Final report: FDI / 236 /2005
Enhancing Surveillance of Potentially Foodborne Enteric Diseases in New Zealand:
Human Campylobacteriosis in the Manawatu

prepared by Professor Nigel French
and the Molecular Epidemiology and Veterinary Public Health Group

30th September 2008

Hopkirk Institute
Institute of Veterinary, Animal and Biomedical Sciences
College of Sciences
Massey University
New Zealand
Email - N.P.French@massey.ac.nz
Phone - +64 (06) 356 9099 extn 81188
Fax - +64 (06) 350 5714

Contents

1 Executive summary

Human health surveillance is an essential element of the monitoring and review component of the New Zealand Food Safety Authority's risk management framework. This study was undertaken as part of the human surveillance strategy, to inform the aetiology of human campylobacteriosis in a sentinel area, using novel approaches to source attribution. The molecular epidemiology of human campylobacteriosis was studied for a three-year period in the Manawatu between March 2005 and February 2008. Using the advanced molecular genotyping technique of Multilocus Sequence Typing (MLST), combined with statistical modelling tools, we have estimated the relative contribution of different food and environmental sources to human infection with *Campylobacter jejuni*. Further, by linking 89.5% of the human cases with genotyping data to information provided by the EpiSurv surveillance system, we have gained new insight into the epidemiology of this disease in New Zealand.

Source attribution modelling of 521 human cases over the 3-year period identified poultry as the most important source of infection. Four approaches were used to estimate the relative contribution of each source to the burden of human cases, of which three (the Dutch, Hald and Island models) provide a direct estimate of the proportion of cases attributable to each source. These three models estimated the contribution of poultry to be 52% (Dutch model), 67% (Hald model) and 75% (Island model). The next most importance source was estimated to be cattle, contributing 17% (Dutch), 23% (Hald) and 17% (Island) to human cases. Smaller contributions were estimated to come from sheep, wild birds and environmental water. Each model is based on different underlying assumptions and these are described in detail in the body of the report and the appendices.

A single multilocus sequence type, namely ST-474, accounted for 28%, 34% and 27% of the human cases in 2005, 2006 and 2007 respectively. To find such a dominant strain, and one that is relatively rare internationally, was surprising and unexpected. This strain was one of several that were strongly associated with poultry, whereas others, such as ST-61, ST-42 and ST-2026, were associated with ruminant sources. Mapping of the human case rates at meshblock level, using EpiSurv data, revealed that ST-474, along with other poultry-associated strains such as ST-45 and ST-48, was particularly prevalent in urban areas, whereas the human cases associated with ruminant strains were predominantly found in rural areas. There were major differences in the age distribution and occupation of ruminant-associated cases compared to poultry associated cases. Compared to poultry-derived cases, ruminant associated cases were relatively more common ($P=0.02$) in young children (<10 years of age) and in adults with an occupation that would bring them in contact with ruminant faeces ($P<0.00001$). This suggests that environmental and occupational exposures may be relatively more important for exposure to ruminant genotypes.

Using a sentinel surveillance site, these studies have provided the most conclusive evidence to date of the importance of poultry as a source of human infection, most notably in urban areas. Further, they have underlined the importance of ruminants as a source of infection, especially in young children and in rural areas. They suggest that a control programme aimed at reducing contamination of fresh poultry could at least halve the number of human campylobacteriosis cases in the Manawatu and, given the comparable data in other areas such as Auckland and Christchurch, New Zealand as a whole. In order to reduce the incidence beyond what can

be achieved by implementing controls in the poultry industry, interventions aimed reducing childhood and occupational exposure to ruminant faeces may need to be considered.

2 Introduction

Evidence from outbreak investigations and epidemiological studies of human enteric diseases are being used increasingly as sources of data for risk assessments. However their application is often restricted by the strength of the evidence presented and its interpretation. They also rarely recognise that most investigations are performed, analysed and interpreted in the context of urgent disease control needs rather than as planned aetiological studies. A range of reports has described deficiencies in the present public health investigation and management of identified cases of human enteric diseases.

NZFSA Science Strategy has identified human health surveillance as an essential element of the monitoring and review component of its risk management framework. Recently a multi-agency Human Enteric Disease Surveillance Steering Committee has been established to provide a strategic direction of human enteric disease surveillance to ensure there is a co-ordinated system in New Zealand which assists in the reduction of the disease burden of human enteric disease. A paper, Enhanced Sentinel Surveillance for Enteric Disease in New Zealand: the advantages, disadvantages and feasible options, was circulated for comment. Based on the comments received, three priorities have been identified. Among these is: 'developing a prototype sentinel surveillance programme for campylobacteriosis'.

Sentinel surveillance systems involve selected reporting site(s) from which information collected may be extended to the general population. A concentration of resources in the defined site(s) aims to produce a rich source of information, producing more accurate final estimates than those available normally from broader national surveillance programmes. More comprehensive than traditional passive approaches, this surveillance includes:

- sustained co-ordination among the local health system;
- specialised laboratory involvement/partnerships;
- standardised microbiological and epidemiological methods, which provide useful information for pathogen or disease risk assessments; and
- timely and effective reporting and communications.

Source attribution is the process of determining what proportion of a particular disease is acquired from a given source (e.g. chicken) and through a given pathway (e.g. food, water, person-to-person transmission). While it is critical to drawing informed conclusion about safety of food in New Zealand, source attribution of pathogens is rarely accomplished. Reasons for this include inconsistencies in the traditional methods of gathering data on sporadic cases of communicable diseases and disease outbreaks, the fact that smaller outbreaks often go unnoticed, and the difficulties of conducting laboratory analysis of both human and environmental (including food and water) samples.

C. jejuni has been isolated from a wide range of food sources including poultry, red meat and milk as well as from environmental sources such as wild birds and water. Poultry is the most commonly implicated source of human infection from epidemiological and microbiological studies; however the evidence for this association is often weak and circumstantial. The capacity to attribute cases of human disease to a food vehicle or another source responsible for illness is critical for the identification and prioritisation of food safety interventions and a variety

of approaches are used worldwide [1]. In addition, the identification of effective food safety interventions requires an understanding of the relationship between food and pathogen from farm to consumption. This assessment of the origin of human infection is a major global public health issue and classical epidemiological approaches such as case-control studies [7] have been conducted to fulfill this purpose. These methods may give ambiguous or conflicting results and often fail to provide risk managers with sufficient information to deploy effective mitigation strategies to lower the burden of human disease.

The most straightforward way to quantify the effect of exposure would be to estimate the numbers of cases that were caused by this exposure. This number is not estimable from ordinary incidence data, because the observation of an exposed case does not reveal the mechanism that caused the disease [11]. In contrast to an outbreak situation, where the attributable risk fraction for an identified risk factor would be very high, the source of infection in sporadic cases is more difficult to identify [18].

Different species of bacteria can have a different epidemiology, as illustrated for *Campylobacter jejuni* and *C. coli* [10]. This could also apply to different strains and would limit the inferences made at the species level by classical epidemiological investigation. Additionally, the success of epidemiological methods may depend on the study's sensitivity, particularly when seeking to associate statistically very small risks relative to the background, such as when analysing sporadic cases of endemic diseases. This is why the study of endemic disease risks becomes much more difficult, in particular when the epidemiology of disease is complex, including a large amount of risk factors and potential sources.

The recent advances in molecular biology enable us to classify bacteria on a much more discriminatory level, and identify genotypes associated with particular host animals. Genotyping allows for computer assisted data analysis in a numeric manner, catering for data sharing and easing the processing of large numbers of samples [5]. For *Campylobacter* Multilocus Sequence Typing (MLST) has recently emerged as a typing technique and has provided valuable insight into the population structure and evolutionary behaviour of this pathogen. It is based on the sequence of seven, relatively conserved, housekeeping genes and is an efficient and accurate way of isolate characterisation that is highly reproducible and its results can be shared via the Internet [5]. This technique has sufficient resolution to accommodate a high level of diversity present in a bacterial pathogen population, but can also rationalize this diversity into groups of isolates with related genotypes. It is particularly suited to long-term and global epidemiology as it identifies variation that is accumulating slowly within a population, but can be used to investigate individual outbreaks, especially when combined with other data [4]. Data generated can be applied to the investigation of population structure and evolutionary mechanisms [3], as well as to determine the relative importance of the food chain, and its individual components, as sources of human *Campylobacter* infection.

An NZFSA funded PhD programme was started in February 2005 and is being carried out by a student (Petra Mullner) registered at Massey University under the supervision of Professor Nigel French, Dr Cord Heuer and Professor Steve Hathaway in collaboration with Dr Phil Carter at ESR, Kenepuru. The PhD programme included a pilot study assessing the feasibility of using multilocus sequence typing (MLST) of human, food and environmental *Campylobacter jejuni*, combined with modelling to assess the relative contribution of different pathways to human

infection.

The new project described in this report extended the work carried out in the PhD programme by providing the following:

- an assessment of the feasibility of using several model-based tools including the 'Hald' Bayesian Risk Attribution Model, and models based on population genetics, to assess the relative contribution of different food pathways to human campylobacteriosis in the Manawatu; and
- greater understanding of the epidemiology of campylobacteriosis in New Zealand (e.g. seasonality of different genotypes, spatial distribution and risk factors) by enhancing and incorporating quality epidemiological information on each case.

Criteria used to select the sentinel site include:

- a population of around 250,000;
- an urban/rural mix;
- laboratory capacity for *Campylobacter* spp - both diagnostic and multilocus sequence typing;
- innovation in both local public health and academic communities; and
- willingness to participate.

The Manawatu was chosen because of existing interests in campylobacteriosis that lie within both the MidCentral Health Public Health Service and Massey University. In addition MidCentral Health Public Health Services has separate offices serving Manawatu and Whanganui Health Districts, allowing the comparison of any human surveillance initiatives piloted.

The Manawatu study consists of structured parallel studies of isolates from human cases of campylobacteriosis and environmental and foods sources in a defined geographical area of New Zealand over a 3 year period. This period included a large un-seasonal national epidemic that occurred in the winter of 2006 [15]. Using the dataset from this study, several risk attribution models and risk research tools have been developed and used to assess the relative contribution of food and non-food sources to the burden of human disease. Isolates found in New Zealand are also compared with those from overseas using the PubMLST database (http://pubmlst.org/). Since each technique is based on different assumptions, a combination of methods is used to gain a better understanding of the epidemiology of this pathogen.

3 Methods

C. jejuni were isolated from human clinical cases, and food and environmental sources over a three-year period between March 2005 and February 2008 in the Manawatu. The isolates were genotyped using multilocus sequence typing (MLST) and the genotype distributions were compared using four methods, of which three were model-based tools designed to estimate the number of human cases attribuatble to each source, with an estimate of uncertainty. The output from the three model-based source attribution tools were compared. Further comparisons were made with external validation studies conducted in Auckland and Christchurch, and a subset of samples were subjected additional sequence typing of the *fla*A gene. Data derived from the Episurv surveillance system were linked to the typing data and used to gain further insight into the epidemiology of campylobacteriosis in the Manwatu.

3.1 Data collection for the Manawatu study

3.1.1 Human faecal samples

Human specimens submitted to MedLab Central, Palmerston North that were positive for *Campylobacter* by ELISA (ProSpecT®, Remel, USA) were sent to the Hopkirk Molecular Epidemiology laboratory over the 3 year period 1st March 2005 to 29th February 2008. Faecal swabs were made using Amies Charcoal transport swabs (Copan, Italy).

3.1.2 Epidemiological data from human cases

Three years of epidemiological human data were acquired - initially collected using routine Public Health Service (PHS) methods, followed by a sentinel surveillance approach with data collection targets. The MidCentral Public Health Services (MCPHS), working with ESR Ltd, provided the project with a link to the national disease database (Episurv) where human epidemiological data gathered in the field are logged. Information gathered by MCPHS between February 2005 and June 2006 was acquired using both questionnaires and telephone interviews. From the 1st July 2006 the MCPHS received concurrent funding from NZFSA to enhance the quality of the surveillance data gathered within Manawatu. From that date, contact of notified cases of campylobacteriosis was via telephone with a target of 95% of cases interviewed and 95% of data sets completed. The interviews were conducted using the Episurv Case Report Form (CRF) format. The MedLab number was recorded in the comments field of the CRF. From the November 2006 the CRF questions were enhanced with additional questions relating to meat eaten and the consumption of unpasteurised milk during the cases incubation period. The results of these questions were recorded in a separate Access database, including an 'unknown' and 'no contact' option, lnked via the Episurv and MedLab identification numbers.

During this time period there was also a significant change to Episurv with the launch of SURVINZ Episurv V 7.2.1 in April 2007. This resulted in a change in the format of the data collected, and initiated the recording of investigation methods.

3.1.3 Food samples

Over the three-year period from March 2005 to Feb 2008, 12-18 fresh whole poultry carcases were sampled each month from retail outlets in Palmerston North, representing the different poultry suppliers. In addition, 45 samples of fresh red meat and offal (pork, beef and lamb mince and lamb and beef liver) were sampled monthly from retail stores in Palmerston North.

3.1.4 Environmental samples

Six sites, identified as high use recreational swimming spots by Horizons Regional Council, were sampled every two weeks over the three-year period, and cattle and sheep faeces from farms in the catchment of these river sources were sampled. These sites are shown in Figure 1. Water was collected in sterile 200ml bottles approximately 1 m from the waterway's edge and transported in the dark at ambient temperature to the laboratory for testing. 100ml of water was passed through a sterile 0.45μm filter (Sartorius, Germany) and the filter was immersed in 20ml of Bolton Broth.

Water samples were collected from the following sites. Recreational swimming sites: Mangapapa stream, Woodville; Manawatu River, Hopelands Picnic reserve, Hopelands; Oroua River, Timona Park, Feilding; Manawatu River, Albert Street, Palmerston North; Tokomaru River, Horseshoe Bend, Tokomaru and Kaikokopu Stream, Himatangi Beach. Faecal swabs were collected from the centre of fresh faecal pats using transport swabs with Amies Charcoal. Four swabs were taken from each farm site.

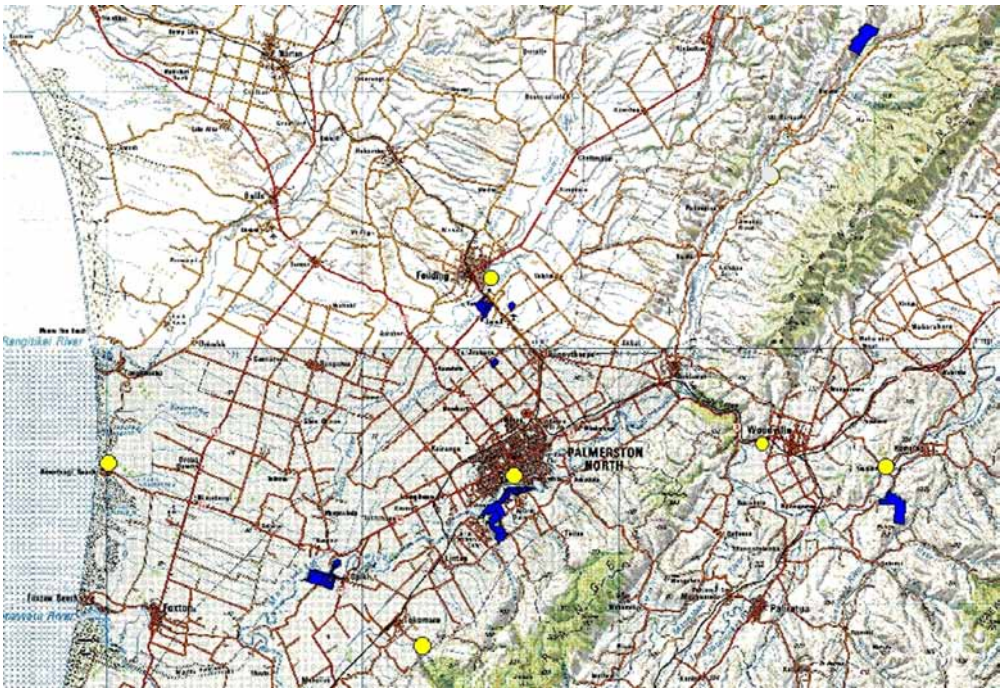

Figure 1: Showing the location of river sampling points (yellow circles) and ruminant faecal material (blue polygons)

3.1.5 Laboratory methods

Samples were cultured for *Campylobacter* spp. in a microaerophillic chamber using standard methods. Isolates were then confirmed as *C. jejuni* by PCR, and genotyped using multilocus sequence typing. A subset were further typed by *fla*A sequencing.

3.1.5.1 Sample preparation and culture techniques Human faecal swabs were cultured on modified Cefoperazone Charcoal Deoxycholate agar (mCCDA) plates (Fort Richard, Auckland) and in Bolton Broth (Lab M, Bury, England) and incubated at 42°C in a microaerobic atmosphere (85% N_2, 10% CO_2, 5% O_2) for 2 days. A single colony resembling *Campylobacter* species was subcultured to Blood Agar (BA) (Fort Richard, Auckland) and incubated microaerobically at 42°C for 2 days before DNA preparations were made. Cultures were frozen at -80°C in Glycerol Broth (Difco, USA).

Chickens were washed and massaged in 200ml of Buffered Peptone Water (BPW) (Difco, USA). The chicken wash was centrifuged (10,000 rpm, 6°C, 35 mins, Sorvall RC5B) and the resultant pellet resuspended in 5ml of BPW. Approximately 3ml of the resuspended pellet was added to 90ml of Bolton Broth, which was incubated at 42°C microaerobically for 2 days. After incubation, the broth was subcultured onto mCCDA and incubated microaerobically at 42°C for 2 days. Single colonies resembling *Campylobacter* species were subcultured to BA and incubated microaerobically at 42°C for 2 days before DNA preparations were made. Cultures were frozen at -80°C.

