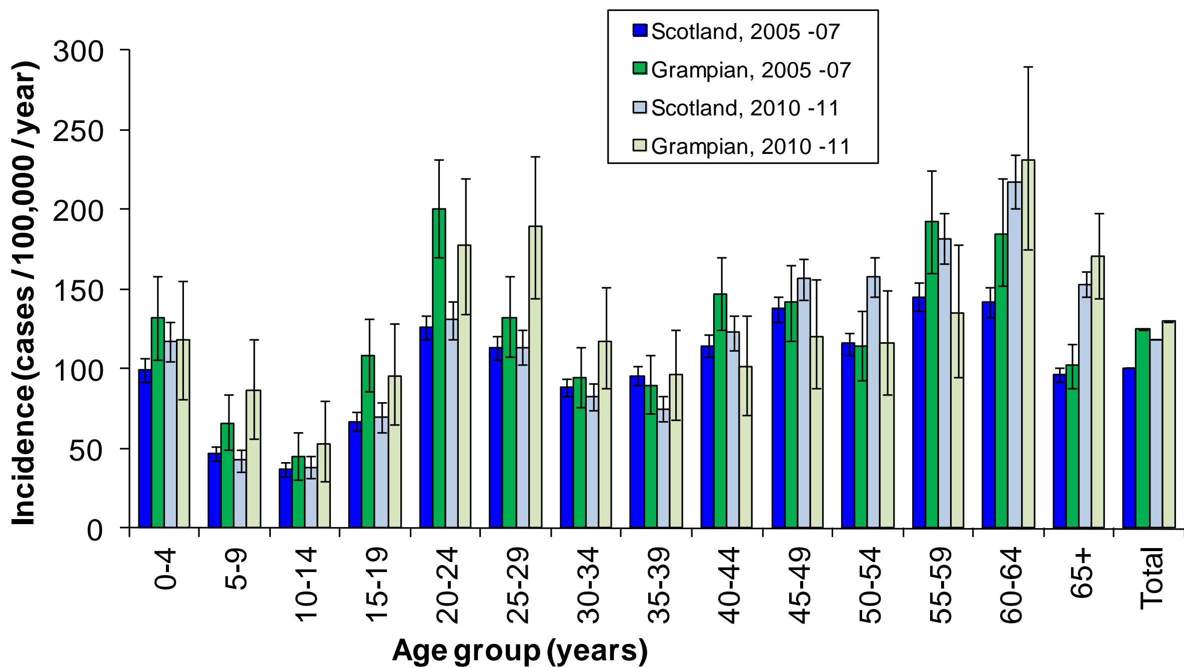
Data from Health Protection Scotland.

Figure 4. Regression of campylobacteriosis incidence in Grampian vs Scotland from 1990 to 2010.



Data from Health Protection Scotland

Figure 5. Age structured incidence of campylobacteriosis in Grampian and Scotland for 2005- 06 and 2010- 11.



Data from Health Protection Scotland. Error bars are 95% CI.

1.3 Aims

This Research Requirement seeks to estimate the proportion of clinical *Campylobacter* isolates that are attributable to retail chicken sources and to compare this with the previous CAMPS study. This attribution is dependent on appropriate source isolates typed by MLST. The CaMPS study (3,15) identified that those from chicken, cattle and sheep were of greatest relevance. This Research Requirement will establish baseline data against which the success of future interventions, over a number of years, at many points along the 'farm to fork' pathway to chicken consumption will be measured. It is therefore important that this baseline dataset includes contemporaneous chicken, cattle and sheep isolates. These will all be sourced predominantly from Grampian which we have shown previously to be typical for cattle and sheep strains (13) when compared with other Scottish regions, whilst retail chicken is both sourced and distributed all around the UK (our survey of abattoir locations displayed on retail chicken products in Grampian shows these to be sourced from across the UK).

There is evidence to suggest that *Campylobacter* strains (by ST and by allelic type) recovered from these host species are quite stable over large geographical scales (14) and timeframes (3). Hence, it is plausible that attribution of clinical strains to source can utilise pre-existing MLST data from these hosts. This will also be investigated in this research requirement.

2. Materials and Methods

2.1 Isolates Collections

All available clinical isolates in Grampian for the 12 month period 1 April 2010 –31 March 2011 (n=783) were collected (Table 1). Contemporaneous *Campylobacter* isolates from the principal source hosts were also collected: retail chicken, cattle, sheep (Table 1).

Table 1. Number of specimens collected, number of presumptive *Campylobacter* spp. isolated, number of MLST-confirmed *Campylobacter* spp, number of MLST 7 locus isolates.

	Specimens collected	*Campylobacter* positive specimens[a]	Campylobacter prevalence[a]	Presumptive-*Campylobacter* spp. isolated	MLST-confirmed *Campylobacter* spp[h]	Isolates with 7 locus MLST
Human	798	-	-	783[f]	775[g]	697
Chicken	238	215	90.3%	228[d]	211[e]	181
Cattle	142	47	33.1%	96	78[b]	77
Sheep	167	88	52.7%	103	101[c]	100

a: Confirmed *Campylobacter* spp. by MLST
b: Five isolates were presumptive *Arcobacter*, two isolates were presumptive *C. fetus*, the remainder did not provide sequence from all or most loci.
c: Two isolates were presumptive *C. fetus*
d: Includes 4 isolates from Glasgow and 9 taken as 2nd isolate from muti-portion packs.
e: Two isolates were presumptive *Arcobacter*, one isolates was presumptive *C. fetus*, the remainder did not provide sequence from all or most loci.
f: 15 isolates did not grow or were observably not *Campylobacter* spp.
g: Two isolates were presumptive *C. lari*, the remainder did not provide sequence from all or most loci.
h: A very few specimens yielded two isolates.

2.2 MLST of isolates

Multi-locus Sequence Typing was carried out on all isolates by the method detailed in the 2005-06 CaMPS report (3). These are summarised in Table 1. Not all presumptive isolates were confirmed to be *Campylobacter jejuni/ coli* by MLST (Table A) and this was most probably due to the difficulty of achieving this by visual inspection of colonies and latex sero-agglutination testing. Some *Campylobacter jejuni/ coli* confirmed isolates could not be successfully sequence typed at all loci (Table 1), even when alternative oligonucleotide PCR or sequencing primers were used.

2.3 Host reservoir isolate datasets

The poultry, cattle and sheep data were compared with that obtained in CaMPS 2005/6 (3) using Nei's genetic distance (6) and rarefaction to establish whether the species data can be combined from the two different years. This will also provide evidence of the stability or otherwise of sources over time.

To maximise the use of availabile source datasets, typed clinical and reservoir isolates from the 2005-06 Scottish study and clinical isolates from the overlapping 27 month period July 2005 -Oct 2007 (n=1452) from Grampian were used in the molecular attribution analyses (Table 2).

Table 2. Isolate datasets.

Host dataset 1. 2005 -06 Scottish-wide hosts

Host		*C.jejuni*	*C.coli*	TOTAL
Cattle	2005 -06	336	25	361
Sheep	2005 -06	91	56	147
Chicken	2005 -06	255	47	302
Wild Birds	2005 -06	176	12	188
Pigs	2005 -06	7	33	40

Host dataset 2. 2010 -11 Grampian-wide hosts

Host		Total
Cattle	2010 -11	77
Sheep	2010 -11	100
Chicken	2010 -11	181
Wild Birds	2005 -06	as above
Pigs	2005 -06	as above

Host dataset 1&2. Combined 2005 -07 Scottish-wide plus 2010 -11 Grampian-wide hosts

Host		Total
Cattle	2010 -11 & 2005 -06	438
Sheep	2010 -11 & 2005 -06	247
Chicken	2010 -11 & 2005 -06	483
Wild Birds	2005 -06	188
Pigs	2005 -06	40

Clinical isolate datasets.

Period	Region	Total
2005 -06	Scotland	5674
2005 -07	Grampian	1452
2010 -11	Grampian	697

2.4 Molecular attribution methods

Attribution by microbial sub-typing is a relatively new area of research. The term "source attribution" has been defined (11) 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 (2). Three of these methods will be used in the current study (Table 3) and are detailed below.

The Dutch Model (4) 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 as in 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.

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.

STRUCTURE (12) 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 programme 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.