From October 2006, *Campylobacter* on all chicken carcases were enumerated using a Wasp spiral plater (Don Whitley, England) and a manual spread plate. Duplicate mCCDA plates were inoculated with 50μl (spiral plater) or 1ml (spread plate) aliquots of chicken wash or 100μl (spiral plater) aliquots of resuspended chicken wash pellet. The plates were incubated microaerobically at 42°C for 2 days. Colonies were counted manually or by using a plate reader (aCOLyte, Synbiosis, England).

For red meat sampling, 10g (+/- 1g) of mince or liver was aseptically removed from the packet and stomached briefly with 90ml of Bolton Broth. The Bolton Broth plus meat was incubated at 42°C microaerobically for 2 days. After incubation the broth was subcultured onto mCCDA and incubated microaerobically at 42°C for another 2 days. Single colonies resembling *Campylobacter* species were subcultured to BA and incubated microaerobically at 42°C for a further 2 days before DNA preparations were made. Cultures were frozen at -80°C.

For the environmental water samples, the broth containing the filter was incubated at 42°C microaerobically for 2 days. After incubation, the broth was subcultured onto mCCDA and incubated microaerobically at 42°C for 2 days. Single colonies resembling *Campylobacter* species were subcultured to BA and incubated microaerobically at 42°C for 2 days before DNA preparations were made. Cultures were frozen at -80 °C.

The four swabs from each of the cattle and sheep environmental sites were immersed together in 20ml of Bolton Broth and incubated microaerobically at 42°C for 2 days. After incubation the broth was subcultured onto mCCDA and incubated microaerobically at 42°C for 2 days. Single colonies resembling *Campylobacter* species were subcultured to BA and incubated microaerobically at 42°C for 2 days before DNA preparations were made. Cultures were frozen at -80°C.

3.1.5.2 Species confirmation by PCR Species confirmation by PCR. DNA was extracted from freshly grown cultures by boiling for 10 min in the presence of 2% Chelex (Biorad), followed by centrifugation to remove both cell debris and the Chelex, which inhibits PCR. The supernatant, containing the nucleic acids, was transferred to a fresh tube and used for amplification PCR and MLST. The isolates of *Campylobacter* were speciated by multiplex PCR to detect genes associated with either *C. jejuni* or *C. coli*. The mapA gene was shown to be found only in *C. jejuni* [21], so primers MapA-F (5'-CTTGGCTTGAAATTTGCTTG-3') and MapA-R (5'-GCTTGGTGCGGATTGTAAA-3') were designed to target this gene for its identification. Detection of *C. coli* was performed using primers (COL3 and MDCOL2)[2]. These two sets of primers were combined into one PCR reaction (multiplex PCR) for the simultaneous identification of the two species of *Campylobacter*. Amplification was performed in a 20 μl reaction containing 1 unit of Platinum Taq Polymerase (Invitrogen), 100 μM of each dNTP, 200 nM of each primer (MapA-F, MapA-R, COL3, and MDCOL2), and 1.5 mM MgCl2. The reactions were carried out in an Applied Biosystems 9700 Thermocycler by heating the sample to 96°C for 2 mins, followed by 38 cycles of 96°C for 30 sec, 58°C for 30 sec and 72°C for 30 sec, with a final extension of 72°C for 2 mins. The PCR products were visualised by subjecting a 5μl aliquot to electrophoresis in a 1% agarose gel in TBE buffer, which was then stained with ethidium bromide and exposed to UV light. The presence of a 603bp product indicated *C. jejuni* while a 462bp one indicated *C. coli*.

3.1.5.3 Multilocus sequence typing After speciation, MLST of *C. jejuni* isolates was performed using seven house-keeping genes: *asp*A (aspartase A), *gln*A (glutamine synthase), *glt*A (citrate synthase), *gly*A (serine hydroxymethyltransferase), *pgm* (phosphoglucomutase), *tkt* (transketolase) and *unc*A (ATP synthase alpha subunit) based on the method as outlined by Dingle et al., 2001[5]. Each amplification reaction comprised 2μl of the DNA preparation, 5 pmoles of both forward and reverse amplification primers, 12.5μl of ABI 2x AmpliTaq Gold PCR Mastermix and water to make up to a total volume of 25μl. Amplification was performed on a Corbett Palm Cycler under the following conditions: Initial denaturation was for 15 mins at 94°C followed by 35 cycles of 94°C denaturation for 30 sec, 50 °C annealing for 30 sec and 72°C extension for 90 sec. Final extension was for 72°C for 7 mins. PCR products were precipitated with 25μl 20% PEGS/2.5 M NaCl solution, washed with 80% EtOH, dried and taken up in 12μl H_2O and screened on agarose gels. Sequencing reactions were performed using 2μl of PCR product, 3.2pmoles primer, 2μl ABI BigDye, 2μl of x5 BigDye buffer and water to a total volume of 10μl. Reaction were performed under the following conditions. Initial denaturation at 96°C for 3 min. then 25 cycles of 96°C for 15 sec, 50°C for 15 sec and 60°C for 4min. Sequenced products were precipitated with 0.1M Na acetate/78% EtOH solution, washed with 70%EtOH, dried and taken up in 12μl H_2O and the sequence read at ESR, Kenepuru, on an ABI 3130XL automated DNA sequencer using ABI BigDye v3.1 (Applied Biosystems) following the manufacturer's instructions. Sequence data were collated by Dr Phil Carter at ESR, and alleles assigned using the Campylobacter PubMLST database (http://pubmlst.org/campylobacter/). Novel alleles and sequence types were submitted for allele and ST designation as appropriate and alleles that did not give clear results were re-amplified and sequenced using primers sets published by Miller et al., (2005)[16] using the same protocol as above.

3.1.5.4 *fla*A short variable region typing In addition to the seven housekeeping gene MLST scheme, we assess whether further insight can be gained by the addition of a less conserved gene. A subset of 69 isolates belonging to ST-45 and ST-474, from a range of sources were selected for this analysis. PCR was performed with a primer set targeting the Short Variable Region (SVR) of the flagellin A gene (*fla*A) in *C. jejuni*[4]. The outer forward primer FLA4F (5'-GGATTTCGTATTAACACAAATGGTGC-3') and inner reverse primer FLA625RU (5'-CAAGWCCTGTTCCWACTGAAG-3') were used in this study. All PCR reactions (20μl) initially contained 1.5mM $MgCl_2$, 100μM of (each) dATP, dCTP, dGTP and dTTP; 1μl (2pmole) of each primer (Invitrogen), 1x PCR buffer and 1 U of Platinum Taq DNA polymerase (Invitrogen). In the cases where only a weak amplicon was obtained, the reaction was repeated but with the MgCl2 concentration raised to 2.5 mM. The PCR was performed in a GeneAmp® PCR System 9700 (Applied Biosystems) for 37 cycles. Each cycle included denaturaturing at 94°C for 30 sec, annealing at 64°C for 30 sec and extension at 72°C for 45 sec. The PCR product was analysed by electrophoresis (1 hour at 100V) on 1% Ultrapure^{TM} agarose (Invitrogen) in 1x TBE buffer. The gel was stained with ethidium bromide solution then photographed under UV. The product size was 640bp. The PCR product was purified using PureLink^{TM} PCR Purification Kit (Invitrogen). The purified amplicon was quantified using Low Mass DNA Ladder (Invitrogen). The quantified product was then sent for sequencing service (Allan Wilson Centre, Massey University, Palmerston North). The peptide sequences encoded by the SVR nucleotide sequences were analysed. The 321bp unique amino acid sequence was assigned *fla*A SVR allele variant number using the Oxford Campylobacter *fla*A database (http://hercules.medawar.ox.ac.uk/flaA/).

3.2 Comparative studies in Auckland and Christchurch

Three additional studies were conducted to assess the generalisability of the study conducted in the Manwatu; one in Auckland in 2007, and two in Christchurch (2007 and 2008). All were designed to provide MLST data that could be compared with the ongoing study in the Manawatu.

Between 24 June to 28 August 2007, isolates from medical laboratories in Auckland and Christchurch were collected on a weekly basis and sent to the ESR Public Health Laboratory in Christchurch for purification, identification and DNA extraction. Each isolate was collected on a charcoal swab (Amies swab, Copan Innovations, Italy) from the primary *Campylobacter* positive culture plate (mCCDA) streaked from the diarrhoeal specimen of clinical cases submitted to the laboratory by General Practitioners for microbiological diagnosis. In Auckland, there is only one community provider of medical diagnostic services so all clinical isolates were collected from one laboratory. In Christchurch, the two community providers of medical diagnostic services were asked to provide a selection of isolates each week. Using the sampling plan, up to 96 isolates from human diarrhoeal patients were anticipated from each city.

Over the same time period whole chilled chickens were purchased from supermarkets and other retail outlets twice a week for eight weeks and tested using similar methods to those described for the Manawatu study. In Auckland and Christchurch poultry was available from poultry suppliers A, B and D.

Full details of the sampling strategy and results from these concurrent studies are provided in two ESR technical reports: 'MLST Genotyping of *Campylobacter* isolated from human cases

and retail poultry in Canterbury following Sanova intervention at a poultry processing plant'; and 'Comparison of Human and Poultry *Campylobacter jejuni* isolates utilising Multi-Locus Sequence Typing in Auckland and Christchurch with those isolated from the Manawatu'.

3.3 Data analysis

3.3.1 Enumeration of *Campylobacter* on poultry carcases

This study was started in October 2006. The aim was to estimate both the proportion of carcases positive and the levels of *Campylobacter* present on positive carcases. The laboratory techniques are standard, but the method of analysis employs a novel application of recently developed statistical tools for the analysis of count data where there are a large proportion of zeros, and several replicates at the sample level.

3.3.1.1 Statistical analysis of *Campylobacter* count data Bacterial count data has two components: firstly whether the bacteria present or absent, and secondly, if it is present, how much of it is there. In order to capture these two components in our model we used a Bayesian zero-inflated Poisson model. The model features two linear predictors: one informing the probability that the bacteria will be present and the other informing the amount of bacteria that will be obtained given that it is present. Full details of the model are provided in Appendix A. The output from these models is presented as a series of graphs describing the probability of a carcase containing *Campylobacter*, by supplier and by quarter, and the estimated number of viable *Campylobacter* on positive carcases - again by supplier and quarter. This method ensures that all the individual replicate counts for each sample are analysed appropriately.

3.3.2 Minimum spanning trees

Minimun spanning trees (MSTs) are a graphical tool available in the BioNumerics software (Applied Maths; http://applied-maths.com/bionumerics/bionumerics.htm). The technique uses a clustering algorithm, designed for use with MLST data, that incorporates a simple but realistic model of the way in which bacterial clones emerge and diversify to form clonal complexes. MSTs were developed as an alternative way of visualising the relationships among isolates from bacterial populations that overcomes some of the problems associated with other phylogenetic tools. This approach does not impose a tree-like pattern of descent, which is particularly important for investigating species with frequent recombination events like *Campylobacter*. In the analysis presented here, MSTs are a convenient way of representing the distribution and diversity of sequence types (STs) amongst the difference sources, whereby each ST is represented as a pie chart. The size of the circle indicates the number of isolates, and each coloured segment represents the proportion from each source.

3.3.3 Source attribution modelling

Four methods were compared to assess the relative contribution of each source to the burden of human infections in the Manawatu. The first (Proportional Similarity Index) simply assesses the area of overlap of the genotype distributions from each source with that of the human genotype distribution. The other three estimate the number of human cases attributable to

each source using models (Dutch, Hald and Island models) that are based on different underlying assumptions.

3.3.3.1 Proportional Similarity Index The proportional similarity index (PS) is an objective and simple estimate of the area of intersection between two frequency distributions [19]. In this context, the PS estimates the similarity between the frequency distributions of STs of each source and the distribution of STs amongst human cases. The values for PS range from 1, for the highest possible similarity, to 0 for distributions with no common types.

$$\mathrm{PS} = 1 - 0.5 \sum_i |p_i - q_i|$$

where $|p_i - q_i|$ is the modulus of the difference in the relative frequency of MLST genotype i in source p compared to source q.

Bootstrap confidence intervals for this measure were estimated based on the approach of Garrett et al. [9].

3.3.3.2 Dutch model The Dutch method compares the number of reported human cases caused by a particular bacterial subtype with the relative occurrence of that subtype in each source. The number of reported cases per subtype and source is estimated by:

$$\lambda_{ij} = \frac{p_{ij}}{\sum_j p_{ij}} x_i,$$

where p_{ij} = relative occurrence of bacterial subtype i in source j,
x_i = estimated number of human cases of type i per year,
λ_{ij} = expected no. of cases / year of type i from source j.

A summation across subtypes gives the total number of cases from source j, denoted by λ_j:

$$\lambda_j = \sum_i \lambda_{ij}.$$

The method of Garret et al. [9] was extended to provide bootstrap confidence intervals for the Dutch model.

3.3.3.3 Modified Hald model We modified the Bayesian risk assessment model originally developed to quantify the contribution of different food sources to the number of human cases of salmonellosis in Denmark [12]. The original model compares the number of human cases caused by different 'types' with their prevalence in different food sources, weighted by the amount of food source consumed. This model is a further development of the frequentist model described in section 3.3.3.2 and requires a heterogeneous distribution of some types among animal and food sources. Like the Dutch model, this approach compares the number of human cases caused by different bacterial subtypes with their prevalence in different food sources. However, by using a Bayesian approach, the Hald model can explicitly include and quantify the uncertainty surrounding each of the parameters. In our study the Hald model was adapted to overcome some of the problems associated overparameterisation and to incorporate uncertainty in the

prevalence matrix. Further, the food consumption terms was removed to enable the inclusion of environmental sources of campylobacteriosis.

The core of the modified Hald model is the model equation $\lambda_{ij} = p_{ij} q_i a_j$ where

- p_{ij} is the prevalence of type i in source j,
- q_i is the bacteria ST dependent factor for sequence type i,
- a_j is the food source dependent factor for food source j.

Full details of the model are provided in Appendix B.

3.3.3.4 Island model This new method, published September 2008 [6], is based on coalescent models, which are different from classical phylogenetic methods in their explicit considerations of the genealogical history of sampled alleles [20]. This approach is fundamentally different from the Dutch and Hald models in that it explicitly models the genealogy of all isolates, using their allelic profiles and taking into account the relatedness of STs.

Island models were first proposed by Wright, 1931 [22] and are models of gene flow derived from population genetics. The technique devised by Wilson et al 2008 [6] reconstructs the genealogical history of the isolates, based on their allelic profiles, and estimates mutation and recombination rates, as well as the 'migration' rates from each source into the human 'Island'. It is these migration rates that are used to estimate the relative contribution from each source. Importantly this technique has one major advantage over the other methods; it can assign human cases that have no identified reservoir in the animal or environmental reservoirs. Further details of the model are provided in Appendix C.

3.3.4 Comparison of the epidemiology of ruminant and poultry associated human cases

3.3.4.1 Case-case comparison of risk factors The probability that each case arose from either a poultry or ruminant source is derived using population genetics software designed to determine the structure of populations, and the association of particular clusters with common sources [17]. STs with a probability of 0.6 or greater of having been derived from either a poultry or ruminant source, were assigned to that source. We then compared the features of poultry-associated cases with those of ruminant associated cases using contingency tables. We also considered the cases who were overseas travellers during the incubation period and examined the ST distribution associated with these cases compared to the non-overseas travellers.

3.3.4.2 Spatial and temporal analysis The spatial location of each notified human case was given at the meshblock level: the smallest regions defined for the census, each containing between zero and about 200 individuals. In urban areas this gives a very fine spatial resolution, and even in rural areas where the meshblocks have a much larger area, the resolution appears to be sufficiently fine for this analysis. The relative risk of being a case of campylobacteriosis in each meshblock is described using a Bayesian hierarchical model. Relative risk surfaces are prepared for human cases attributable to the common poultry-associated STs, and for ruminant associated STs. Details of the model are provided in Appendix D.

The smoothed time series is obtained by averaging the number of cases in the three weeks before, the three weeks after and the current week. At the ends of the time series, missing values are ignored. This smoothing enables us to see the underlying trends more clearly without the distraction of noise.

4 Results

4.1 Sample details

Table 1 shows the number of samples processed between 1/3/2005 and 29/2/2008.

Source	Number of samples processed
Environment	602 (335 water, 267 ruminant faeces)
Human	773 primary samples (1035 in total)
Poultry	568
Red meat and offal	1312
Total	3517

Table 1: The number of samples processed over the 3 year study period

4.2 Human cases

4.2.1 Primary samples

A detailed examination of laboratory, hospital and EpiSurv data revealed that only 773 (75%) of the 1035 human samples submitted between 1/3/05 and 29/2/08 could be considered "primary samples" that originated in the Manawatu. The remainder were duplicate samples from the same individual, or samples from outside the Manawatu region. Of these, 661 (85%) were culture positive and 585 (83%) of these were confirmed as *C. jejuni* by duplicate PCR and 9 are confirmed *C. coli* by PCR. A total of 572 primary cases provided MLST data, and of these 521 provided complete allelic profiles.

Figure 2 shows the number of primary human samples submitted, culture positive (presumptive) and sequenced over the three-year period. The higher incidence in the summer months and the national winter epidemic in 2006 [15] are clearly visible.

4.2.2 Distribution of MLST genotypes in the Manawatu

Fifty six different multilocus sequence types (STs) were isolated from 521 human samples. The relative frequencies of all STs occurring in human cases in the Manawatu between 1/3/05 and 29/2/08 are summarised in Figure 2. Sequence type ST-474, a member of clonal complex 48, was the most dominant throughout the study period: this ST accounted for the largest proportion of cases in 2005, 2006 and 2007 (Figure 3), and accounted for 30.1% of all human cases over the study period. The second most common human sequence type was ST-48, the founder strain for clonal complex 48. This ST accounted for 8.4% of human cases. Other common STs, occuring in over 20 cases, included ST-45, ST53, ST-190 and ST-354. The frequency of ST-190 isolations appeared to decline over the 3 year period, whereas ST-354 notifications increased over the same time period (Figure 3).

ST-474 was the most prevalent ST during the winter epidemic (May-July 2006), accounting for 33% (14/42) of isolates typed from that period. Both ST-190 and ST-48 accounted for approximately 12% each of the remaining isolates from the epidemic period. These findings are very similar the other DHBs investigated by ESR Ltd during the winter 2006 epidemic. Only

Table 2: Sequence types from human cases in the Manawatu between 1/3/05 and 28/2/08

Allelic profile										
aspA	*glnA*	*glt*	*gly*	*pgm*	*tkt*	*uncA*	ST	CC	Number of cases	% cases
7	2	5	2	10	3	6	5	353	1	0.2
2	1	1	3	2	1	5	21	21	7	1.3
4	7	10	1	1	7	1	25	45	1	0.2
2	4	2	2	6	1	5	38	48	13	2.5
1	2	3	4	5	9	3	42	42	19	3.6
4	7	10	4	1	7	1	45	45	42	8.1
2	4	1	2	7	1	5	48	48	44	8.4
2	1	12	3	2	1	5	50	21	24	4.6
7	17	2	15	23	3	12	51	443	1	0.2
9	25	2	10	22	3	6	52	52	17	3.3
2	1	21	3	2	1	5	53	21	28	5.4
1	4	2	2	6	3	17	61	61	16	3.1
2	4	2	2	6	3	17	81	61	1	0.2
4	7	10	4	42	7	1	137	45	1	0.2
2	1	5	3	2	3	5	190	21	22	4.2
1	4	2	2	6	3	1	219	61	1	0.2
9	2	4	62	4	5	6	257	257	12	2.3
8	10	2	2	11	12	6	354	354	23	4.4
10	27	16	19	10	5	7	403	403	1	0.2
2	1	5	3	2	5	5	422	21	3	0.6
7	21	5	62	4	61	44	436	U/A	4	0.8
2	1	2	3	2	3	5	451	21	8	1.5
1	2	3	3	5	9	3	459	42	1	0.2
2	4	1	2	2	1	5	474	48	157	30.1
2	1	12	88	2	1	5	520	21	8	1.5
1	4	2	2	2	3	17	578	61	1	0.2
4	7	10	4	42	51	1	583	45	10	1.9
2	4	2	4	19	3	6	658	658	2	0.4
10	81	50	99	120	76	52	677	677	6	1.2
33	39	30	82	113	43	17	829	828	1	0.2
2	165	73	147	220	190	104	1457	U	1	0.2
8	10	149	2	11	12	6	1517	354	4	0.8
129	66	30	82	189	47	17	1581	U/A	1	0.2
9	2	5	2	11	3	1	1707	607	1	0.2
10	1	16	19	10	5	7	2026	403	11	2.1
9	53	2	53	11	3	3	2140	574	1	0.2
10	7	10	4	1	7	1	2219	45	1	0.2
2	4	5	2	10	1	5	2343	48	1	0.2
2	4	5	2	2	5	5	2345	206	3	0.6
2	4	1	2	2	5	5	2350	48	2	0.4
2	15	4	48	360	25	23	2391	1034	1	0.2
33	39	30	82	104	173	68	3072	828	1	0.2
33	283	44	82	189	44	17	3222	U/A	1	0.2
47	2	4	2	6	5	17	3538	U/A	1	0.2
1	307	3	4	5	9	3	3676	42	2	0.4
166	2	1	10	17	3	1	3728	U/A	1	0.2
2	1	12	3	357	1	5	3715	21	2	0.4
2	1	21	3	62	1	5	3717	21	1	0.2
2	1	12	3	11	1	6	3713	21	1	0.2
1	2	4	62	4	5	17	3711	257	2	0.4
9	NEW	4	62	4	5	6	u257a	257	1	0.2
8	NEW	2	2	11	12	6	u354	354	1	0.2
1	2	215	4	90	24	8	3712	362	2	0.4
3	1	5	10	11	11	6	3720	49	1	0.2
2	4	1	4	1	1	5	3718	48	1	0.2
48	7	10	4	183	7	1	3727	45	1	0.2

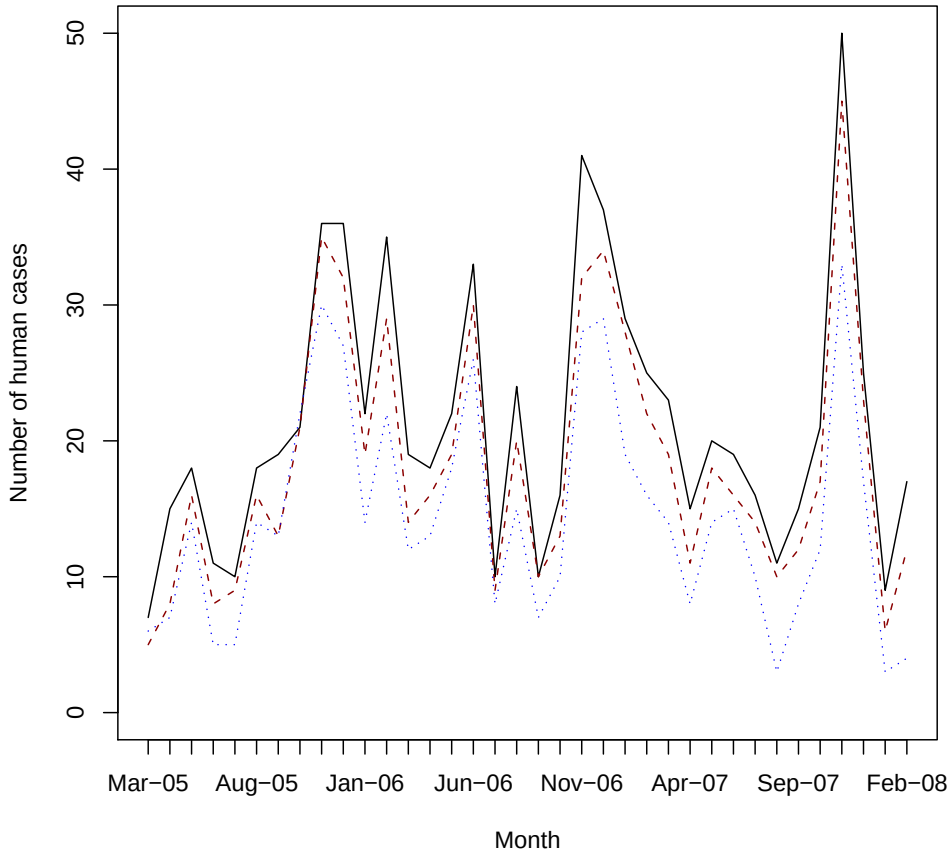

Figure 2: The number of ELISA +ve samples (solid line), culture positive (dashed red line) and fully sequenced (dotted blue line) primary human samples submitted to Massey in each month between March 2005 and Feb 2008.

eight cases between January and February 2008 were confirmed as *C. jejuni*, three of these were ST-474.

4.3 Food and environmental source prevalence and genotype distribution

4.3.1 Prevalence and level of contamination

4.3.1.1 Poultry Retail poultry samples had a high prevalence of contamination with *Campylobacter* spp. throughout the 3-year period. Some 81% of carcases were positive for *Campylobacter*, and 74% were positive for *C. jejuni*. The prevalence of *Campylobacter* spp. in each supplier is shown in Table 3.

The estimated proportion of positive carcases (Figure 4), and the counts conditional on being positive (Figure 5), are shown for each supplier and each quarter from October 2006 February 2008. There is a moderately high probability of contamination for all suppliers throughout the

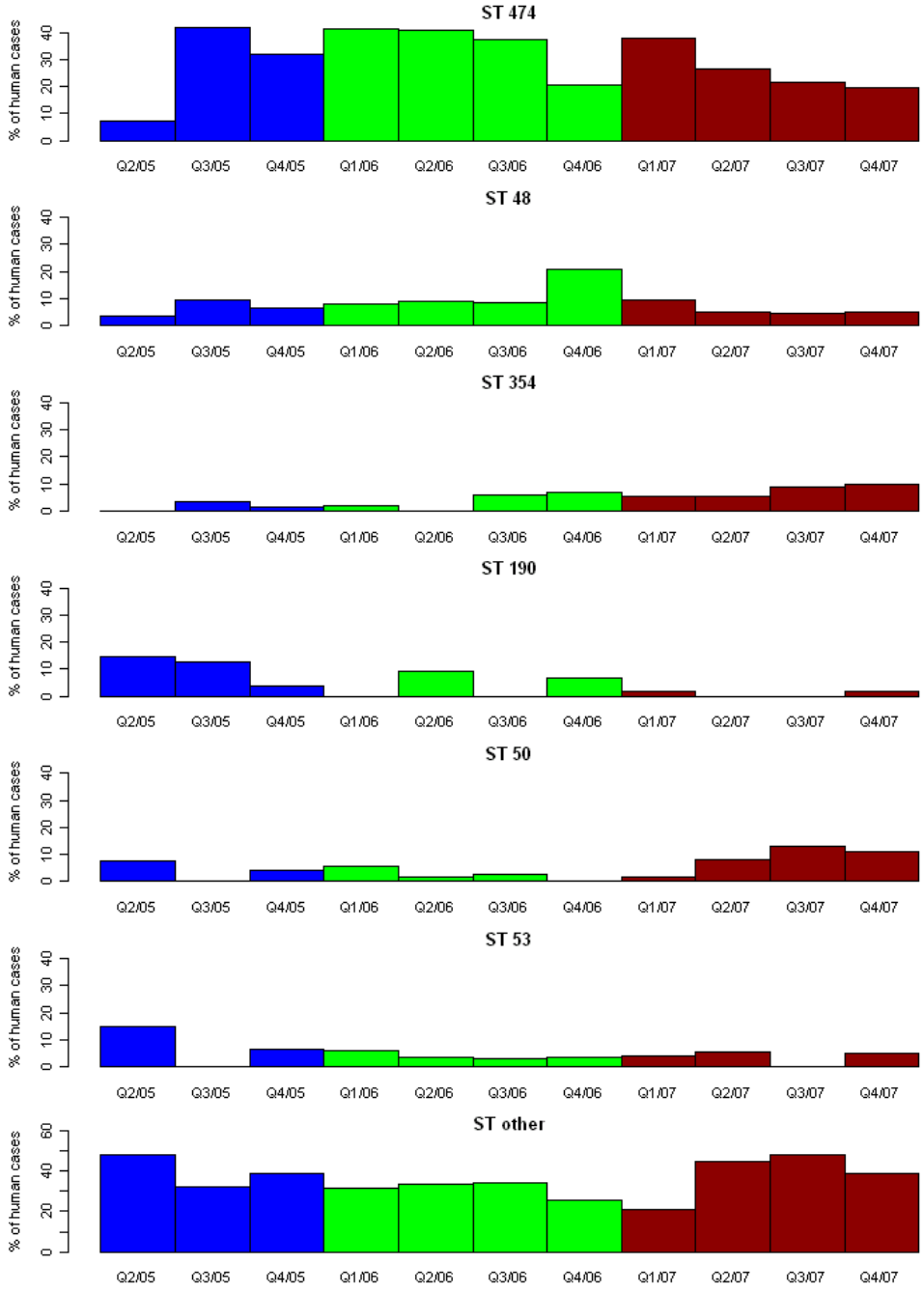

Figure 3: Frequency of all human STs that occurred in more than 20 clinical cases by quarter.

Table 3: Proportion of samples positive for *Campylobacter* spp. in the Manawatu March 2005-Feb 2008.

	Number of samples	Number culture positive	Prevalence (%) presumptive *Campylobacter*	Number confirmed *C. jejuni*	Prevalence (%) *C. jejuni**
Human (primary samples)	773	661	85.5	585	75.7
Chicken	562	454	80.8	415	75.3
Supplier A	239	203	84.9	189	80.7
Supplier B	196	136	69.4	117	60.1
Other suppliers	127	115	90.6	109	88.9
Red Meat	1309	167	12.8	142	11.6
Pork mince	441	11	2.5	6	1.8
Beef mince	171	10	5.8	7	5.3
Beef liver	279	34	12.2	26	9.7
Lamb mince	139	2	1.4	2	1.4
Lamb liver	279	110	39.4	101	38.0
Farm	278	211	75.9	141	50.7
Sheep faeces	133	92	69.2	61	51.1
Cattle faeces	145	119	82.1	80	59.3
Water	335	140	41.8	82	32.2
Total samples	3257	1633		1365	
Total isolates	**4506**				

*% samples speciated

study period, with over 80% of carcases from Supplier A being positive throughout 2007, and over 60% of carcases postive for most suppliers in most quarters (Figure 4).

The average number of *Campylobacter* cfus on positive carcases started high (over 10^4 cfu/carcase) in supplier A, but has declined steadily over the study period, whereas the levels in suppliers B and C have remained at a relatively constant, lower level throughout the study period (Figure 5).

4.3.1.2 Red meat and offal Liver samples were more likely to be positive for *C. jejuni* than any of the red-meat mince samples. The highest prevalence was observed in lamb's liver (38%) followed by beef liver (10%). Of the mince samples, beef, pork and lamb mince were positive on only 5%, and 2% and 1% of occasions respectively (Table 3).

4.3.1.3 Sheep and cattle faeces Over 70% of pooled faecal samples from cattle and sheep were positive for presumptive *Campylobacter* spp. The proportion PCR positive for *C. jejuni* was lower (55% of samples tested).

4.3.1.4 Environmental water The proportion of presumptive *Campylobacter* positive samples varied from 23.6% (13/55) at Tokomaru to 54% at Albert St Palmerston North. Overall 42% of samples have been culture positive and 32% were confirmed as *C. jejuni* positive (Table 3).

4.3.1.5 Wild bird A total of 24/192 samples (12.5%) were positive for *C. jejuni*; 4/60 dried samples (6.7%) and 20/132 fresh samples (15.2%).

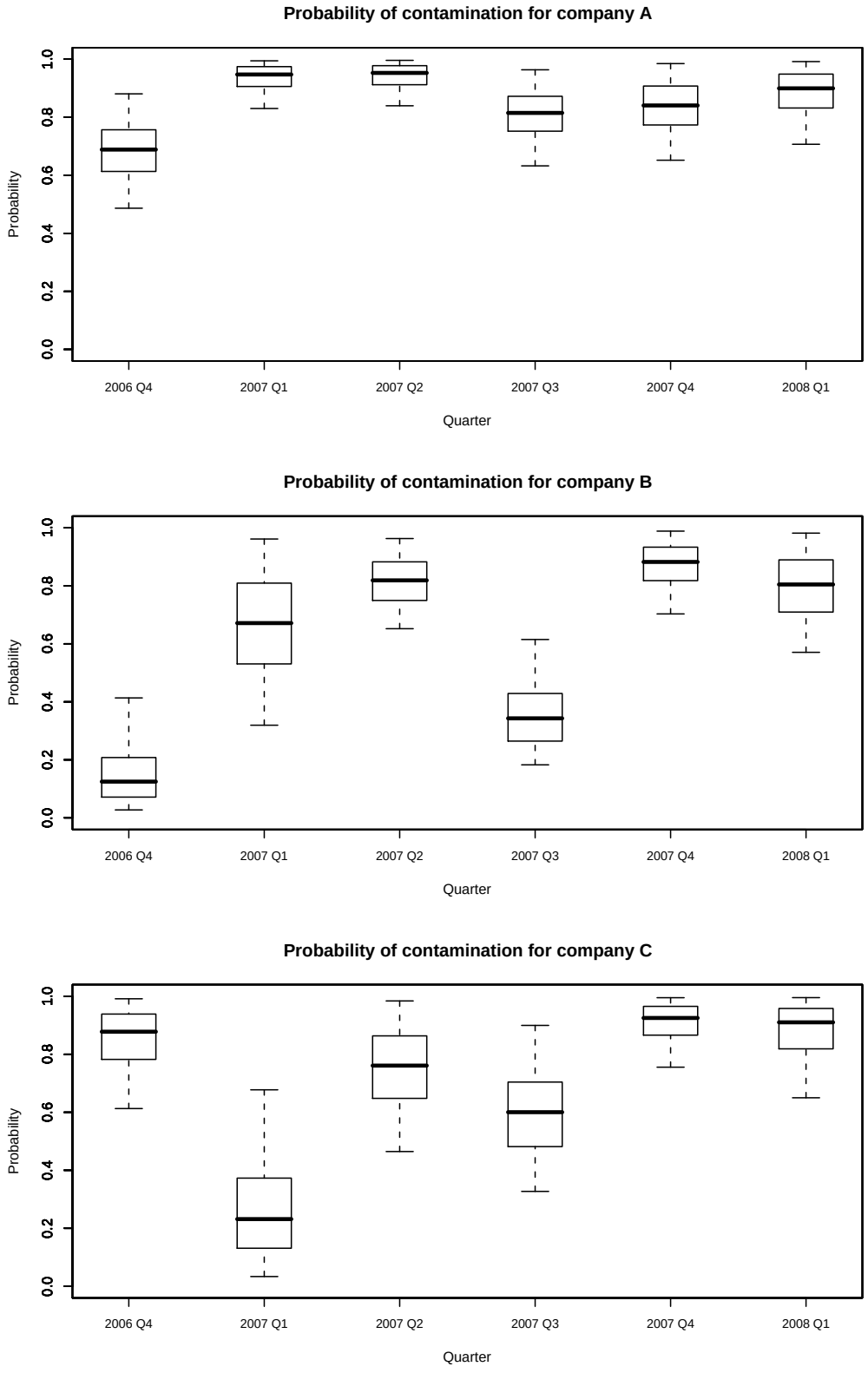

Figure 4: *Campylobacter* on poultry carcases by quarter: probability of contamination for each supplier