The Asymmetric Island (AI) Model (21) incorporates a Bayesian approach and uses the allelic profile of the sequence subtypes to reconstruct the genealogical history of the isolates. The host populations are considered to exist on separate "islands" (e.g. the sheep island). Mutations and recombination occur on each island. Migrations from between each reservoir (island) and into the human population are used to estimate the degree of attribution to each source. This model has been applied to MLST data from England (21), Scotland (15) and New Zealand where 56%, 78% and 75% of human cases were attributed to poultry respectively.

The Asymmetric Island model incorporates recombination and mutation, uses MLST data at the allele level and achieves relatively high values for self attribution. However, the model appears to be complicated and the current explanations of its operation difficult to comprehend. The Asymmetric Island model assigns each human case to the potential source populations on the basis of DNA sequence similarity. By comparing human isolates to a panel of reference sequences of known source (e.g. cattle, sheep, chickens, pigs, wild birds and the environment), each human case can be assigned a probability of originating in each source population. The source attribution probabilities are calculated using a statistical model of the way the DNA sequences evolve in the populations of bacteria. In the statistical model, there are parameters representing the processes of mutation, DNA exchange between bacteria (recombination or horizontal gene transfer) and zoonotic transmission between populations. These processes lead to differences in gene frequencies between the source populations, facilitating source attribution. The model can be trained, by estimating the parameters exclusively from the sequences of known source, before using it to calculate source attribution probabilities for human sequences.

Self attribution is a key performance measure for these models. This is the average percentage accuracy that any given isolate from a reservoir can be correctly attributed back to its own reservoir. This can be performed in a number of ways. However, one simple approach is to use a jacknife method to predict the source of an isolate that was unknown to the model and known to the user. This is then repeated for all the source isolates a number of times (e.g. up to 10,000) so that an average, and confidence intervals, can be calculated. Self attribution ranges are reported as between 62-97% for between 5-7 hosts for the Asymmetric Island model (15,21) and 38-70% for STRUCTURE (15). Note that by chance you would expect a correct self attribution of 20% and 14% for 5 and 7 sources respectively. The poorest self attribution in these methods is environment, which is likely to contain isolates from a number of hosts. These data demonstrate that there are differences in the frequencies of MLST types between hosts and that this information can be used for source attribution.

Table 3. Molecular attribution methods used.

Method	Genetic unit of assessment
Dutch proportional	ST allele
STRUCTURE	ST allele
Assymetric Island	allele

3. Results and Discussion

3.1 Has the prevalence of *Campylobacter* in food and animal reservoirs changed over time?

The prevalence of *Campylobacter* in the different reservoirs sampled and in retail chicken was compared between the 2005 -06 study and 2010 -11 study (Table 4). There was a significant increase in prevalence (OR >1 and P<0.05) for all reservoirs and retail chicken. This increase could provide an explanation for the increase in clinical cases over the last few years.

Table 4. *Campylobacter* prevalence in 2005 -06 and 2010 -11 in cattle, sheep and retail chicken.

Reservoir	2005 -06 +ve/total (%)	2010 -11 +ve/total (%)	OR[b]	P-value[c]
Cattle	104/474 (21.9)	47/142 (33.1)	1.76	0.005
Sheep	97/292 (33.2)	88/167 (52.7)	2.24	<0.001
Chicken	142/222 (64.0)	215/238 (90.3)	5.27	<0.001

[a]Taken from Table 1
[b]Odds ratio (if >1.0 indicates an increase in prevalence with time)
[c]Calculated by Fisher's exact test

3.2 Do strain types change over time?

The extent to which the isolates from sources represented the maximum hypothetical diversity was characterised using rarefaction. Rarefaction is a data re-sampling technique that indicates whether diversity has reached a plateau or is still rising at the total sample size, i.e., at the end of collection. A rarefaction curve that has reached a plateau indicates that all diversity (i.e. all MLST genotypes) has been sampled whereas an increasing slope indicates that some diversity remains unsampled (i.e. there are likely to be MLST types in the reservoir that have not yet been sampled). This method assumes that the dataset represents a random sample taken from a closed system characterised by a constant, stable spectrum of types. As in the 2005 -06 study (3), the rarefaction curves for all clinical, environmental and food sources were still rising, even at the maximum sample sizes (Figure 6). This is because the system being studied is open to immigration (e.g. for human clinical strains there will be immigration by foreign travel) and also that the sampling size is not sufficiently large to be comprehensive. Both the clinical and chicken strains have similar levels of diversity. However, the cattle and sheep strains exhibited less diversity (for both the 2005 -06 and 2010 -11 studies) than those from retail chicken and human clinical strains.

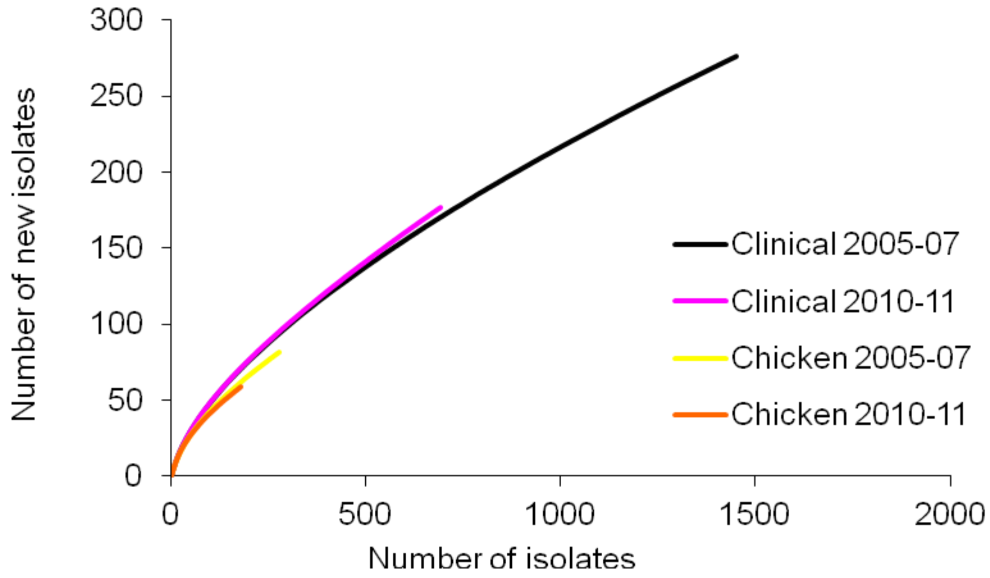
It was apparent visually that the proportions of clinical strains were changing rapidly with time (Figure 7). Table 5 presents how the relative abundance of the main clinical strains changed over time. Comparing the twelve month period up to July 2006 with the following fifteen month period showed that quarter of the 16 most abundant strains had changed significantly in abundance. By the time of the 2010 -11 study half of the major strains had significant changes in relative abundance: the most abundant strain (ST257) decreased in relative abundance by half, the third most abundant strain (ST45) by quarter. A dramatic exemplar of changing abundance is ST5136 (clonal complex CC464) which was undetected during the 2005 -07 study and became the sixth most abundant strain in the 2010 -11 study. A single ST5136 isolate was first recorded, in the same year, 2010. in the pubMLST database as a UK clinical isolate from human stool collected in Oxford. During the 2010 -11 study there were 26 clinical (4%) and 14 chicken (8%)

isolates reported. The chicken isolates originated from a single Company but from three different plants throughout the UK. It is worth noting that in New Zealand there are strains which are associated with particular poultry companies (9) and this appears to be the case in the UK with ST5136. All ST5136 clinical isolates were isolated from September 2010 up to the end of the study period in 2011.

These descriptive changes in the *Campylobacter* population which were apparent at a strain level were also examined by calculating, Nei's, genetic distance between isolates from each source from the two study periods (Table 6). Nei's genetic distance is a measure of the overlap in the genetic content of populations and this was measured at both strain level (a single measure of similarity using ST number) and at allele level (similarity measured across the seven MLST loci). Again significant differences were observed both between hosts and between the two study periods.

The most parsimonious explanation of this strain diversity is that *Campylobacter* present in clinical, environmental and food sources in Scotland represents an extremely large pool of strains that is continually being augmented: internally by mutation and recombination and externally by strain input from human travel and migrating wildlife. Since the Nei genetic distance findings imply that the host datasets are only somewhat genetically similar, then combining datasets for a particular host from the two periods may be problematical, however small sample size (cf rarefaction) will have contributed to this. Accordingly, attribution analyses have used a combination of 2005 - 06 and 2010 - 11 host datasets as indicated in Table 2 and in the analyses below.