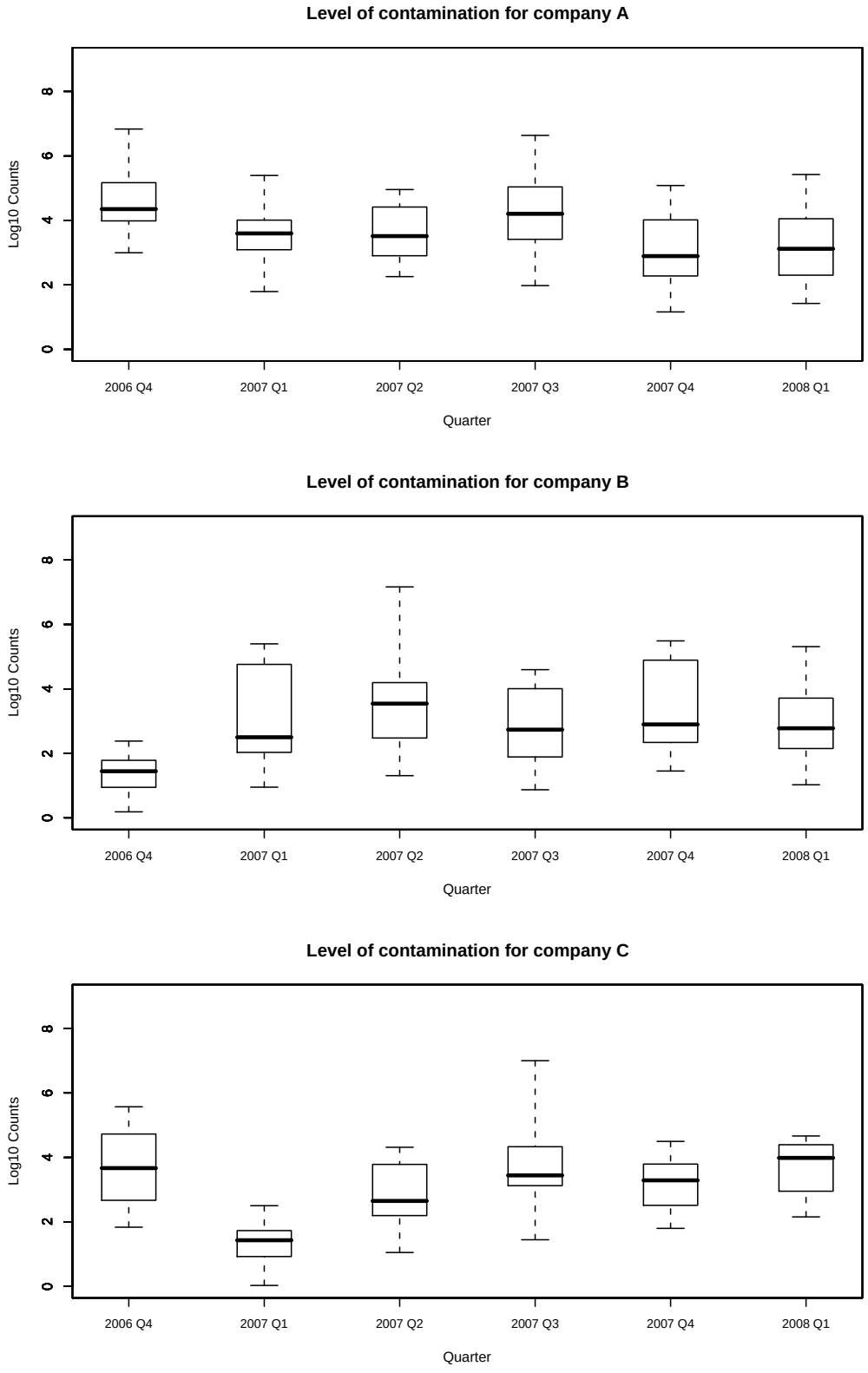

Figure 5: *Campylobacter* on poultry carcases by quarter: level of contamination on positive carcases for each supplier

4.3.2 Distribution of MLST genotypes

4.3.2.1 Poultry Some sequence types were common to all poultry suppliers (e.g. ST-45) whereas others were found predominantly, or uniquely in individual suppliers (Appendix E, Table 12.

4.3.2.2 Cattle and sheep sources Several STs were strongly associated with ruminants, including ST-42, ST-61, ST-422 and ST-2026. These were also identified in human cases. See (Appendix E, Table 13)

4.3.2.3 Environmental water Although some ruminant-associated genotypes were isolated from environmental water, the majority belonged to genotypes associated with wildlife. The most common was ST-2381 - this genotype has not been identifed anywhere other than in New Zealand water. See (Appendix E, Table 14)

4.3.2.4 Wild birds The most common genotype isolated from wild birds was ST-45. This ST, along with two other members of the ST-45 complex (ST-137 and ST-583) were the only wild bird genotypes associated with human infection. See (Appendix E, Table 14)

4.3.2.5 Comparing ST-474 and ST-45 using *fla*A sequencing While the seven housekeeping gene MLST scheme is adequate for characterising many host- and source-associated strains, there is increasing evidence that certain sequence types, such as the founder strain for clonal complex ST-45, are quite diverse. Here we explore the diversity of selected strains archived over the last 3 years from multiple sources in the Manawatu. We see if further insight into source attribution can be gained from finer-resolution genotyping through the addition of a more variable gene, the *fla*A Short Variable Region (SVR), to the current scheme. We focus on ST-45 and ST-474, two of the most common sequence types in New Zealand. A total of 69 isolates from a range of sources were typed using *fla*A sequencing – all were either ST-45 or ST-474. Eleven *fla*A nucleotide alleles and 8 unique peptide alleles were identified among the 49 ST-45 isolates (Table 4). The commonest allele was nucleotide allele *fla*A-21, which was found in all poultry suppliers and a sheep isolate, but not in humans. All ten of the ST-45 isolates from supplier B were *fla*A-21. Human isolates grouped with isolates from wild birds, poultry suppliers A and D and sheep. The largest group of human isolates, *fla*A-2, grouped with an isolate from supplier A and an isolate from sheep (Figure 6).

The twenty ST-474 isolates grouped into four nucleotide and 3 peptide alleles. The human isolates belonged only to *fla*A-14, a group which contained all isolates from supplier A and single isolates from environmental water and pork (Table 4 and Figure 6).

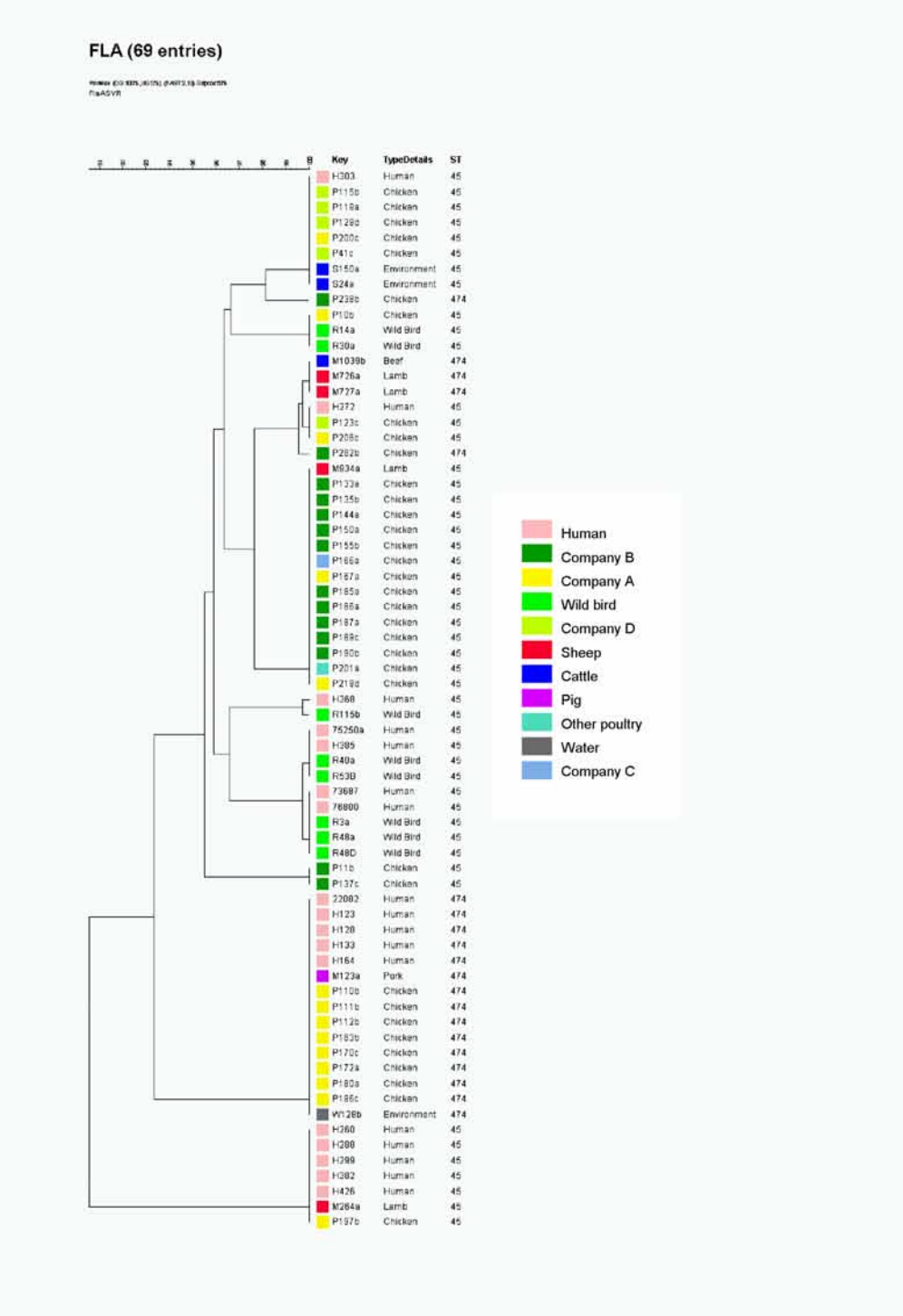

Figure 6: Dendrogram showing the relationship between isolates based on the *fla*A gene. Isolates were from a range of sources, belonging to ST-45 and ST-474

Table 4: Distribution of *fla*A alleles amongst ST-45 and ST-474 isolates

Sequence type	*fla*A Nucl. Allele	Peptide Allele	Total	Sources
45	8	1	3	Human (1), Supp A(1), Supp D(1)
45	22	1	8	Human (1), Supp A (1), Supp D (4), Sheep (2)
45	21	2	15	Supp A (2), Supp B (10), Supp C (1), Other (1), Sheep (1)
45	5	5	1	Human (1)
45	15	5	4	Human (2), Wild bird (2)
45	70	5	5	Human (2), Wild bird (3)
45	264	15	2	Wild bird (2)
45	284	15	1	Supp A (1)
45	18	20	2	Supp B (2)
45	2	27	7	Human (5), Supp A(1), Sheep (1)
45	5*	partial	1	Wild bird (1)
474	49	1	1	Supp B (1)
474	516*	1	3	Sheep (2), Cattle (1)
474	126	5	1	Supp B (1)
474	14	11	15	Human (5), Supp A (8), Water (1), Pork (1)

4.4 Distribution of MLST genotypes in Auckland and Christchurch studies

In Auckland, 125 isolates of *C. jejuni* were isolated for genotyping by MLST; 60 isolates from retail chicken and 65 isolates from human campylobacteriosis cases. ST-474 (41.5%), ST-354 (13.8%) and ST-50 (13.8%) were the most common STs found in human cases in Auckland. Figure 7 shows the prominence of these sequence types represented as a Minimum Spanning Tree, whereby the size of each pie chart represents the total number of isolates, and the charts are arranged to repesent the relatedness of each genotype.

In Christchurch in 2007, 99 isolates of *C. jejuni* were genotyped by MLST; 57 from retail chickens and 42 from humans. Eighteen STs were identified of which ST-50 was the most prominent accounting for 34.9% of human STs during the eight week study period. ST-474 (14% of cases) and ST-45 (11.6% of cases)were the other prominent STs in Christchurch found in humans (Figure 9).

In the Christchurch study conducted in 2008, ST-45 (15% of human cases) and ST-53 (10% of cases and 8.2% from chicken) were the two most common human STs (Figure 8).

The Minimum Spanning Trees (MSTs) for each region show a similar distribution of both human and poultry isolates in Auckland and Christchurch, to that of the Manawatu. The Manawatu MST includes isolates from non-poultry sources and clearly shows the presence of ruminant associated clusters (e.g. ST-61, ST-422, ST-2026 and ST-42). The unique environmental water cluster is also clearly visible (ST-2381)

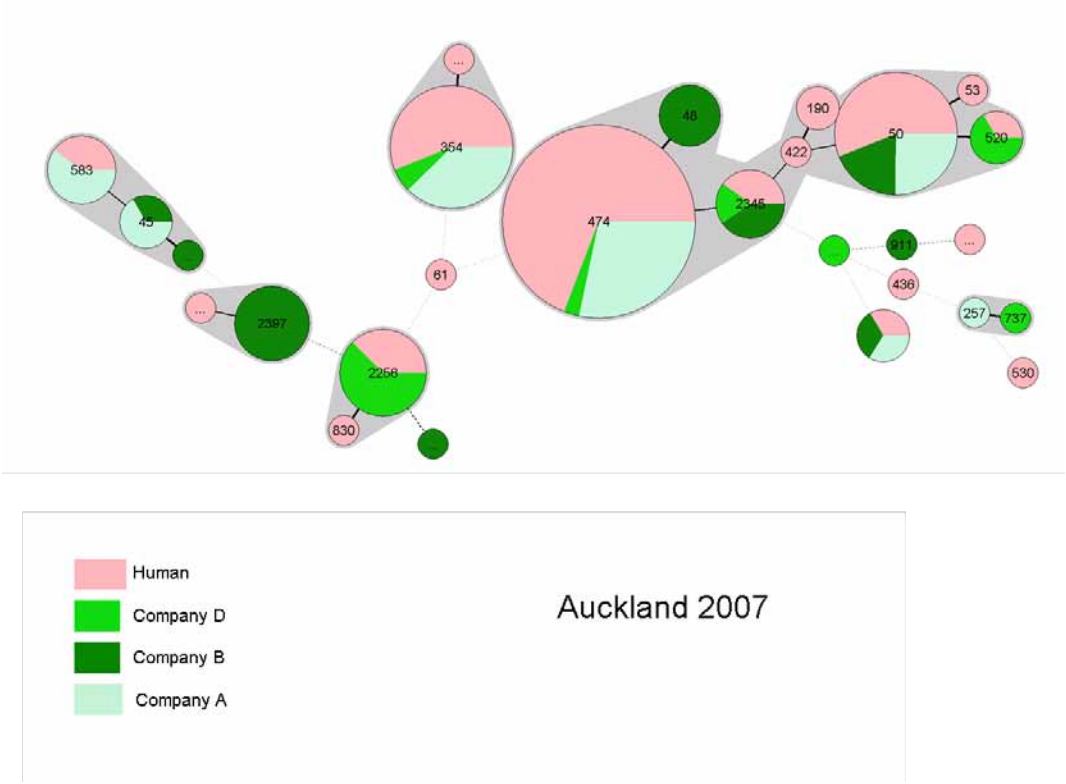

Figure 7: Minimum spanning tree showing the interrelationships between isolates from poultry and human cases in Auckland. The size of the circle represents the number of isolates. Note the association between Supplier A poultry and the common human STs 474 and 354.