Figure 6. Rarefaction (saturation) analysis.



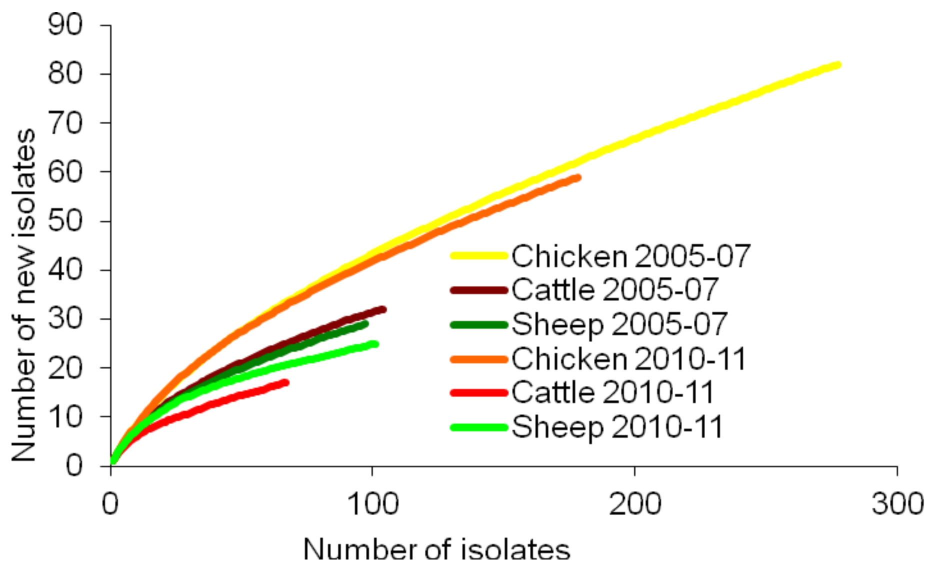


Figure 7. Abundance of MLST Sequence Types of clinical isolates from Grampian.

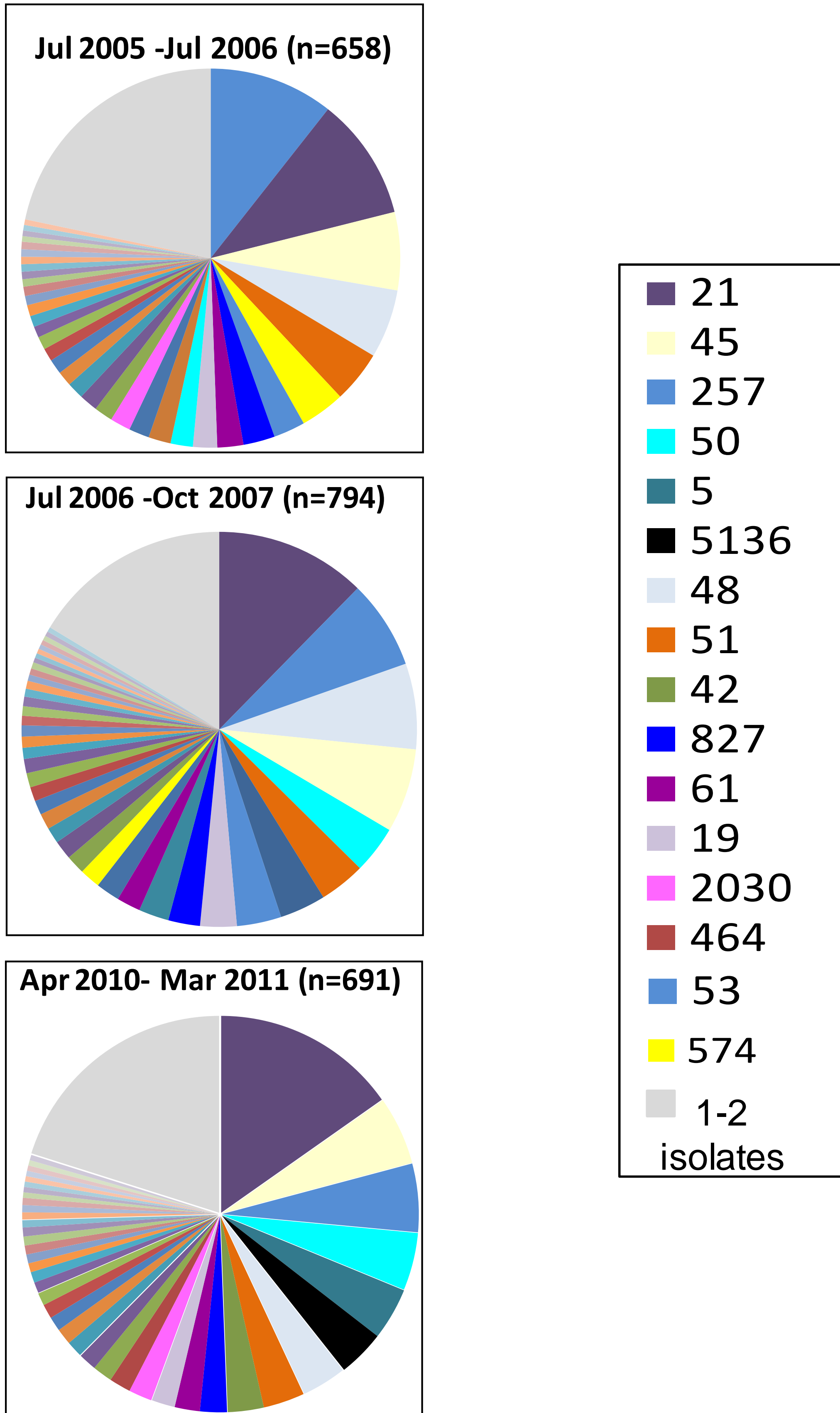


Table 5. Comparing the relative abundance of the most prevalent clinical MLST types over time.

Comparison	ST	OR[a]	P-value[b]
Jul 2005-Jul 2006 with Jul 2006-Oct 2007	21	1.30	0.070
	45	1.12	0.340
	257	0.71	0.046
	50	2.24	0.012
	5	0.88	0.508
	5136	∞	1.000
	48	1.30	0.136
	51	0.91	0.478
	42	1.06	0.537
	827	1.04	0.527
	61	0.94	0.638
	19	1.57	0.129
	2030	0.72	0.302
	464	0.73	0.413
	53	1.46	0.140
	574	0.47	0.019
	1-2 isolates	0.76	0.017
Jul 2005-Jul 2006 with Apr 2010-Mar 2011	21	1.68	0.001
	45	0.23	<0.001
	257	0.55	0.003
	50	2.78	0.001
	5	4.32	<0.001
	5136	∞	<0.001
	48	0.67	0.082
	51	0.82	0.297
	42	2.05	0.045
	827	0.89	0.441
	61	1.01	0.565
	19	1.01	0.568
	2030	1.20	0.409
	464	2.04	0.113
	53	0.59	0.127
	574	0.45	0.021
	1-2 isolates	0.91	0.252

[a] Odds ratio (if >1.0 indicates an increase with time)
[b] Calculated by Fisher's exact test

Table 6. Genetic distances (Nei) between isolates in the two study periods for all sources.

Within hosts

ST level

Group	Genetic Distance (p-value)
Cattle 2005 -07 vs. 2010 -11	0.5655 (0.0001)
Sheep 2005 -07 vs. 2010 -11	0.4519 (0.0007)
Chicken 2005 -07 vs. 2010 -11	0.6221 (<0.0001)
Clinical 2005 -07 vs. 2010 -11	0.4096 (<0.0001)

Allele level

Group	Genetic Distance (p-value)
Cattle 2005 -07 vs. 2010 -11	0.3735 (0.0004)
Sheep 2005 -07 vs. 2010 -11	0.2378 (0.0247)
Chicken 2005 -07 vs. 2010 -11	0.3089 (<0.0001)
Clinical 2005 -07 vs. 2010 -11	0.1288 (0.0028)

Between hosts

ST level

Group	Genetic Distance (p-value)
Cattle 2010 -11 vs. Sheep 2010 -11	0.6875 (<0.0001)
Cattle 2010 -11 vs. Chicken 2010 -11	0.8605 (<0.0001)
Sheep 2010 -11 vs. Chicken 2010 -11	0.8232 (<0.0001)
Cattle 2010 -11 vs. Clinical 2010 -11	0.8428 (<0.0001)
Sheep 2010 -11 vs. Clinical 2010 -11	0.7651 (<0.0001)
Chicken 2010 -11 vs. Clinical 2010 -11	0.5355 (<0.0001)

Allele level

Group/Group	Genetic Distance (p-value)
Cattle 2010 -11 vs. Sheep 2010 -11	0.4983 (<0.0001)
Cattle 2010 -11 vs. Chicken 2010 -11	0.5939 (<0.0001)
Sheep 2010 -11 vs. Chicken 2010 -11	0.5955 (<0.0001)
Cattle 2010 -11 vs. Clinical 2010 -11	0.5383 (<0.0001)
Sheep 2010 -11 vs. Clinical 2010 -11	0.4713 (<0.0001)
Chicken 2010 -11 vs. Clinical 2010 -11	0.2752 (<0.0001)

Nei's genetic distance take a value of 0.0 where the genetic distance between the two populations is completely overlapping, and 1.0 when the two populations are completely genetically distinct.

3.3 Does the host source dataset or the attribution method influence attribution rates?

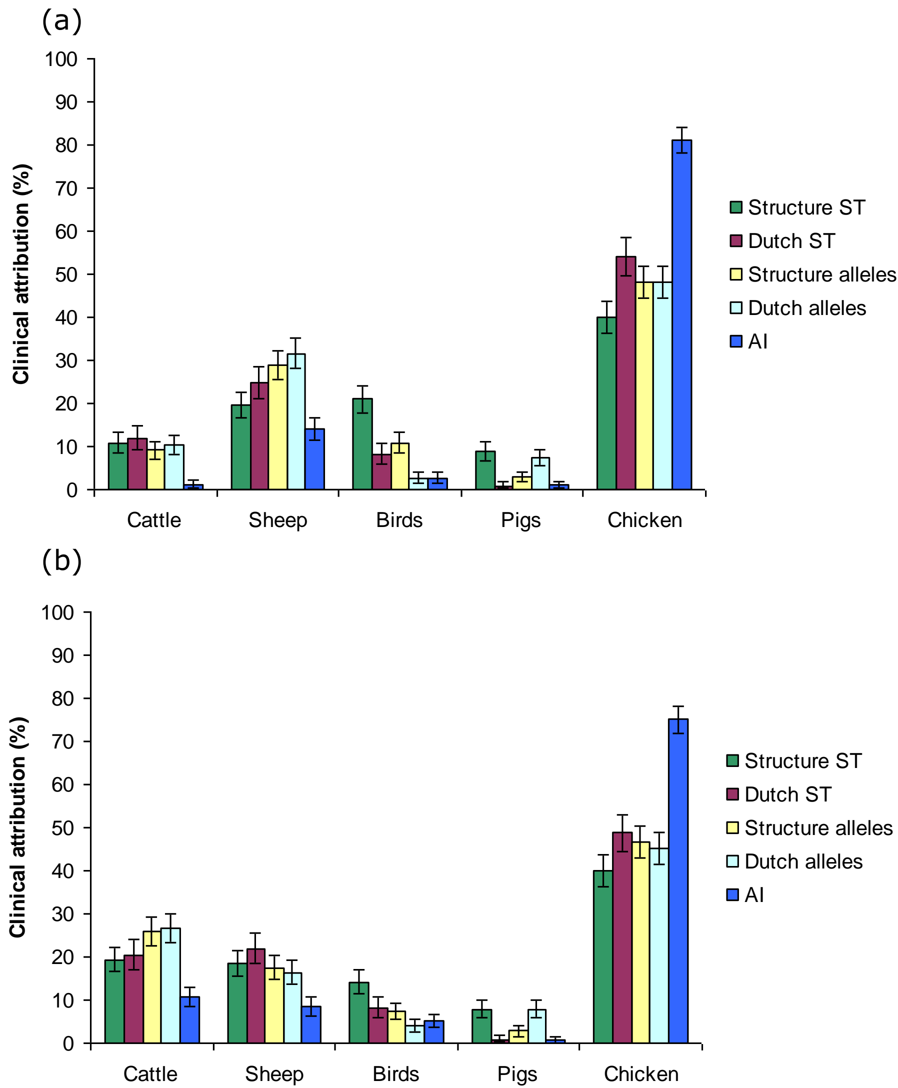
Section 3.2 noted that there were differences between the strain distributions in the host datasets from 2005 -07 and 2010 -11 (Host Dataset 1 and Dataset 2). The extent that these differences could influence attribution were for 2010 -11 Grampian clinical isolates tested by comparing the contemporaneous but small Host Dataset 2 with the larger Host Dataset 1&2. Figure 8 indicates that although there were differences these were not statistically significant, other than for the relative distribution between the two ruminant species.

The attribution methods were described in Section 2.4. The choice of attribution method was assessed by comparing five variations of the three different attribution methods (Table 3). Core to all three of the methods is the idea that the relative abundance of strains in each reservoir is used to assess the likelihood that a clinical isolate will have come from one or other reservoir on the basis of the relative abundance of that strain in each reservoir. Two of the five models (Table 3) used strain definitions based on a single number, the ST number, for this classification, the other three models used a strain definition based on seven numbers, the seven allele numbers. The former has the merit of simplicity but cannot attribute a reservoir to a clinical isolate where its ST number is not found in any of the host reservoirs, and will poorly attribute where there are few reservoir isolates with that ST. The latter three models are more complex, Asymmetric Island particularly so, however they have the advantage that the attribution of a clinical isolate is based on the smaller variety of alleles available and also by averaging the attribution scores from each of the seven loci; there is thus less chance of a rare clinical strain which has no common alleles in the different host reservoirs.

Comparison of the two graphs in Figure 8 suggests that there is less inter-model variation when larger reservoir datasets are used. Secondly, most of the models gave broadly similar attributions to the different reservoirs suggesting there was broad agreement between them due to an underlying common biology. The Asymmetric Island model consistently gave a higher attribution to retail chicken (75% to 81%) than the other models (40% to 54%), as has also been reported elsewhere (15), however the parameters behind this method have not been published and so the reason for the attribution differences are not clear.

In conclusion, the different models and the different host datasets gave broadly similar molecular attribution results. Hence, further work in this study will use the combined 2005 -07 and 2010 -11 host dataset (Host Dataset 1&2) and both STRUCTURE with alleles and Asymmetric Island models.

Figure 8. Source attribution of 2010 -11 Grampian clinical isolates using (a) the 2010 -11 host datasets or (b) the combined 2005-06 + 2010 -11 host datasets.



Source attribution of 2010 -11 Grampian clinical isolates using Host Dataset 1&2. 95% CI.

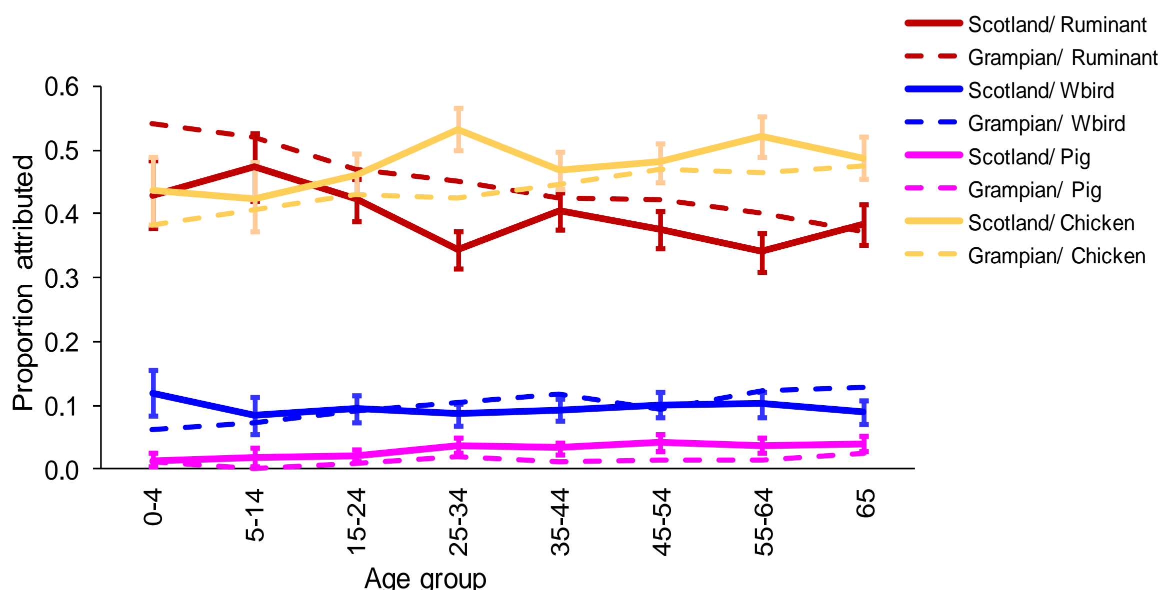
3.4 Does attribution in Grampian mirror that in the rest of Scotland?