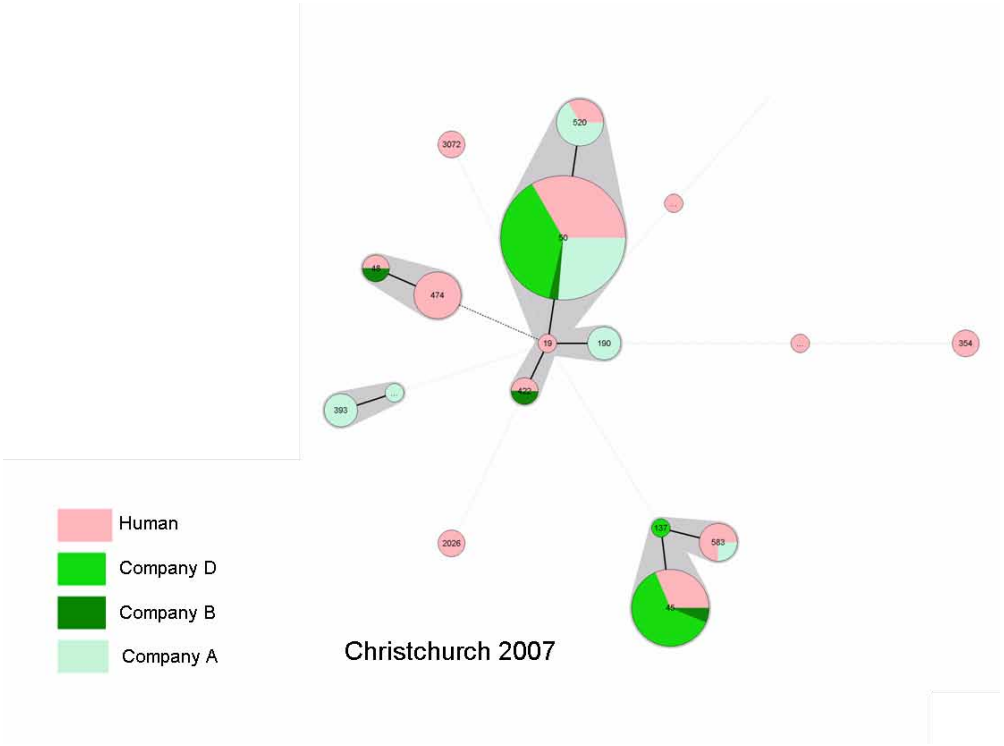

Figure 8: Minimum spanning tree showing the interrelationships between isolates from poultry and human cases in Christchurch in 2007. The size of the circle represents the number of isolates. Note the lower diversity among isolates in this region compared to Auckland and Manawatu and the dominance of STs 50 and 45.

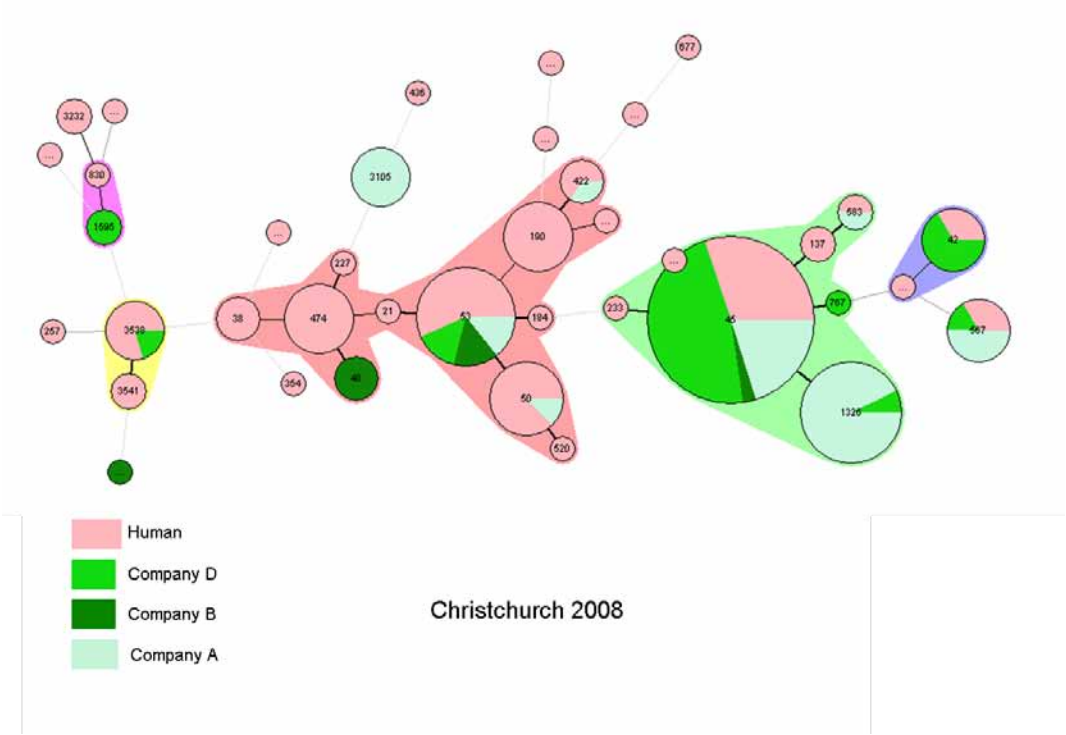

Figure 9: Minimum spanning tree showing the interrelationships between isolates from poultry and human cases in Christchurch in 2008. The size of the circle represents the number of isolates. Note the greater diversity among isolates in this region in 2008 compared to 2007.

4.5 Source attribution modelling

4.5.1 Proportional similarity index

Although this technique does not provide a direct estimate of the number or proportion of human cases attributable to each source, it does give a measure of the degree of similarity between the STs identified with each source and the human cases. Table 5 shows a strong similarity between the STs found in poultry and the human cases (PS = 0.51) and the fact that the confidence intervals do not overlap with the next highest PS index (Cattle, PS = 0.35) suggest this value is significantly different from all other sources. The lowest PS indexes were for wild birds and water isolates.

Table 5: The Proportional Similarity index for each source compared to the distribution of human genotypes, with 95% bootstrapped confidence intervals. Higher values indicate a strong similarity between the STs identified in the source and the human cases

Comparison source	Proportional Similarity index	Lower 95% CI	Upper 95% CI
Poultry	0.51	0.45	0.55
Cattle	0.35	0.28	0.40
Sheep	0.30	0.24	0.34
Water	0.15	0.08	0.20
Wild bird	0.10	0.06	0.13

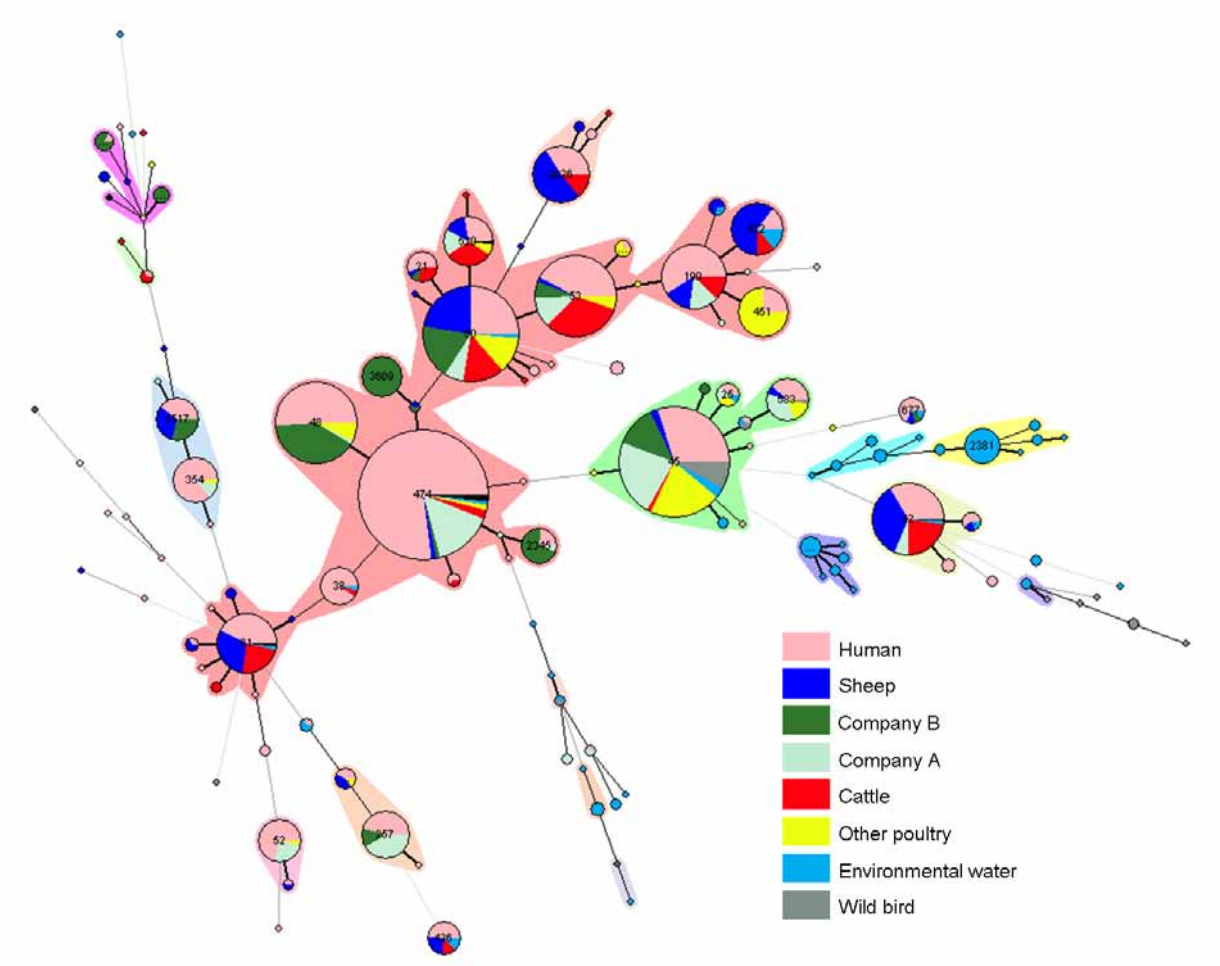

Figure 10: Minimum spanning tree for all isolates from human, food and environmental sources in the Manawatu between March 2005 and Feb 2008.

4.5.2 Dutch model

Table 6 shows the estimated proportion of cases attributable to each source using the Dutch model. Using this approach, only the 495 cases with a ST from an identified reservoir can be attributed. The highest proportion of cases is attributed to poultry (52%, 95% CI 49-64%) followed by cattle (17%, 95% CI 13-22%).

Table 6: Source attribution using the Dutch model: the proportion of human cases attributable to each source with 95% bootstrapped confidence intervals.

Source	Proportion of cases	Lower 95% CI	Upper 95% CI
Poultry	0.52	0.49	0.64
Cattle	0.17	0.13	0.22
Sheep	0.10	0.09	0.14
Wild bird	0.05	0.03	0.06
Water	0.11	0.04	0.18

4.5.3 Modified Hald model

Using the modified Hald model, the estimated contribution from each source is provided in Table 7. This model attributes all 495 cases in which a reservoir can be identified. Poultry was estimated to contribute 67% of cases. The 95% Bayesian credible intervals were wide for all sources using this model. Cattle were estimated to contribute more cases than sheep; 23% and 7% of cases respectively, whereas wild bird and environmental water contributed to only 1% of cases.

Table 7: Source attribution using the modified Hald model: the proportion of human cases attributable to each source with 95% Bayesian credible intervals.

Source	Proportion of cases	Lower 95% CrI	Upper 95% CrI
Poultry	0.667	0.402	0.833
Cattle	0.227	0.030	0.499
Sheep	0.075	0.004	0.223
Wild bird	0.010	0.003	0.059
Water	0.006	0.002	0.036

4.5.4 Island model

Table 8 shows the estimated source attribution using the Island model. This technique allowed all 521 cases to be attributed to a source. Again, poultry was identified as the most important source of human infection, accounting for 75% of human cases (95% credible intervals 65 to 85%). The next highest contributor to the burden of human cases was ruminants, with cattle estimated to contribute to 17% of cases (95% CrI 9-25%).

Figure 11 is a matrix plot showing the assignment probabilities for each of the 521 cases. Each isolate is represented by a column which is coloured according to the frequency with which each ST has been identified in each source. It is evident from this diagram that there is a block of human cases to the left of the diagram that are highly likely to have come from cattle or

Table 8: Source attribution using the Island model: the proportion of human cases attributable to each source with 95% Bayesian credible intervals

Source	Proportion of cases	Lower 95% CrI	Upper 95% CrI
Poultry	0.749	0.644	0.854
Cattle	0.172	0.088	0.249
Sheep	0.044	0.003	0.114
Wild bird	0.02	0.002	0.046
Water	0.004	0.000	0.028

sheep sources (i.e. a mixture of red and blue bars). In contrast, there is a block of cases that are highly likely to have come from poultry sources (yellow). Environmental water (pale blue) and wild bird (green) isolates form a small block of largely unique STs.

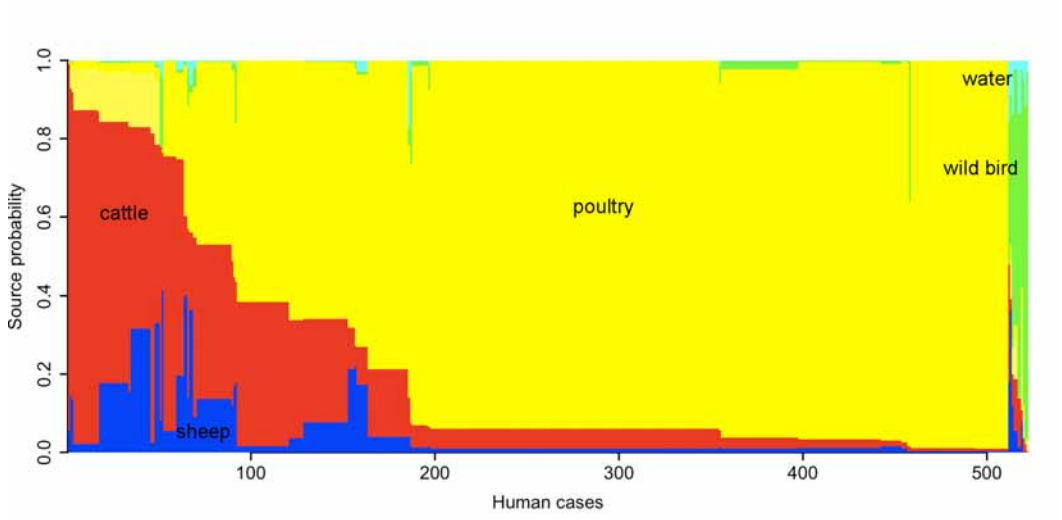

Figure 11: Output from Island model represented as a matrix plot. Each human case is a vertical column coloured according to the probability it came from each source

A comparison of the output from all three models, the Dutch, Hald and Island models, is shown in Figure 12.

4.5.5 Changes in source attribution over time

Source attribution estimates were calculated for each year of the study using the Dutch and modified Hald models (Table 9). These show a small decline in the proportion of cases attributable to poultry in year 3, and an increase in the proportion attributable to ruminants – although these estimates should be treated with caution due to the wide confidence intervals.

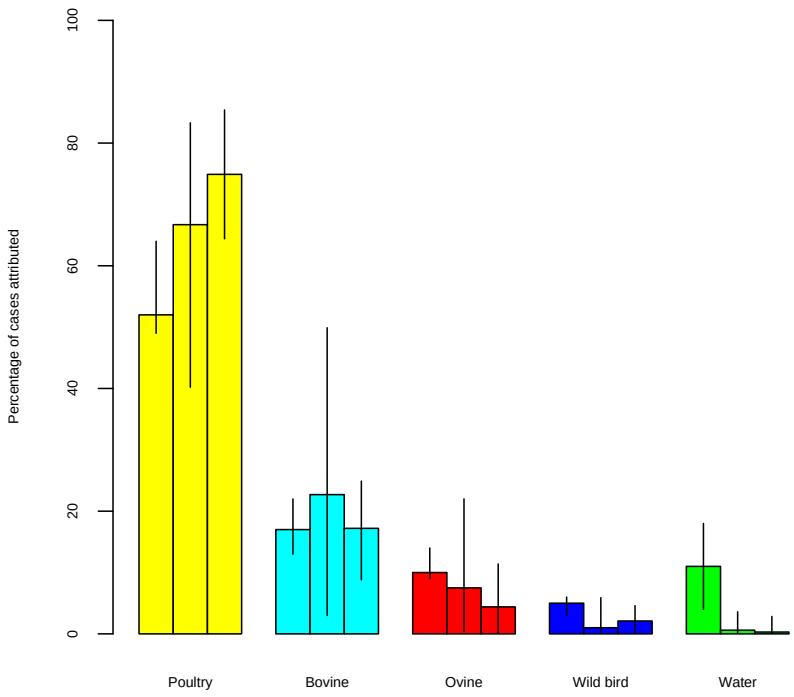

Figure 12: Proportion of human cases attributable to each source: comparing, from left to right, the Dutch, modified Hald and Island models. Error bars represent 95% confidence/credible intervals

Table 9: Source attribution estimates with upper and lower 95% confidence/credible intervals in year 1 (March 2005-Feb 2006), year 2 (March 2006-Feb 2007), and year 3 of the study (March 2007-Feb 2008) using the Dutch and modified Hald model

	Dutch model								
	Year 1			Year 2			Year 3		
	Propn	Lower CI	Upper CI	**Propn**	Lower CI	Upper CI	**Propn**	Lower CI	Upper CI
Poultry	**0.59**	0.52	0.66	**0.64**	0.58	0.72	**0.57**	0.51	0.65
Cow	**0.22**	0.16	0.27	**0.17**	0.11	0.22	**0.19**	0.12	0.23
Sheep	**0.12**	0.09	0.17	**0.09**	0.06	0.13	**0.14**	0.10	0.19
Wild bird	**0.03**	0.01	0.05	**0.05**	0.03	0.07	**0.05**	0.02	0.08
Water	**0.04**	0.01	0.07	**0.05**	0.01	0.09	**0.06**	0.02	0.09
	modified Hald model								
	Propn	Lower CrI	Upper CrI	**Propn**	Lower CrI	Upper CrI	**Propn**	Lower CrI	Upper CrI
Poultry	**0.73**	0.48	0.89	**0.80**	0.57	0.93	**0.69**	0.47	0.86
Cow	**0.15**	0.01	0.42	**0.10**	0.00	0.32	**0.09**	0.00	0.34
Sheep	**0.08**	0.00	0.24	**0.05**	0.00	0.19	**0.16**	0.02	0.35
Wild bird	**0.01**	0.00	0.06	**0.01**	0.00	0.08	**0.01**	0.00	0.09
Water	**0.01**	0.00	0.05	**0.01**	0.00	0.05	**0.01**	0.00	0.07

4.6 The epidemiology of human cases

EpiSurv information was successfully matched to 692 of the 773 primary samples received (89.5%). This has enabled the relative frequency of different genotypes to be examined for different human sub-populations. As examples we describe the relationship between ruminant- and poultry associated genotypes and age and occupation. For this we consider ruminant genotypes to be ST-61, ST-42 and members of the ST-403 complex (ST 403 and ST-2026). These are ruminant associated both in our study and overseas studies [8, 13, 14]. Using similar criteria, poultry associated genotypes are ST-257, ST-354, ST-45, ST-474 and ST-48. Further, using Markov random field modelling of relative risk surfaces we compared the geographical distribution of ruminant and poultry associated cases.