Comparison of the molecular attribution of clinical isolates from Scotland vs Grampian was performed. This used clinical and host datasets from the period 2005 -07 and the STRUCTURE model utilising MLST data at the allele level. Whilst previous analyses were stratified into five hosts – chicken, cattle, sheep, pigs, wild birds; the similarity of cattle and sheep strains to each other and the similarity in human exposure to these sources suggested that these two hosts could be combined together as ruminants. Patient age is known to be an indicator of differing incidences of campylobacteriosis and so this was included here. Confidence intervals have been graphed at 90% to make comparison of findings more apparent; most obviously in the larger ranges seen with the smaller Grampian dataset. It should be noted that one inevitable consequence of increasing partitioning of the data into smaller groups (e.g. by age) will be larger error bars.

The molecular attribution of clinical isolates from Scotland vs Grampian using clinical and host datasets from the period 2005 -07 using STRUCTURE with alleles is presented in Figure 9 (full data presented in Supplementary Figure 1) . Overall, for each of the four hosts there is a very close agreement in the molecular attribution of clinical isolates from Scotland vs those from Grampian given the overlap in error bars. Indeed it is only in age group 25-34 where there is a significantly greater attribution for chicken in Scotland than Grampian; and conversely for ruminants in Grampian.

Overall these findings, together with the earlier points that the trends in changing incidence are similar and that the 2005 -06 study (3) demonstrated that there were similar proportions of the 25 most common ST's across most health boards suggests that Grampian is a suitable region from which to model Scottish-wide trends.

Figure 9. Patient age vs attributed host source of Scottish (2005 -06) and Grampian (2005 -07) clinical isolates.



Summary of Supplementary Figure 1. Attribution based on Host Dataset 1. Error bars are 90% CI.

3.5 The sources of human campylobacteriosis in Grampian

The above analyses have established:

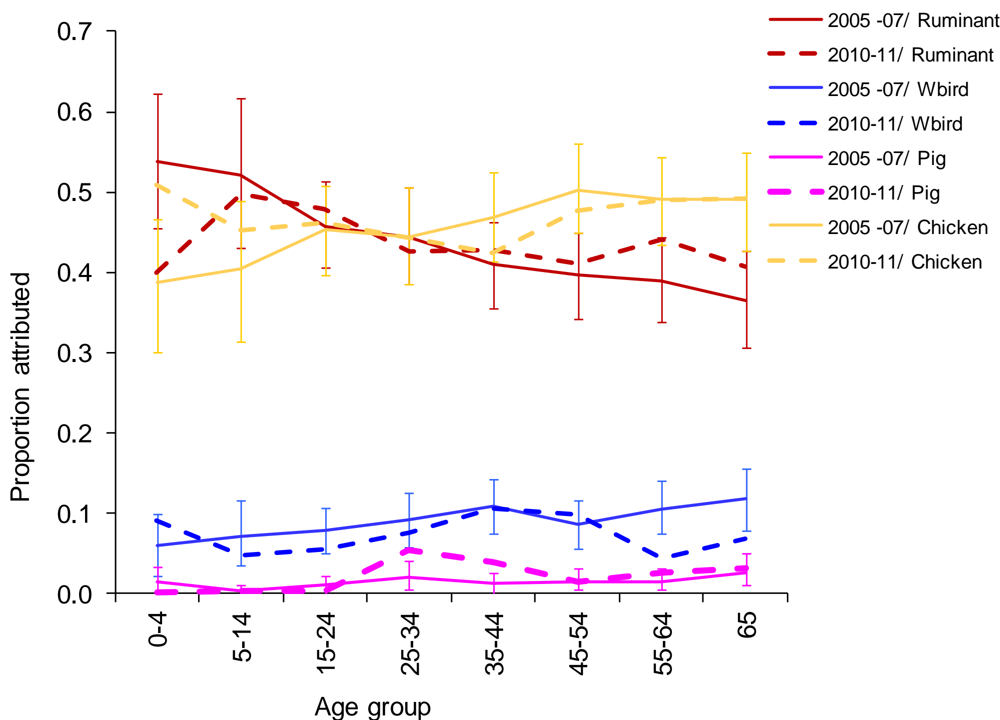
- Grampian is representative of Scotland for the understanding of the sources of campylobacteriosis
- Larger host datasets are more informative for attribution than smaller datasets
- STRUCTURE with alleles and the Asymmetric Island Models are appropriate for attribution analyses

In Grampian, neither pigs, nor wild birds contributed significantly to the burden of campylobacteriosis and this burden is constant with age (Figure 10). In contrast, ruminant and retail chicken sources were both significant contributors (Figure 10).

There is an age dependent increase in attribution to retail chicken sources at the expense of ruminant sources. Since there is a trend of increasing number of cases in the elderly population in recent times (5) this could be explained by poultry sources. From the single year of data collected in 2010 -11 it is difficult to confirm whether this is actually the case, however the continuance of this study may clarify this hypothesis. All of these age dependent trends were observed in both study periods.

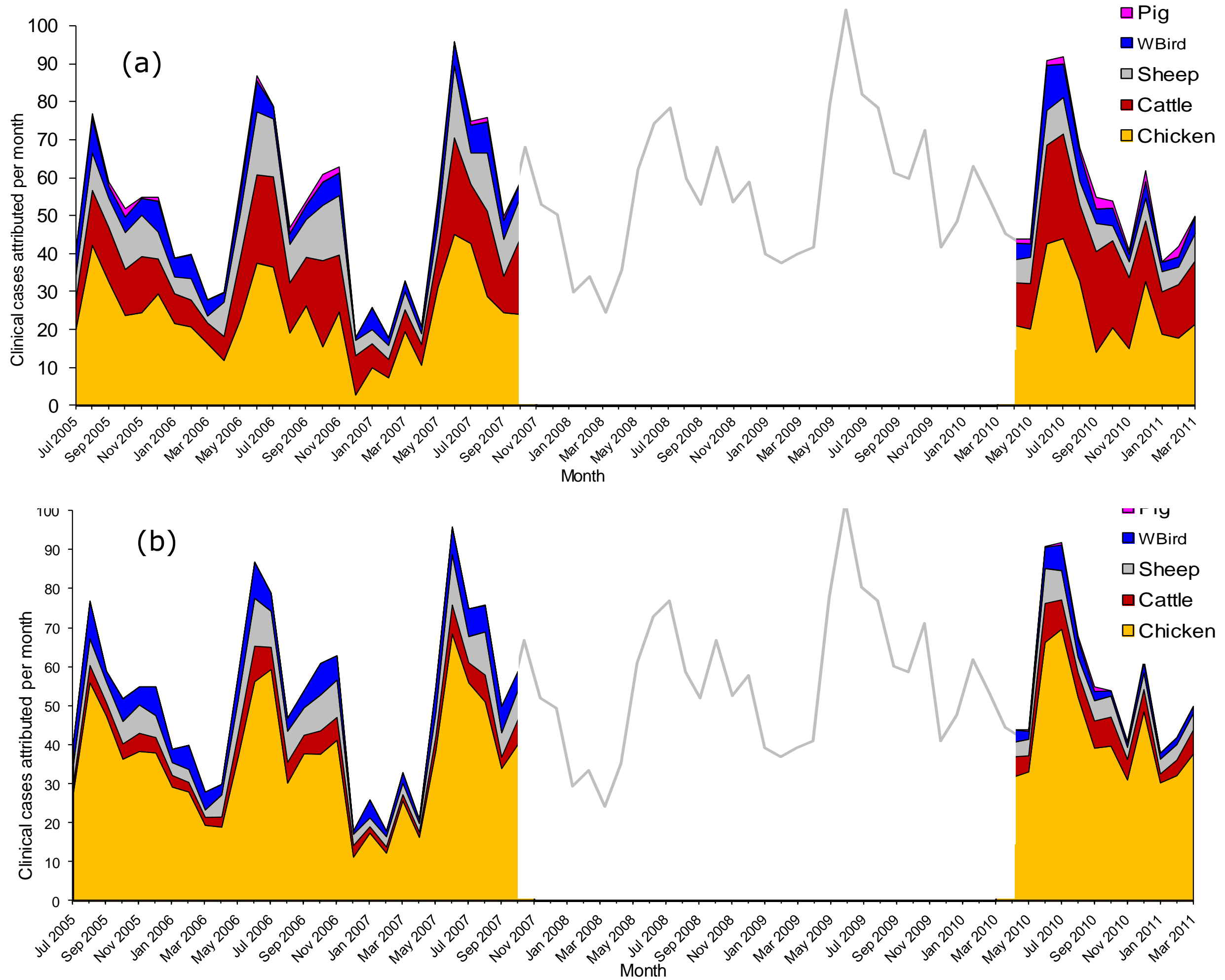
Little change in the relative importance of these sources was seen between the two study periods which were separated by 29 months (Figures 11, 12; full data in Supplementary Figures 2, 3). For example, STRUCTURE with alleles showed little overall difference in attribution between the two study periods (e.g. chicken 46% and 44%; cattle/sheep combined 42% and 43%). However, there was an increase in the proportion of cases that were associated with chicken (70% compared with 75%) using the Asymmetric Island model.