4.6.1 Completeness of epidemiological data collection

Table 10 shows the completion rates for a selection of fields both before and after the enhanced surveillance programme was initiated. Only confirmed cases were followed up by the MidCentral PHS (i.e. within the MidCentral DHB). For most fields there was a marked improvement in completion after July 2006. A significant improvement in contact rate was noted during the trial when MCPHS Health Protection Officers (HPO) adopted the use of a late afternoon and early evening shift to attempt to contact and interview cases.

Table 10: The completeness of EpiSurv data fields before and after the MidCentral Public Health Services (MCPHS) enhanced surveillance project was initiated in July 2006

Data fields	1/07/06 to 1/03/08	1/02/05 to 30/06/06
Report date*	100.0%	100.0%
Sex*	100.0%	99.0%
Age*	100.0%	100.0%
Hospitalisation*	92.8%	68.4%
Ethnicity*	94.3%	70.6%
Linked to an outbreak	98.0%	99.0%
Onset date*	92.6%	66.0%
Occupation*	94.7%	84.0%
Contact with another case	91.4%	63.5%
Fits clinical description*	95.1%	70.1%
Contact Farm Animals	94.0%	67.2%
Sick Animals	91.0%	63.3%
Drunk Untreated H2O	81.0%	52.9%
Food Premises	88.0%	56.6%
Other Symptomatic People	93.0%	59.0%
Recreational Water contact	95.5%	67.0%
Overseas during incubation	93.0%	68.2%
Linking Lab number in comments section	97.0%	0.0%
Cases contacted phone/visit/questionnaire	95.3%	69.1%

* core Episurv fields

Between 1st of November 2006 and the 1st of March 2008 HPO's attempted to include the additional questions relating to meat eaten and the consumption of unpasturised milk during the case's incubation period. In 416 interviews only 13 cases were not contacted resulting in a contact rate of 96% and high completion rates (98.0% for unpasteurised milk consumption, and

99.8% for chicken consumption during the incubation period).

One of the limitations of the Episurv database identified during the data collection phase related to the "unknown" option. As an example: an HPO during an interview can chose between "yes" "no" or "unknown". However, "unknown" could mean either the case didn't know the answer to the question, or the HPO failed to ask the question: this may have resulted in a lower completion percentage for more subjective questions.

4.6.2 Temporal trends in human case data

Figure 13 shows the smoothed number of cases per week in the Manawatu attributable to the poultry associated ST-474, other poultry associated STs and ruminant associated STs. There is clear evidence of seasonality in the poultry associated cases, and the winter poultry-associated epidemic of 2006 (May - July 2006) is clearly visible. A recent decline in the weekly incidence of ST-474 associated cases is evident.

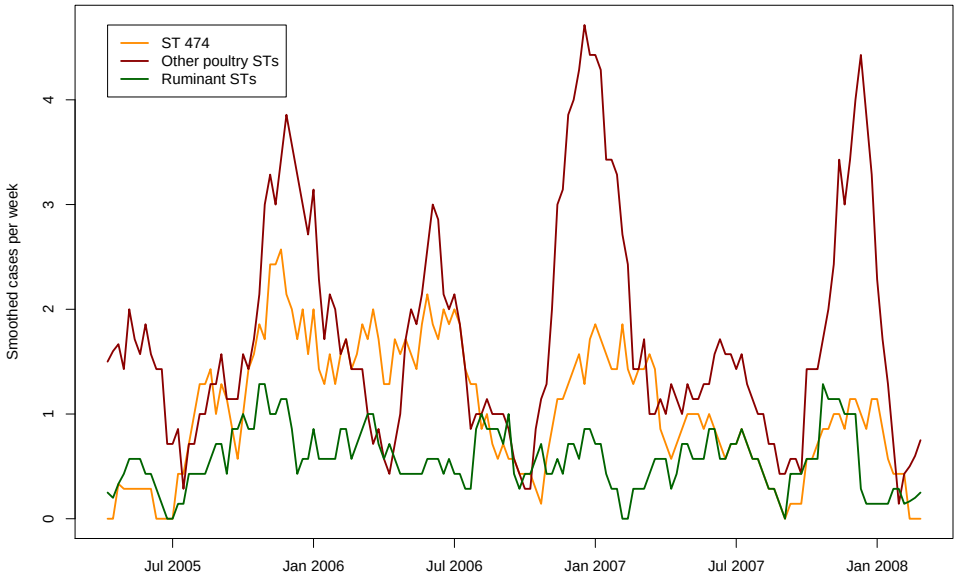

Figure 13: Temporal trends in human cases attributable to poultry and ruminants in the Manawatu between March 2005 and Feb 2008.

4.6.3 Overseas travel cases

Only ten of the 393 human cases who provided information on foreign travel reported being overseas during the incubation period. Of these, six (60%) cultured sequence types with no known animal reservoir in New Zealand, and three were unique to this study and were therefore new to the PubMLST database. This was significantly different from non-overseas travellers, for which only 17 (4.4%) were sequence types not recovered from animals in New Zealand.

4.6.4 Case-case comparison of poultry and ruminant-associated genotypes

4.6.4.1 Age distribution The relationship between age and ruminant versus poultry genotype is shown in Figure 14. The difference in age-profiles is significant (χ^2 16.6, 7d.f, P=0.02). Ruminant genotypes are relatively more common in children under 10 years of age compared to poultry genotypes.

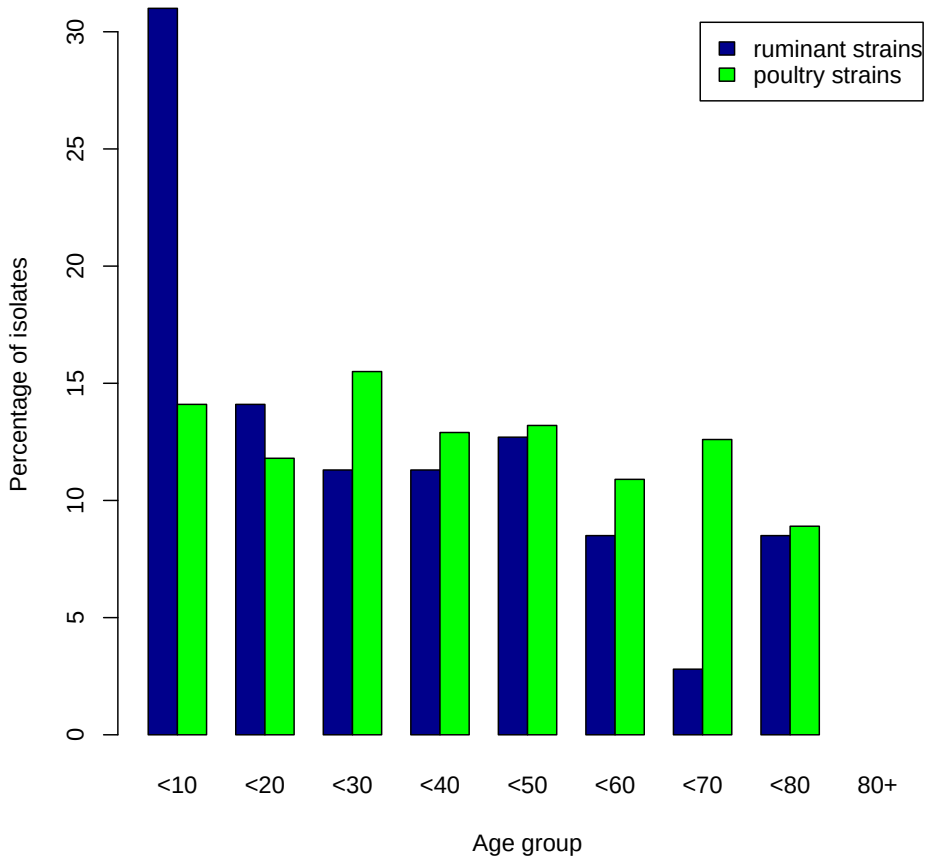

Figure 14: The proportion of cases associated with ruminant (blue bars, N=71) and poultry (green bars, N=348) genotypes in each age category.

4.6.4.2 Occupational exposures A total of 190 occupations were listed for the 480 EpiSurv linked human cases. However for analysis, these were grouped into a number of smaller categories, with a related to their likely exposure to ruminants. For example the Cattle/sheep worker group included occupations stated as: Dairy cattle farm worker, Dairy cattle Farmer, Dairy farm hand, Dairy farm worker, Dairy Farmer, Beef cattle farmer, Farm worker, Farmer, Farmer and Farm Manager, Freezing worker, Meat Inspector, Meat Process Worker, Meat Worker, Sheep weigher, Slaughterman/processor, Trainee Meat Inspector. We considered children and retired individuals as a separate group, and all other occupations as a single group.

Ruminant genotypes were relatively more common in children under 10 and adults with farm animal exposure (χ^2 4d.f. P<0.000001) compared to poultry genotypes (Table 11). For example 21.1% of human cases identified as ruminant genotypes were in individuals with an occupation involving contact with ruminant faeces, compared to only 4.9% of cases identified as poultry genotype.

Table 11: The distribution of human cases attributable to ruminant and poultry genotypes by occupation

	Number of cases		% row total		% column total	
Occupational group	Cattle genotypes	Poultry genotypes	Cattle genotypes	Poultry genotypes	Cattle genotypes	Poultry genotypes
Cattle/sheep worker	15	17	46.9	53.1	21.1	4.9
Other occupation	15	160	8.6	91.4	21.1	46.0
Child	28	78	26.4	73.6	39.4	22.4
Retired	7	51	12.1	87.9	9.9	14.7
Unknown	6	42	12.5	87.5	8.5	12.1

4.6.5 Spatial distribution of poultry and ruminant-associated human cases

Figures 15, 16 and 17 show the relative risk surfaces in the Manawatu for poultry associated genotypes ST-474, all other poultry associated genotypes and ruminant genotypes. Poultry associated cases are more prevalent in urban areas - ST-474 is is particularly prevalent in Palmerston North (Figures 15 and 16). In contrast, ruminant-associated cases are relatively more prevalent in rural areas (Figure 17).