Figure 10. Attributed host sources of clinical isolates from Grampian in 2005-06 and from 2010 -11 partitioned by patient age.



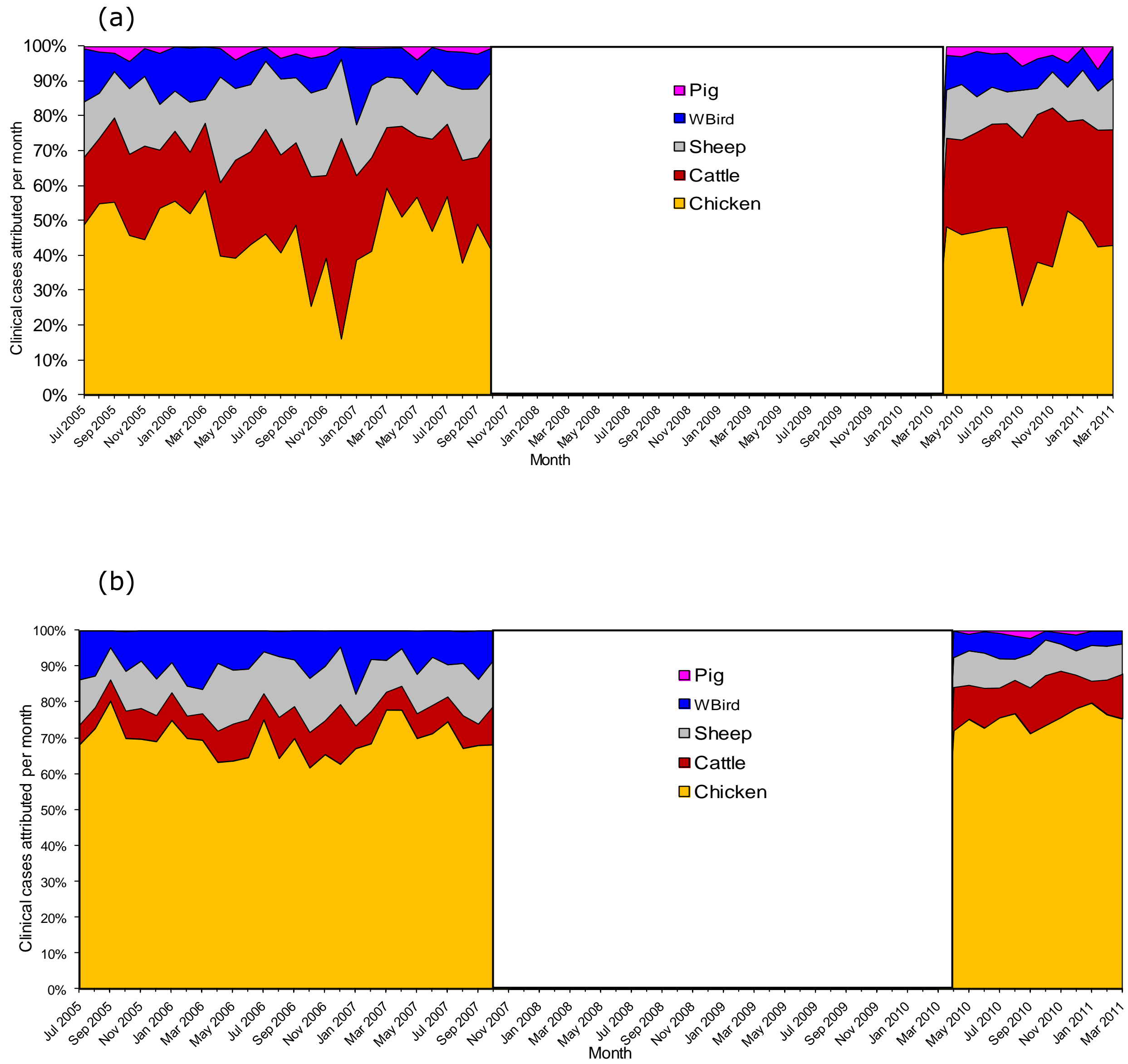
Attribution based on Host Dataset 3. 90% CI.

Figure 11. Attribution to five potential host reservoirs of clinical *Campylobacter* cases in Grampian per month by (a) STRUCTURE with alleles Model, (b) Asymmetric Island Model.



Graph stacked to total number of clinical cases. Attribution based on Host Dataset 3. 90% CI.

Figure 12. Attribution to five potential host reservoirs of clinical *Campylobacter* cases in Grampian per month by (a) STRUCTURE with alleles Model, (b) Asymmetric Island Model.



Graph stacked to 100% of clinical cases. Attribution based on Host Dataset 3. 90% CI.

3.6 Conclusions

There has been an increase in prevalence of *Campylobacter* in animal reservoirs (cattle, sheep and retail chicken) and this may explain the increase in human disease incidence.

There continues to be extensive population diversity of *Campylobacter* strains in farm animals, in retail chicken and in human isolates.

The relative abundance of the strain types found in these reservoirs changed over time (the relative abundance of half of the commoner strains changed between the first year of the 2005 -07 study and the 2010 -11 study), even over periods as short as one year. Indeed, in only the 28 months between the two study periods strain ST5136 rose in abundance from undetected in c.6000 Scottish clinical isolates to become the sixth most abundant (4%) clinical strain. Further, this strain was also only found in the latter period's retail chicken samples (8%). That this strain has expanded its population size rapidly and may also have evolved recently is suggested by all of these isolates having the same alleles at the hypervariable *flaA flaB porA* loci.

Host attribution modelling of putative sources of human infection suggests that there is broadly the same proportional attribution in Grampian compared to Scotland as a whole with retail chicken making the largest contribution.

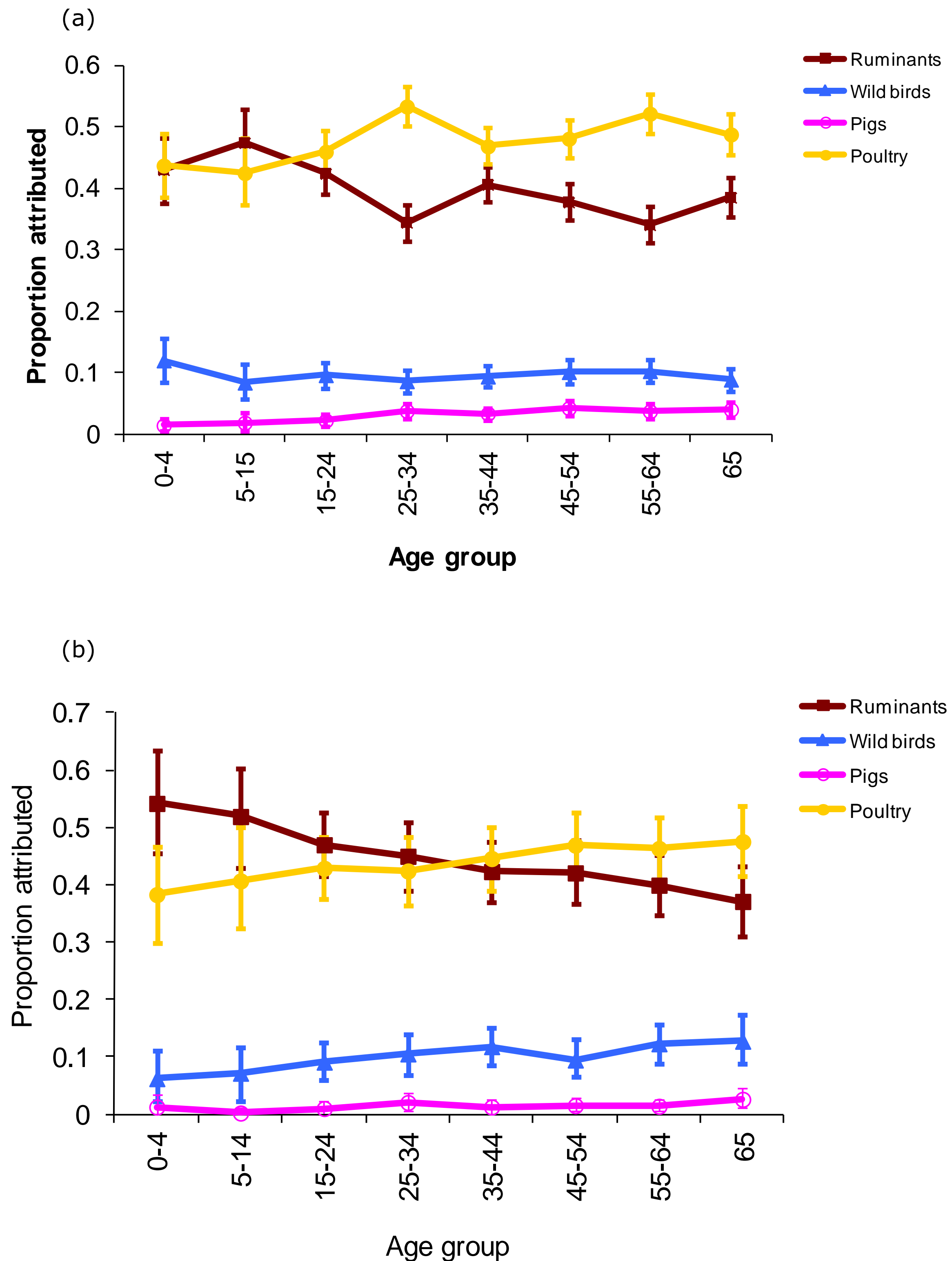
Host attribution modelling of putative sources of human infection suggests that there has been broadly the same proportional attribution over the two study periods with retail chicken making the largest contribution. The Asymmetric Island model, alone, suggested that retail chicken may now be making a greater contribution than in the past.

The impact, in Scotland and the UK, of forthcoming intervention strategies to reduce human campylobacteriosis originating from the poultry food chain should be observable by a decrease in human cases and confirmed by a subsequent decrease in the proportion of clinical isolates associated with chicken.

The current study has highlighted the dynamic nature of *Campylobacter* and the requirement to monitor prevalence, counts and strain types.

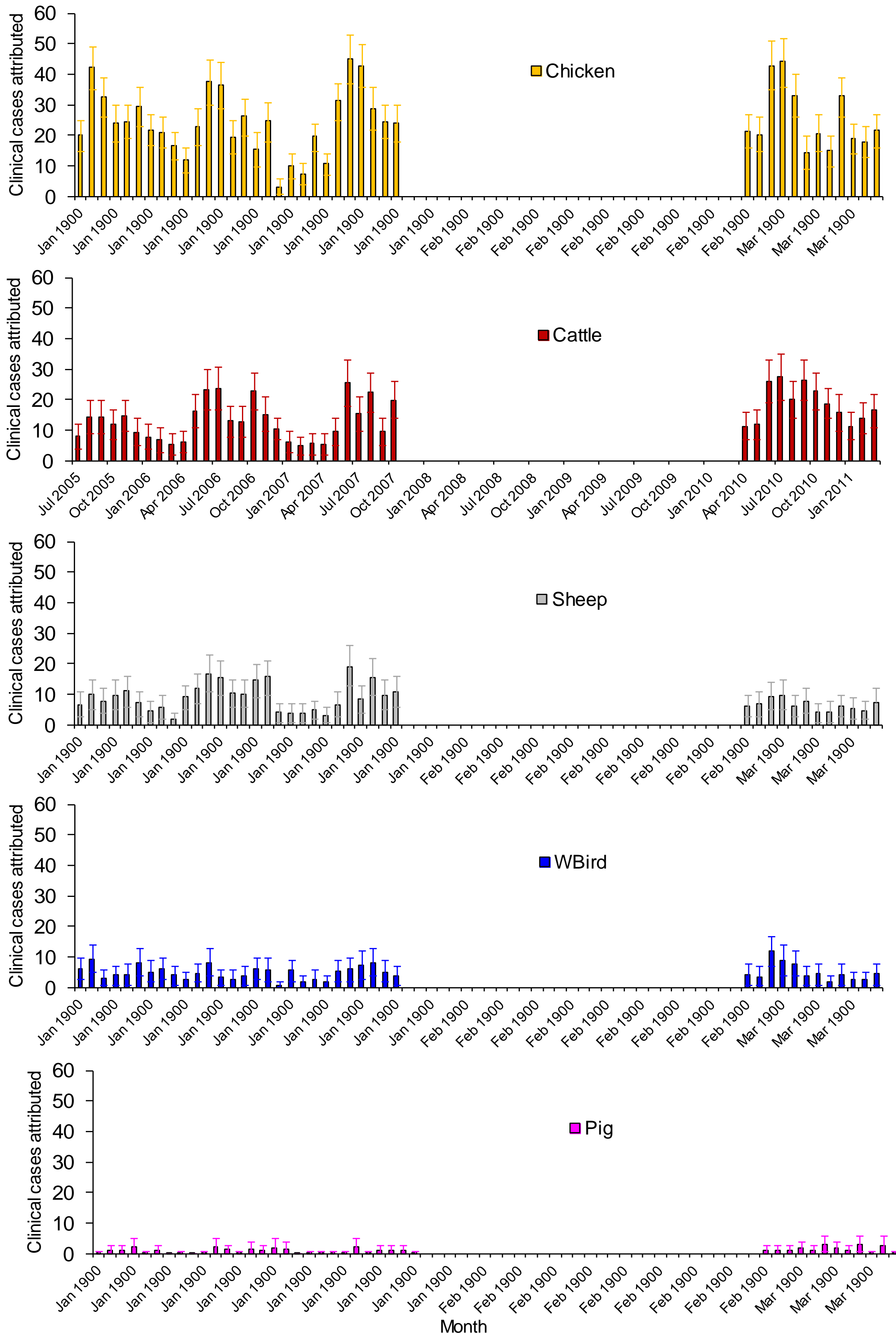
4. Supplementary Data

Supplementary Figure 1. Patient age vs attributed host source of (a) Scottish clinical isolates (2005 -06) or (b) Grampian clinical isolates (2005 -07).