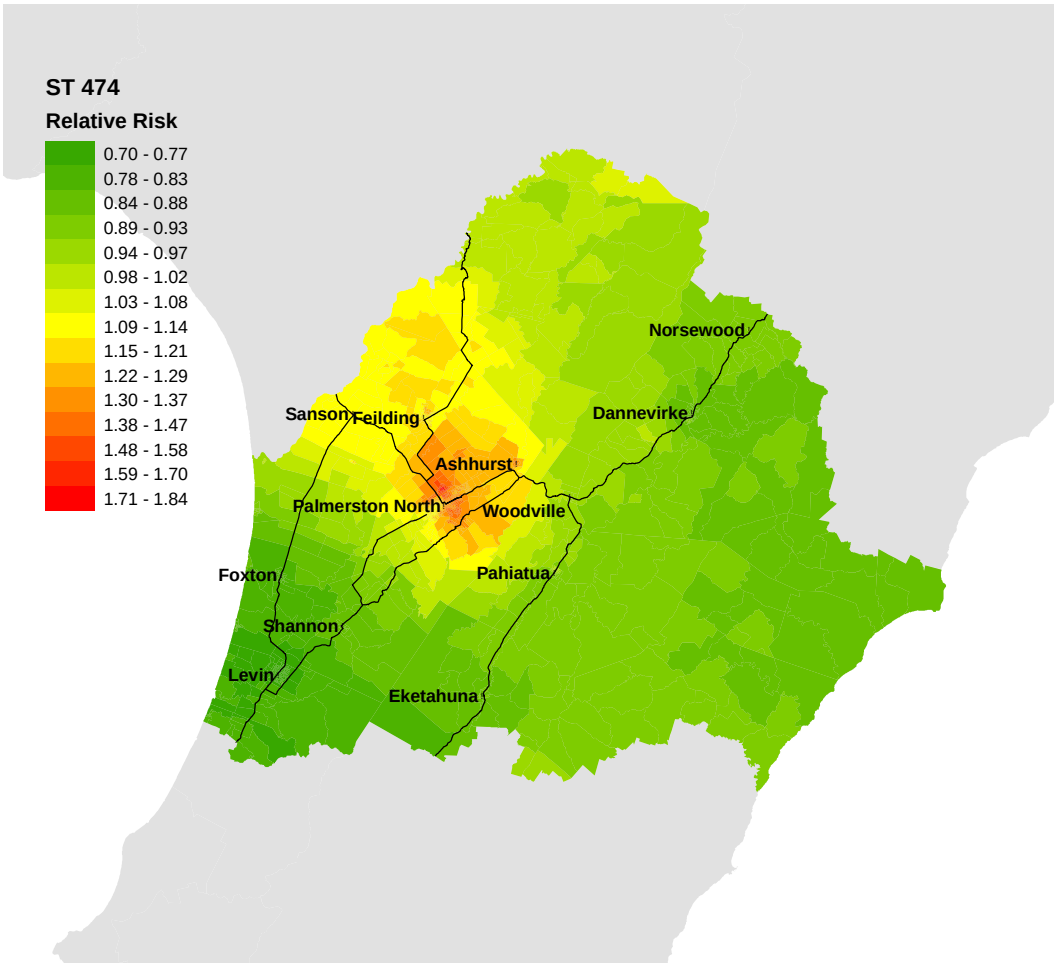

Figure 15: Relative risk surface of human cases of ST-474

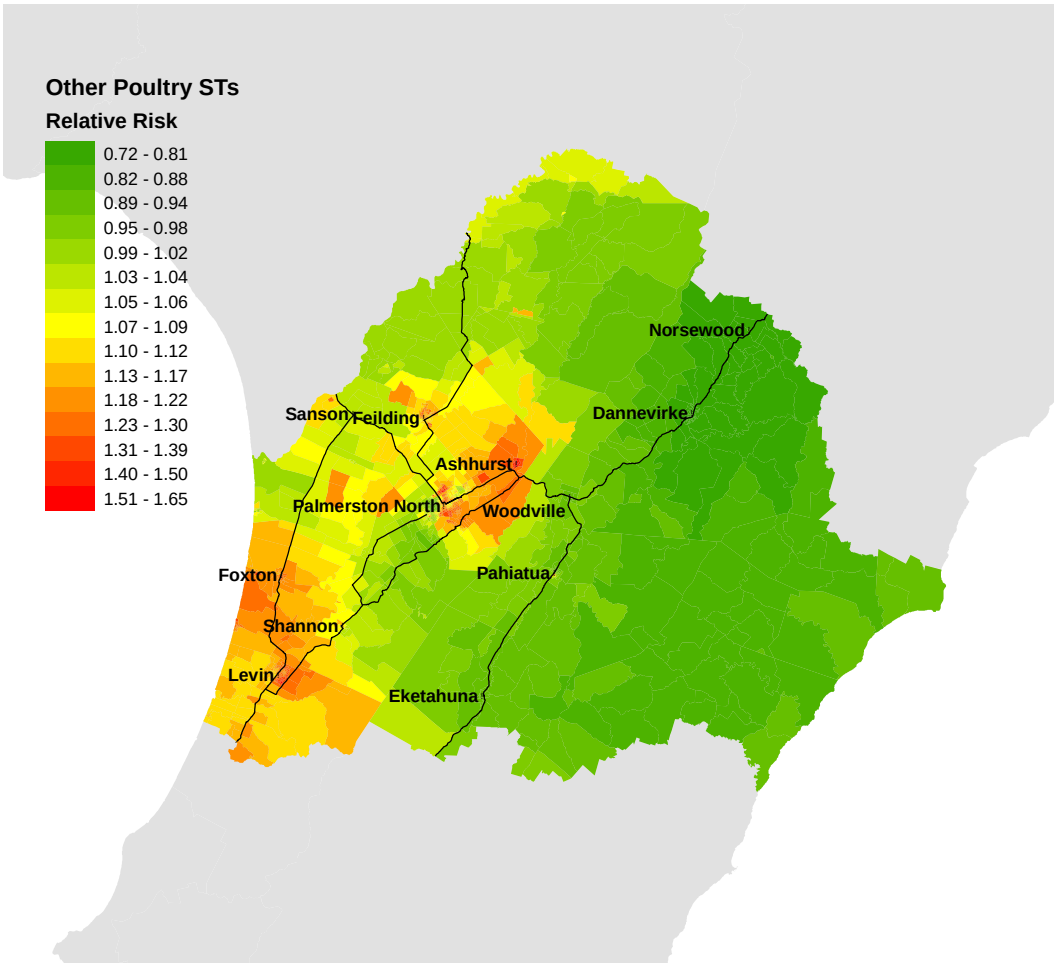

Figure 16: Relative risk surface for other poultry STs

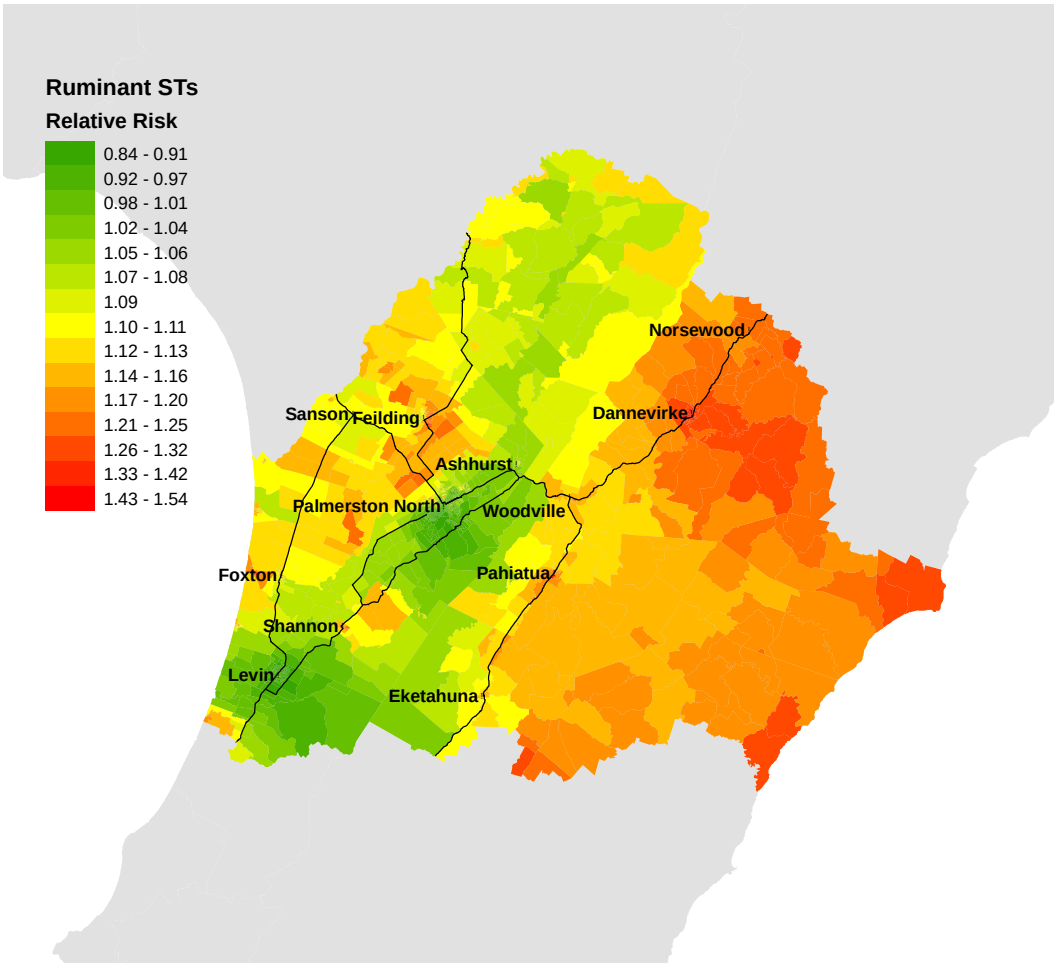

Figure 17: Relative risk surface for ruminant STs

5 Discussion

The purpose of this study was two fold: Firstly the development of the attribution techniques using *Campylobacter jejuni* as the prototype, and secondly robust estimation of source attribution proportions, that will inform the *Campylobacter* risk management strategy. We have shown that four published source attribution techniques [12, 19, 9, 6] can, with some modifications, be applied to *C. jejuni* MLST data to estimate the contribution of animal reservoirs to the burden of human infection, and inform public health policy. Further, by improving the quality of epidemiological information gathered on each case, we can gain further insight into the relative roles of food and environmental sources of human infection. Given appropriate information on other pathogens, some of which is already gathered on a routine basis, these techniques have the potential to be applied to inform decision making for the control of other zoonoses, such as salmonellosis and cryptosporidiosis. Further, these tools can be further modified to enable continuous, dynamic assessment of source attribution, enabling seasonal variation in infection sources to be identified, and an assesment of the effectiveness of control programmes.

5.1 Development of source attribution tools

Rather than focus on a single approach, we have modified four existing techniques and applied each technique to the same dataset. As each is based on different underlying assumptions this enables us to compare and contrast the output from each model. Although we have no gold standard, such an approach tells us whether the models are providing either reasonably consistent, or wildly different, estimates. In this study all models provided similar estimates of the contribution of each source to the burden of human infection, and this increases our confidence in the attribution estimates, and the inferences that can be drawn. It is important that consistency is not confused with accuracy - all models could be telling the same story, but be equally wrong - however, the results are consistent with the findings of earlier studies conducted in New Zealand [7].

It is worth considering the different assumptions underlying each model, as this may explain any observed differences in attribution estimates. The proportional similarity index is the simplest to calculate, and possibly the easiest to interpret: it is simply the area of intersection of each pair of genotype probability distributions. If the distribution of genotypes in a particular source is very similar to the distribution in human cases, it will have a PS index close to 1.0, if none of the genotypes found in the source appear in humans, the PS index will be zero. The technique doesn't estimate the number or proportion of human cases attributable to each source, and therefore doesn't make any of the assumptions underlying the three model-based techniques (such as variation in virulence and the nature of the food or environmental source of infection that are discussed below). We have adapted the techique to provide both bootstrapped confidence intervals and a method of making formal comparisons of PS indices. The advantages of this technique are therefore 1) its simplicity and ease of calculation, and 2) its ability to assess formally changes in the PS index for single sources. We therefore recommend this is a useful first-step in the assessment of source attribution. The major disadvantage is that PS index doesn't provide an estimate of the number of cases attributable to each source, simply the correlation between source distributions.

The Dutch model does provide a direct estimate of the number of cases attributable to each source and, like the PS index, can be calculated quickly and easily using a simple spreadsheet. We have adapted the techniqe to provide some estimate of uncertainty, but this is restricted to sampling uncertainty and does not include parameter uncertainty. Further, the technique does not consider variation in the ability of different strains to cause human disease, and does not take account of differences between food sources. In other words, all strains are considered to have an equal probability of causing human disease, at a given dose in a given food, and all sources contaminated with *C. jejuni* are assumed to have an equal probability of causing disease in humans. It is highly likely that some STs are more virulent than others, and therefore more likely to cause human disease at a given dose. Further, some food sources may be considered to be more 'risky' than others due to variation in, for example, the level of contamination, properties of the food matrix and the effects of cooking on thermal inactivation (some foods are more likely to be undercooked than others).

The modified Hald model does take into account both strain-type and food-source factors, and incorprates both sampling and parameter uncertainty using Bayesian inference. In our modified approach, we extend the model described by Hald et al [12] to include uncertainty in the source prevalence matrix. In order to reduce the number of parameters in the model we have made further assumptions concerning the distribution of bacteria-dependent factors. The advantages of this approach are therefore the direct estimates of the number of cases attributable to each source, the comprehensive representation of uncertainty in these estimates, and the explicit consideration of both bacteria and source-dependent factors. A disadvantage of this method, that is shared by the Dutch method, is that only human cases with an ST that has been identified in a source or animal reservoir can be attributed. This is because neither method takes into consideration the relatedness of STs to each other, and therefore cannot draw any inference based shared host-associations arising from genotypic similarity. This disadvantage is overcome by taking a population genetics approach such as the Island model.

The Island model can assign all human isolates, regardless of whether thay have been identified in a reservoir, to a probable source. In the Manawatu approximately 5% of human cases have STs that have not been identifed in a reservoir - using the Island model these can now be assigned. Like the Dutch and modified Hald models, the Island model provides estimates of the number of cases attributable to each source, with a measure of uncertainty that, in common with the modified Hald model, captures both sample and parameter uncertainty using Bayesian inference. The credible intervals for the Island model are narrower than for the modified Hald model, most likely because more information (allelic profile) is used to fit the models.

We conclude that all models must be interpreted after taking into consideration the underlying assumptions, and that it is preferable to fit more than one model, calculate the PS index - and then compare the results.

5.2 Source attribution estimates

All three models lead to similar estimates, and these are consistent with the PS index. All methods suggest that the majority of cases over the three-year period could be attributed to poultry, providing further evidence of the importance of the contribution of this food source to the burden of campylobacteriosis in New Zealand [7]. The next most important source was

cattle, followed by sheep, with relatively minor contributions from wild birds and water. These findings are remarkably similar to the conclusions of a similar study conducted in Lancashire, UK [6], the major difference being the relatively higher contribution from cattle (35%). In the UK study 56.5% were attributed to poultry, 4.3% from sheep, 1.7% from wild birds and 0.9% from environmental water. Additional information from the comparative studies conducted in Auckland and Christchurch show a similar distribution of genotypes in both humans and poultry, suggesting the source attribution in the Manawatu is likely to be similar to that in other regions of New Zealand.

The estimated contribution from poultry was consistently the highest for each year of the study, although we could only obtain yearly estimates using the Dutch and modified Hald models. This work could be extended to develop the methodology for 'dynamic source attribution' in continuous time, using multiple models. The consistently high proportion of cases attributable to poultry provides further support for targetting this food source as part of the *Campylobacter* risk management strategy. Although there is some suggestion of a decline in the proportion of cases attributable to poultry in year 3, it is insufficient evidence to assess the likely effectiveness of recent measures to control *Campylobacter* in the food chain. More conclusive information could be provided by sequence typing and dynamic modelling of a longer time series by extending the study to include the 12 months following the observed decline in the levels of contamination on poultry, and the introduction of mandatory standard and performance targets.

*fla*A sequence typing provided additional insight of the relative contribution of more specific sources to human infection, and could be a valuable extension to the existing MLST scheme. The nested *fla*A study showed that the common genotypes ST-474 and ST-45 could be further discriminated, and the sub-genotypes were associated with particular sources. The addition of this gene could further reduce the uncertainty in source attribution estimates, and be of more value in identifying precise sources in localised outbreaks.

5.3 Further insight gained from epidemiology

The addition of epidemiological data has provided further unique insight into the transmission pathways in New Zealand. We found that poultry associated cases were largely confined to urban dwellers, whereas the ruminant associated cases were predominantly in rural dwellers, and in children and adults with an occupation that is likely to bring them into contact with ruminant faeces. This epidemiological information, combined with the relative contamination levels on food products, suggests that poultry cases are likely to be acquired from food, whereas ruminant-associated cases are more likley to result from direct contact with animal faeces. If the poultry control strategy is successful, this will present new challenges for the control of ruminant-acquired infections and will require a different approach to risk management.

6 Concluding comments

Combining advanced genotyping with multiple model-based source attribution tools has provided reasonably consistent estimates of the relative contribution of different food and environmental sources to the burden of human campylobacteriosis in the Manawatu. Poultry are estimated to account for the majority of cases in this region. This suggests that, if the recent interventions in

the poultry industry substantially reduce human exposure, the poultry risk management strategy is likely to have a major impact on the human notification rates. The long-term effectiveness of the poultry risk management strategy could be tested by sequence typing and modelling the distribution of genotypes in humans, food and environment beyond February 2008. Ruminant sources, namely cattle and sheep, were also identified to be the source of a smaller number of cases, mainly in younger children in rural areas. Given the nature of this exposure, and the relatively low prevalence in food derived from these sources, these ruminant-derived cases were more likely to be the result of environmental rather than food-borne exposures. With further modification of each of the attribution models, this work could be developed to provide 'dynamic source attribution' in continuous time, rather than the three year discrete yearly intervals described in this study. Such an approach could be extended to other pathogens, providing a unique and valuable tool for the national surveillance of human infectious diseases.

7 Acknowledgements

This work was funded by the New Zealand Food Safety Authority and was done in collaboration with ESR Ltd, Kenepuru and MidCentral Public Health Services. We acknowledge the following individuals who contributed to the writing of this report: Dr Daniel Wilson (University of Chicago, formerly Lancaster University, UK), Dr Simon Spencer, Dr Anne Midwinter, Dr Julie Collins-Emerson, Petra Mullner and Tui Shadbolt. The work was carried out by the above and Rebecca Pattison, Rukhshana Akhter, Errol Kwan, Lynn Rogers, Isabel Li, Jim Learmonth, Anthony Pita, Sarah Vaughan, (Massey Molecular Epidemiology group, Hopkirk Institute), Dr Phil Carter, Ruth Pirie and Dr Tecklok Wong (ESR), Dr Grant Hotter (AgResearch). We also acknowledge the following individuals and organisations for their contribution to the study: Palmerston North Hospital, MidCentral PHS, MedLab Central, Graham McBride (NIWA), Poultry Industry, Professor Mike Hedley, Dr Geoff Jones, Dr Alasdair Nobel, Professor Martin Hazelton, Dr Michael Baker, Diane Richardson, Horizons Regional Council, and the farmers in the Manwatu who allowed us to sample cattle and sheep.

References

[1] M. B. Batz, M. P. Doyle, J. G. Morris, J. Painter, R. Singh, R. V. Tauxe, M. R. Taylor, and Dmal Wong. Attributing illness to food. *Emerging Infectious Diseases*, 11(7):993–999, Jul 2005.

[2] M. Denis, J. Refregier-Petton, M. J. Laisney, G. Ermel, and G. Salvat. *Campylobacter* contamination in french chicken production from farm to consumers. use of a pcr assay for detection and identification of *Campylobacter jejuni* and *C. coli*. *J Appl Microbiol*, 91(2):255–67, Aug 2001.

[3] X. Didelot and D. Falush. Inference of bacterial microevolution using multilocus sequence data. *Genetics*, 175:1251–1266, 2007.

[4] K. E. Dingle, F. M. Colles, R. Ure, J. A. Wagenaar, B. Duim, F. J. Bolton, A. J. Fox, D. R. Wareing, and M. C. Maiden. Molecular characterization of *Campylobacter jejuni* clones: a basis for epidemiologic investigation. *Emerg Infect Dis*, 8(9):949–55, Sep 2002.

[5] K. E. Dingle, F. M. Colles, D. R. Wareing, R. Ure, A. J. Fox, F. E. Bolton, H. J. Bootsma, R. J. Willems, R. Urwin, and M. C. Maiden. Multilocus sequence typing system for *Campylobacter jejuni*. *J Clin Microbiol*, 39(1):14–23, Jan 2001.

[6] Wilson D.J, Gabriel P, Leatherbarrow A.J, Cheesbrough J, Gee S, Bolton E, Fox A.J, Fearnhead P, Hart CA, and P. Diggle. Tracing the source of campylobacteriosis. *PLoS Genetics*, in press, 2008.

[7] J. Eberhart-Phillips, N. Walker, N. Garrett, D. Bell, D. Sinclair, W. Rainger, and M. Bates. Campylobacteriosis in new zealand: results of a case-control study. *Journal of Epidemiology and Community Health*, 51(6):686–691, Dec 1997.

[8] N. French, M. Barrigas, P. Brown, P. Ribiero, N. Williams, H. Leatherbarrow, R. Birtles, E. Bolton, P. Fearnhead, and A. Fox. Spatial epidemiology and natural population structure of *Campylobacter jejuni* colonizing a farmland ecosystem. *Environ Microbiol*, 7(8):1116–26, Aug 2005.

[9] N. Garrett, M. L. Devane, J. A. Hudson, C. Nicol, A. Ball, J. D. Klena, P. Scholes, M. G. Baker, B. J. Gilpin, and M. G. Savill. Statistical comparison of *Campylobacter jejuni* subtypes from human cases and environmental sources. *J Appl Microbiol*, 103(6):2113–21, Dec 2007.

[10] I. A. Gillespie, S. J. O'Brien, J. A. Frost, G. K. Adak, P. Horby, A. V. Svan, and et al. A case-case comparison of *Campylobacter coli* and *Campylobacter jejuni* infection: A tool for generating hypotheses. *Epidemiology and Infection*, 128(2):111–118, 2002.

[11] S. Greenland and K. J. Rothman. *Measure of Effect and Measures of Association*, pages 54–58. Lippincott Williams and Wilikins, 2nd edition, 1998.

[12] T. Hald, D. Vose, H. C. Wegener, and T. Koupeev. A bayesian approach to quantify the contribution of animal-food sources to human salmonellosis. *Risk Anal*, 24(1):255–69, Feb 2004.

[13] P. S. Kwan, M. Barrigas, F. J. Bolton, N. P. French, P. Gowland, R. Kemp, A. J. Leatherbarrow, M. Upton, and A. Fox. The molecular epidemiology of *Campylobacter jejuni* populations in dairy cattle, wildlife and the environment in a farmland area. *Appl Environ Microbiol*, in press, 2008.

[14] P. S. Kwan, A. Birtles, F. J. Bolton, N. P. French, S. E. Robinson, L. S. Newbold, M. Upton, and A. J. Fox. Longitudinal study of the molecular epidemiology of *Campylobacter jejuni* in cattle on dairy farms. *Appl Environ Microbiol*, 74(12):3626–33, Jun 2008.

[15] S. M. McTavish, C. E. Pope, C. Nicol, K. Sexton, N. French, and P. E. Carter. Wide geographical distribution of internationally rare campylobacter clones within new zealand. *Epidemiol Infect*, 136(9):1244–52, Sep 2008.

[16] W. G. Miller, B. M. Pearson, J. M. Wells, C. T. Parker, V. V. Kapitonov, and R. E. Mandrell. Diversity within the *Campylobacter jejuni* type i restriction-modification loci. *Microbiology*, 151(Pt 2):337–51, Feb 2005.

[17] J. K. Pritchard, M. Stephens, and P. Donnelly. Inference of population structure using multilocus genotype data. *Genetics*, 155(2):945–59, Jun 2000.