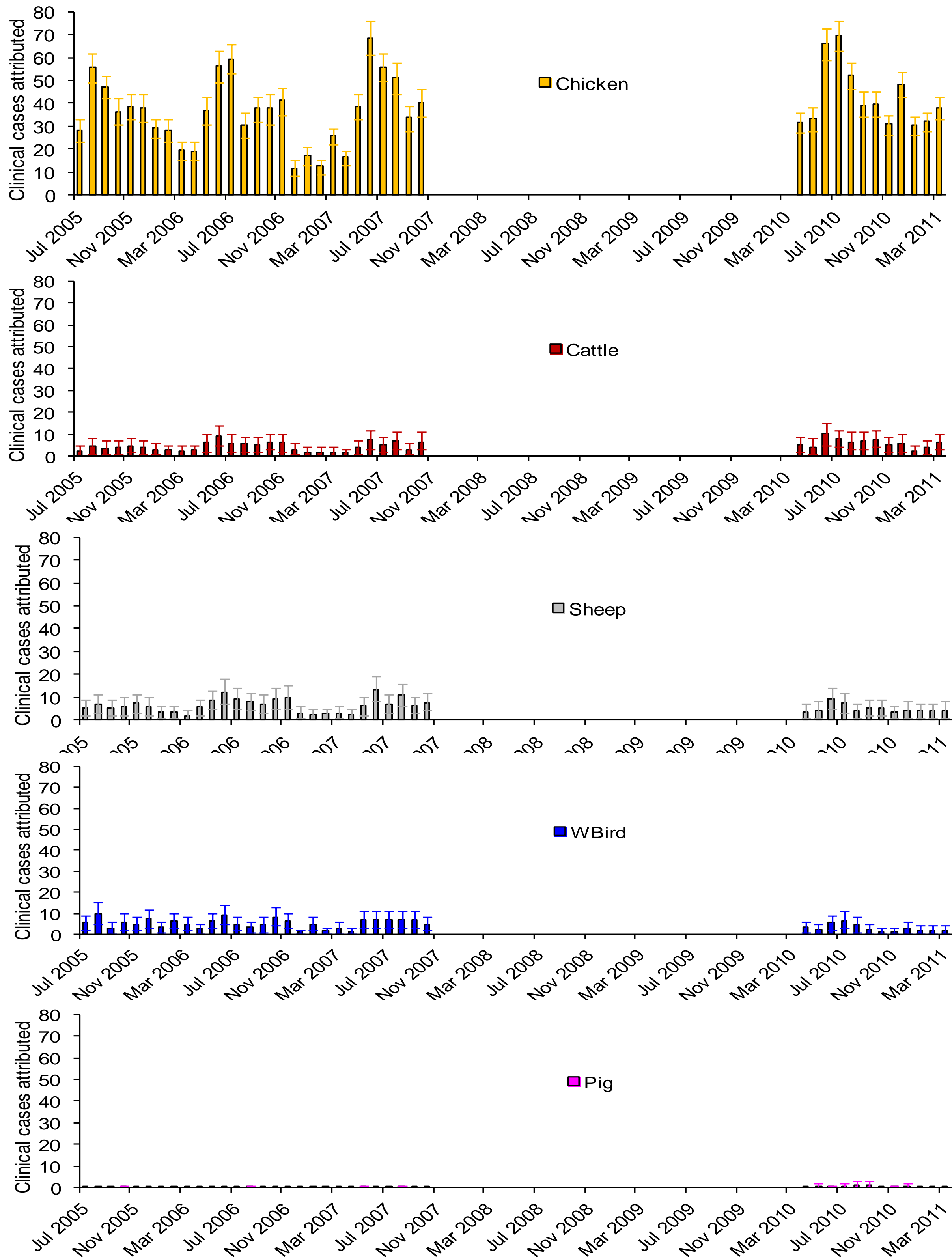
Attribution based on Host Dataset 1. 90% CI.

Supplementary Figure 2. Molecular attribution of clinical *Campylobacter* by STRUCTURE with alleles.



Attribution based on Host Dataset 3. 90% CI.

Supplementary Figure 3. Molecular attribution of clinical *Campylobacter* by Asymmetric Island.



Attribution based on Host Dataset 3. 90% CI.

5. References

1. **Buzby, J. C. and T. Roberts**. 1997. Economic costs and trade impacts of microbial foodborne illness. World Health Statistics Quarterly - Rapport Trimestriel de Statistiques Sanitaires Mondiales **50**:57-66.

2. **EFSA**. 2010. Scientific Opinion on Quantification of the risk posed by broiler meat to human campylobacteriosis in the EU. EFSA Journal **8**:1-89.

3. **Forbes, K. J.** 2009. The Molecular Epidemiology of Scottish *Campylobacter* Isolates from Human Cases of Infection using Multilocus Sequence Typing (MLST). CaMPS -*Campylobacter* MLST Project in ScotlandFood Standards Agency.

4. **French, N. and Molecular Epidemiology and Veterinary Public Health Group**. 2008. Enhancing Surveillance of Potentially Foodborne Enteric Diseases in New Zealand: Human Campylobacteriosis in the Manawatu, p. 1-56. Massey University, New Zealand.

5. **Gillespie, I. A., S. J. O'Brien, and F. E. Bolton**. 2009. Age Patterns of Persons with Campylobacteriosis, England and Wales, 1990–2007. Emerg.Infect.Dis. **15**:2046-2048.

6. **Gormley, F. J., M. Macrae, K. J. Forbes, I. D. Ogden, J. F. Dallas, and N. J. C. Strachan**. 2008. Has retail chicken played a role in the decline of human campylobacteriosis? Appl.Environ.Microbiol. **74**:383-390.

7. **Humphrey, T. J., S. O'Brien, and M. Madsen**. 2007. *Campylobacter* as zoonotic pathogens: a food production perspective. Int.J.Food Microbiol. **117**:237-257.

8. **Little, C. L., F. J. Gormley, N. Rawal, and J. F. Richardson**. 2010. A recipe for disaster: outbreaks of campylobacteriosis associated with poultry liver pâté in England and Wales. Epidemiol.Infect.

9. **Mullner, P., T. Shadbolt, J. Collins-Emerson, A. Midwinter, S. E. F. Spencer, J. Marshall, P. E. Carter, D. M. Campbell, D. J. Wilson, S. Hathaway, R. Pirie, and N. P. French**. 2009. Molecular and spatial epidemiology of human campylobacteriosis in New Zealand - source association and genotype-related risk factors. Epidemiol.Infect. **submitted**.

10. **NELSON, W. A. R. R.** 2010. Campylobacteriosis in New Zealand. Epidemiol.Infect. **138**:1762-1764.

11. **Pires, S. M., E. G. Evers, W. van Pelt, T. Ayers, E. Scallan, F. J. Angulo, A. Havelaar, and T. Hald**. 2009. Attributing the Human Disease Burden of Foodborne Infections to Specific Sources. Foodborne Path.Dis. **6**:417-424.

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

13. **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.

14. **Sheppard, S. K., F. Colles, J. Richardson, A. J. Cody, R. Elson, A. Lawson, G. Brick, R. Meldrum, C. L. Little, R. J. Owen, M. C. J. Maiden, and N. D.**

McCarthy. 2010. Host Association of *Campylobacter* Genotypes Transcends Geographic Variation. Appl.Environ.Microbiol. **76**:5269-5277.

15. **Sheppard, S. K., J. F. Dallas, N. J. C. Strachan, M. Macrae, N. D. McCarthy, D. Falush, I. D. Ogden, M. C. J. Maiden, and K. J. Forbes**. 2009. *Campylobacter* Genotyping to Determine the Source of Human Infection. Clin.Infect.Dis. **48**:1072-1078.

16. **Skirrow, M. B.** 1977. Campylobacter enteritis: a "new" disease. BMJ **2**:9-11.

17. **Strachan, N. J. C. and K. J. Forbes**. 2010. The growing UK epidemic of human campylobacteriosis. Lancet **376**:665-667.

18. **Tam, C. C., L. C. Rodrigues, L. Viviani, J. P. Dodds, M. R. Evans, P. R. Hunter, J. J. Gray, L. H. Letley, G. Rait, D. S. Tompkins, and S. J. O'Brien**. 2011. Longitudinal study of infectious intestinal disease in the UK (IID2 study): incidence in the community and presenting to general practice. Gut .

19. **Tustin, J., K. Laberge, P. Michel, J. Reiersen, S. Dadadottir, H. Briem, H. Harardottir, K. Kristinsson, E. Gunnarsson, V. Friariksdottir, and F. Georgsson**. 2011. A National Epidemic of Campylobacteriosis in Iceland, Lessons Learned. Zoonoses and Public Health **58**:440-447.

20. **van den Brandhof, W. E., G. A. De Wit, M. A. S. de Wit, and Y. Van Duynhoven**. 2004. Costs of gastroenteritis in The Netherlands. Epidemiol.Infect. **132**:211-221.

21. **Wilson, D. J., E. Gabriel, A. J. H. Leatherbarrow, J. Cheesbrough, S. Gee, E. Bolton, A. J. Fox, P. Fearnhead, C. A. Hart, and P. J. Diggle**. 2008. Tracing the Source of Campylobacteriosis. PLoS Genet **4**:e1000203.