[18] L. W. Riley. *Molecular Epidemiology of Infectious Diseases - Principles and Practices*. ASM Press, 2004.

[19] O. Rosef, G. Kapperud, S. Lauwers, and B. Gondrosen. Serotyping of *Campylobacter jejuni*, *Campylobacter coli*, and *Campylobacter laridis* from domestic and wild animals. *Appl Environ Microbiol*, 49(6):1507–10, Jun 1985.

[20] N. A. Rosenberg and M. Nordborg. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nat Rev Genet*, 3(5):380–90, may 2002.

[21] U. Stucki, J. Frey, J. Nicolet, and A. P. Burnens. Identification of *Campylobacter jejuni* on the basis of a species-specific gene that encodes a membrane protein. *J Clin Microbiol*, 33(4):855–9, Apr 1995.

[22] S. Wright. Evolution in mendelian populations. *Genetics*, 16(2):97 – 159, 1931.

List of Figures

List of Tables

A Explanation of the zero-inflated Poisson model for poultry count data

Bacterial count data has two components. Firstly is the bacteria present or absent, and secondly if it is present how much of it is there. In order to capture these two components in our model we have used a Bayesian zero-inflated Poisson model. The model features two linear predictors: one informing the probability that the bacteria will be present and the other informing the amount of bacteria that will be obtained given that it is present.

We begin by describing the zero-inflation part of the model. We assume that sample s is positive for *Campylobacter* with probability p_s. Since p_s is a probability, the logit link function must be used when forming the linear predictor. We wish to model the change in the probability

of contamination quarter-by-quarter for each company, and so we include a random effect for each quarter/company pair. Thus, $\text{logit}(p_s) = \alpha_0 + \alpha_{c(s),q(s)}$ where $c(s)$ is the company that produced sample s and $q(s)$ is the quarter in which sample s was produced. This is a Bayesian model and so priors must be chosen. We assume the following weakly informative priors:

$$\begin{aligned} \alpha_0 &\sim N(0, 10), \\ \alpha_{c,q} &\sim N(0, \kappa_\alpha^{-1}), \\ \kappa_\alpha &\sim \text{Gamma}(2, 2). \end{aligned}$$

The second part of the model describes the amount of *Campylobacter*, given that it is present. We assume that the number of *Campylobacter* in each replicate from a contaminated sample will have a Poisson distribution, with the volume of rinse that is plated as an offset, ie. $Y_i \sim \text{Pois}(V_i\mu_i)$, where Y_i is the number counted from replicate i and V_i is the volume of rinse that is plated. To make a linear predictor for μ_i we use the log link function and include random effects for each company/quarter pair as before. To attempt to separate the laboratory variation from the sample variation, we also include random effects at the replicate level and the sample level. Thus, $\log(\mu_i) = \beta_0 + \beta_{c(i),q(i)} + \gamma_{s(i)} + \delta_i$. Again we assume weakly informative priors:

$$\begin{aligned} \beta_0 &\sim N(0, 10), \\ \beta_{c,q} &\sim N(0, \kappa_\beta^{-1}), \\ \gamma_s &\sim N(0, \kappa_\gamma^{-1}), \\ \delta_i &\sim N(0, \kappa_\delta^{-1}), \\ \kappa_\beta &\sim \text{Gamma}(2, 2), \\ \kappa_\gamma &\sim \text{Gamma}(2, 2), \\ \kappa_\delta &\sim \text{Gamma}(2, 2). \end{aligned}$$

Once we have obtained a large number of samples from the posterior distribution, we have used a trend line to assess the change over time in the probability of contamination and the predicted level of contamination on each chicken. When calculating the predicted level of contamination, we have used only birds that the model suggests are contaminated to arrive at the predicted level of contamination given that contamination has occurred. We therefore do not necessarily expect that these two outcomes will conform to the same pattern. To produce a confidence region around the trend lines, we calculated a trend line for each indivual sample from the posterior distribution and used these trend lines to calculate predicted values for each quarter. The 2.5 and 97.5 percentiles of these predicted values give the boundaries of the confidence region.

B Details of the modified Hald model

We have made several modifications to the Hald model in order to increase its applicability to *Campylobacter* and improve the identifiability. These modifications include

- removing the food consumption weighting M_j,
- incorporating random effects for the q_i parameters,

- incorporating uncertainty in the prevalence matrix p_{ij},
- including rare sequence types that are potentially pathogenic,
- changing the priors for the a_j and the q_i parameters.

Let y_i represent the number of human cases of type i and denote the expected number of cases of *Campylobacter* type i from source j by λ_{ij}. Assume that $y_i \sim \text{Poisson}(\sum_j \lambda_{ij})$ and that $\lambda_{ij} = p_{ij} q_i a_j$.

Parameter	Description
λ_{ij}	expected number of cases / year of type i from source j
p_{ij}	prevalence of type i in source j
q_i	bacteria-dependent factor for type i
a_j	food-source dependent factor for source j
y_i	number of human cases of type i
τ	hyperparameter controlling the variation of the q_i's

We assume the following conditionally independent prior distributions for the parameters.

- $a_i \sim \text{Exponential}(0.002)$,
- $\log(q_i) \sim \text{N}(0, \tau)$,
- $\tau \sim \text{Gamma}(0.01, 0.01)$,
- $p_{ij} \sim \text{Beta}(\alpha_{ij}, \beta_{ij})$.

The parameters α_{ij} and β_{ij} are determined by equating the first two moments of this prior with the first two moments of the posterior distribution obtained from a preliminary Bayesian analysis of the prevalence data for each source. In order to avoid convergence problems we enforced a minimum $\alpha_{ij} = 1$, correcting the corresponding β_{ij} so that the means of the prior and the posterior prevalence are equal. In the preliminary analysis of the prevalence, we assumed that $p_{ij} = \pi_i r_{ij}$, where π_i is the prevalence of *Campylobacter jejuni* in source i and r_{ij} is the relative occurrence of type j in the successfully typed *C. jejuni* isolates from source i. The priors used in this preliminary analysis of the prevalences were $r_{ij} \sim \text{Dirichlet}(1, 1, \ldots, 1)$ and $\pi_i \sim \text{Beta}(1, 1)$.

An estimate of the posterior distribution for this model was obtained using Markov Chain Monte Carlo (MCMC) techniques, run in the software WinBUGS 1.4 called from R 2.5.1 (using the `R2WinBUGS` package). The WinBUGS code was developed from the original code by Hald *et al.* A total of 10,000 samples were taken from five independent Markov chains with widely dispersed starting values, after a burn-in period of 2,000 iterations for each chain. Convergence was monitored using the method developed by Gelman and Rubins. The length of the chain was determined by running sufficient iterations to ensure the Monte Carlo errors for each parameter were less than 5% of the posterior standard deviation.

C Details of the Island model

In the Island model, the host populations are considered to exist on seperate 'islands', with mutation and recombination occurring on each island and also migration occurring between islands. Like the modified Hald model, the Island model is a Bayesian model and inference proceeds in two steps.

In the first step, the evolutionary model described above is used to provide estimates of the mutation, recombination and migration rates for each of the *Campylobacter* host populations (islands). This evolutionary model is quite complex (see [6]) and progress cannot usually be made unless an approximation is used for the likelihood function. This evolutionary model uses only the genotype data from isolates sampled from the host populations. Markov Chain Monte Carlo (MCMC) methods are used to obtain samples from the posterior distribution of the mutation, recombination and migration rates and these are transferred into the second stage of the modelling.

In the second stage, the human isolates are considered to be random samples from one of the island populations. Again MCMC methods are used to obtain the posterior probability that a human sample came from each one of the island populations, based on its genotype. These posterior probabilities are the probability that a human sample can be attributed to each source. The posterior distribution also includes estimates for the underlying proportion (and its uncertainty) of human cases attributable to each source. The strength of this method is its ability to use genotyping information to attribute human cases to a host population, even if the ST of the human case has not been observed in any of the source populations.

D Details of the spatial and temporal Bayesian Hierarchical model

Let $Y_{i,t}$ represent the number of notified cases of campylobacteriosis in meshblock i and week t. We assume that $Y_{i,t} \sim \text{Pois}(n_i \lambda_{i,t})$, where n_i is the usually resident population of meshblock i (obtained from the most recent census) and $\lambda_{i,t}$ represents the expected risk at this point in time and space. Next, we separate the risk into its spatial and temporal components through the relation $\log(\lambda_{i,t}) = R_t + U_i$, where R_t and U_i are the purely temporal and purely spatial components of the risk respectively. For the spatial component a Gaussian Markov Random Field prior is assumed (also called a Gaussian intrinsic autoregression) in which the risk in each meshblock is assumed to be similar to the mean risk of the neighbouring meshblocks. More formally, we assume the following full conditional distributions

$$U_i \sim N\left(\sum_{j \in n(i)} \frac{U_j}{|n(i)|}, \frac{1}{\kappa_U |n(i)|}\right),$$

where $n(i)$ is the set of meshblocks that are neighbours to meshblock i.

For the temporal component we assume a Gaussian second order random walk prior: that the change in risk from week t to week $t+1$ will be similar to the change in risk from week $t-1$

to week t, ie. given $R_1, \ldots, R_t$,

$$R_{t+1} - R_t \sim N\left(R_t - R_{t-1}, \frac{1}{\kappa_R}\right).$$

We assume flat priors for R_1 and R_2 so that the temporal component can absorb the baseline level of risk.

For the hyperparameters κ_R and κ_U we assume the weakly informative and conjugate gamma-distributed priors: $\kappa_R \sim \text{Gamma}(1, 10^{-4})$ and $\kappa_U \sim \text{Gamma}(1, 10^{-2})$. The effect on the posterior of these priors has been explored by considering other weakly informative gamma priors. Other choices of the prior parameters do not appear to influence the posterior distribution very strongly, and so we consider our analysis to be robust to the choice of prior.

E Distribution of sequence types among food and environmental sources

Table 12: Distribution of STs among poultry isolates. Only the STs isolated from poultry are shown, along with the number of human isolates with identical STs

	Allelic profile								Poultry company		
ST	*aspA*	*glnA*	*glt*	*gly*	*pgm*	*tkt*	*uncA*	Human	A	B	Other
45	4	7	10	4	1	7	1	42	36	21	30
474	2	4	1	2	2	1	5	157	32	2	0
257	9	2	4	62	4	5	6	12	12	4	0
53	2	1	21	3	2	1	5	28	11	4	3
190	2	1	5	3	2	3	5	22	10	0	0
50	2	1	12	3	2	1	5	24	8	23	14
583	4	7	10	4	42	51	1	11	8	0	4
520	2	1	12	88	2	1	5	8	6	0	2
52	9	25	2	10	22	3	6	17	6	0	0
42	1	2	3	4	5	9	3	19	4	0	0
354	8	10	2	2	11	12	6	23	3	0	1
25	4	7	10	1	1	7	1	1	2	0	2
3726	37	253	4	48	126	25	3	0	2	0	0
48	2	4	1	2	7	1	5	44	1	35	7
2345	2	4	5	2	2	5	5	3	1	12	0
227	2	4	5	2	2	1	5	0	1	0	0
2391	2	15	4	48	360	25	23	1	1	0	0
3721	8	10	149	2	1	12	6	0	1	0	0
3609	2	4	12	2	7	1	5	0	0	23	0
1517	8	10	149	2	11	12	6	4	0	7	0
1581	129	66	30	82	189	47	17	1	0	5	0
2397	184	39	30	82	113	43	17	0	0	4	0
21	2	1	1	3	2	1	5	7	0	1	0
677	10	81	50	99	120	76	52	6	0	1	0
1818	4	7	10	4	2	7	1	0	0	1	0
1911	7	84	5	10	119	178	26	0	0	1	0
3719	2	4	12	2	2	1	5	0	0	1	0
3725	37	1	57	26	107	29	35	0	0	1	0
451	2	1	2	3	2	3	5	8	0	0	27
3717	2	1	21	3	62	1	5	1	0	0	3
2535	10	2	107	62	120	76	1	0	0	0	1
3230	33	39	30	322	104	85	17	0	0	0	1
3711	1	2	4	62	4	5	17	2	0	0	1

Table 13: Distribution of STs among ruminant isolates. Only the STs isolated from ruminants are shown, along with the number of human isolates with identical STs

	Allelic profile								Ruminant source	
ST	*aspA*	*glnA*	*glt*	*gly*	*pgm*	*tkt*	*uncA*	Human	Cow	Sheep
53	2	1	21	3	2	1	5	28	27	2
50	2	1	12	3	2	1	5	24	17	27
42	1	2	3	4	5	9	3	19	13	25
61	1	4	2	2	6	3	17	16	11	15
520	2	1	12	88	2	1	5	8	10	5
190	2	1	5	3	2	3	5	22	7	8
2026	10	1	16	19	10	5	7	11	5	22
474	2	4	1	2	2	1	5	157	5	3
422	2	1	5	3	2	5	5	3	4	24
21	2	1	1	3	2	1	5	7	4	1
45	4	7	10	4	1	7	1	42	2	4
436	7	21	5	62	4	61	44	4	2	4
3072	33	39	30	82	104	173	68	1	2	0
u60	1	4	2	NEW	6	3	17	0	2	0
38	2	4	2	2	6	1	5	13	1	0
1115	53	39	30	81	118	44	36	0	1	0
2350	2	4	1	2	2	5	5	2	1	0
3714	2	1	12	3	2	1	52	0	1	0
3716	2	1	12	88	2	3	5	0	1	0
3723	33	39	12	82	104	173	5	0	1	0
u403a	NEW	27	16	19	10	5	7	0	1	0
1517	8	10	149	2	11	12	6	4	0	8
3610	2	1	5	88	2	11	5	0	0	3
3711	1	2	4	62	4	5	17	2	0	3
583	4	7	10	4	42	51	1	11	0	2
3232	32	39	30	82	104	324	17	0	0	2
u403	10	1	16	19	NEW	5	7	0	0	2
u61	1	4	2	2	NEW	3	17	0	0	2
u61b	1	NEW	2	2	6	3	17	0	0	2
393	24	17	2	10	23	68	5	0	0	1
618	1	4	2	2	6	3	5	0	0	1
677	10	81	50	99	120	76	52	6	0	1
1191	33	39	30	82	189	47	17	0	0	1
2392	9	25	2	283	22	3	6	0	0	1
3676	1	307	3	4	5	9	3	2	0	1
3719	2	4	12	2	2	1	5	0	0	1
3722	10	1	12	3	2	5	7	0	0	1
3724	33	39	149	2	104	12	6	0	0	1
u33	2	1	12	3	NEW	1	5	0	0	1

Table 14: Distribution of STs among water and wild bird isolates. Only the STs isolated from water and wild birds are shown, along with the number of human isolates with identical STs

	Allelic profile							Environmental source		
ST	*aspA*	*glnA*	*glt*	*gly*	*pgm*	*tkt*	*uncA*	Human	Water	Wild bird
2381	175	251	216	282	359	293	102	0	18	0
1225	27	33	22	49	43	7	31	0	7	0
45	4	7	10	4	1	7	1	42	5	12
422	2	1	5	3	2	5	5	3	4	0
436	7	21	5	62	4	61	44	4	3	0
3659	37	52	57	26	127	29	1	0	3	0
3663	175	6	216	282	261	7	3	0	3	0
50	2	1	12	3	2	1	5	24	2	0
177	17	2	8	5	8	2	4	0	2	0
474	2	4	1	2	2	1	5	157	2	0
526	2	15	4	27	13	80	23	0	2	0
3538	47	2	4	2	6	5	17	1	2	0
3655	1	6	5	282	261	7	3	0	2	0
3656	175	251	216	282	359	293	3	0	2	0
3660	192	295	216	282	359	293	102	0	2	0
3661	27	33	22	49	134	7	31	0	2	0
3662	27	33	22	49	43	110	31	0	2	0
3675	237	2	254	340	435	349	256	0	2	0
u45	4	NEW	10	4	1	7	1	0	2	0
u7	175	6	5	282	261	7	NEW	0	2	0
137	4	7	10	4	42	7	1	1	1	1
2354	37	4	4	48	13	25	23	0	1	1
25	4	7	10	1	1	7	1	1	1	0
42	1	2	3	4	5	9	3	19	1	0
61	1	4	2	2	6	3	17	16	1	0
677	10	81	50	99	120	76	52	6	1	0
694	2	59	4	105	126	25	23	0	1	0
1030	37	4	4	48	13	25	57	0	1	0
1223	27	33	22	49	43	9	31	0	1	0
1243	81	155	30	163	231	43	93	0	1	0
2347	2	4	4	105	10	25	57	0	1	0
2584	2	1	57	26	127	29	35	0	1	0
2619	191	251	216	282	359	293	214	0	1	0
3301	86	155	69	113	276	257	67	0	1	0
3610	2	1	5	88	2	11	5	0	1	0
3640	1	6	5	4	261	7	3	0	1	0
3657	27	33	22	104	134	7	31	0	1	0
3658	1	295	216	282	359	293	102	0	1	0
3664	37	52	4	26	127	29	23	0	1	0
3672	236	306	254	339	433	349	255	0	1	0
3673	175	6	216	4	434	7	3	0	1	0
3674	27	33	22	49	43	350	31	0	1	0
3676	1	307	3	4	5	9	3	2	1	0
2536	35	185	162	5	8	222	21	0	0	2
583	4	7	10	4	42	51	1	11	0	1
681	35	43	9	5	8	46	21	0	0	1
1324	99	128	91	125	170	146	111	0	0	1
2537	17	2	8	5	8	2	143	0	0	1
2538	35	2	8	51	361	2	21	0	0	1
2539	17	2	9	5	8	222	21	0	0	1