Estimating the contribution of different sources to the burden of human campylobacteriosis and salmonellosis

A thesis submitted in partial fulfilment of the
requirements for the degree of
Doctor of Philosophy
at Massey University, Palmerston North,
New Zealand,

by

Petra Müllner

2009

Preface

It was both odd and unjust, said Gauss, a real example of the pitiful arbitrariness of existence, that you were born into a particular time and held prisoner there whether you wanted it or not. It gave you an indecent advantage over the past and made you a clown vis-à-vis the future.

Measuring the World, Daniel Kehlmann 2005.

Seltsam sei es und ungerecht, sagte Gauss, so recht ein Beispiel für die erbärmliche Zufälligkeit der Existenz, da man in einer bestimmten Zeit geboren und ihr verhaftet sei, ob man wolle oder nicht. Es verschaffe einem einen unziemlichen Vorteil vor der Vergangenheit und mache einen zum Clown der Zukunft.

Die Vermessung der Welt, Daniel Kehlmann 2005.

Abstract

This thesis is concerned with the molecular epidemiology of *Campylobacter jejuni* and *Salmonella* in New Zealand and the development of source attribution tools for these pathogens. Although campylobacteriosis is the leading enteric zoonosis worldwide, the pathogen's complex epidemiology and difficulties with existing typing schemes, have posed challenges for the control of this disease.

The first study of this thesis gives an overview of existing approaches to microbial risk assessment and source attribution, with particular respect to campylobacteriosis, and describes their advantages and shortcomings. Further, the chapter discusses pheno- and genotyping techniques for *Campylobacter* spp. and the value of including microbial typing data in risk assessments. In the second study, data from a sentinel surveillance site in the Manawatu region was used to investigate the molecular epidemiology of human campylobacteriosis cases. This analysis revealed the presence of a dominant *C. jejuni* clone, namely sequence type (ST) 474, which accounted for 30.7 % of human cases in the study and identified risk factors for infection with ruminant and poultry associated STs. The third study investigated the link between *C. jejuni* in human cases and samples taken from poultry. By applying epidemiological and population genetic techniques this part of the thesis provided further evidence that poultry is a major contributor to human infection. In the fourth study an existing Bayesian source attribution model was modified and consecutively applied to New Zealand's major foodborne zoonoses: campylobacteriosis and salmonellosis. The majority (80 %) of human campylobacteriosis cases attributable to *C. jejuni* were estimated to have been acquired from poultry sources, whereas wildlife source were estimated to contribute only a minor proportion of cases. In the fifth study the *Salmonella* dataset was descriptively analysed and a large proportion of human cases was found to be caused by 'exotic' *Salmonella* types. In the final study of this thesis four different genetic and epidemiological source attribution methodologies were applied to the same dataset in a comparative modelling framework.

The studies in this thesis show that epidemiological studies combined with molecular tools and modeling can provide valuable risk-based tools to inform the surveillance and control of zoonotic pathogens. Methods from these studies may be readily applied to the control of other (food borne) zoonoses and provide new opportunities for epidemiological investigations and source attribution modelling of major pathogens.

Acknowledgements

I learned many things while working on this thesis and greatly appreciate the excellent training I received at Massey University. But it feels the most valuable improvement has not come from the technical side of things but from a new appreciation I acquired for other disciplines and my colleagues. Clearly my mind has become a broader one in the process.

I owe a lot to my main supervisor Nigel French. You have taught me a great many things and moved me out of my comfort zone to reach peak performance. Thank you for your kindness and encouragement in difficult times, your contagious enthusiasm and guidance.

I would like to thank the New Zealand Food Safety Authority for funding this project as part of their postgraduate programme and for the help of all those I came to know in the process. You provided valuable input for my thesis and it was great to learn from your expertise. Special thanks to the Science Group for their support in the last four years, in particular Steve Hathaway for initiating the program and taking the time to supervise and guide me. To Donald Campbell, Terry Ryan and Roger Cook for their advice, Peter van der Logt for his help and kindness, Paul Dansted for the invaluable noise cancelling headphones, Bob Jackman for his helpfulness, David Tisdall for making me laugh when it was much needed and Audrey Taulalo for those hours of 'girls time' in and out of the office. Thanks all for your annual (failed) attempts to make me understand your sense of humour for the Christmas skit.

Thanks to my supervisor Cord Heuer for getting me over here and started, and always being there when I needed his advice. Thanks to all my fellow students at the Epicentre and Hopkirk Institute for sharing their experiences with me. Thanks to Colleen Blair, Julie Dunlop, Wendy Maharey, Chris Cunningham and Simon Verschaffelt from the EpiCentre for their administrative and computing support and Bruce White for his help with EndNote.Thanks to the many people who made my project possible: the Hopkirk Lab Team (in particular Julie Collins-Emerson, Ann Midwinter, Errol Kwan, Rhukshana Akhter, Rebecca Pattison, Sarah Moore, Isabel Li and Diane

Richardson), MedLab Central (in particular Lynn Rogers), and Tui Shadbolt and her colleagues from MidCentral Public Health. A special thanks to the team at ESR in Kenepuru, in particular Phil Carter, for their help with the project. I owe the Hopkirk Team and Tui some special words for all the hard work and extra hours they put into the project. We pushed through hard times until my German directness and your Kiwi ingenuity melted us into a great team. I will miss working with you.

Thanks to Simon Spencer, Geoff Jones and Alasdair Noble for guiding me through the ups and downs of applied statistics and being key to the success of this project. Thanks to Martin Hazelton for valuable input into the development of the modified Hald model and Jonathan Marshall for his help with the publications. For stimulating discussion and the honour to be part of their group I would like to thank all members of the Enteric Zoonoses Modelling Group.

Thanks to all the people who helped me through the final steps of this thesis in various ways such as providing templates, de-bugging LaTeX and proof reading: Julie Collins-Emerson, Jonathan Marshall, Eve Pleydell, Deb Prattley, Daan Vink and Jacqui Benschop.

Thanks to my friends Jacqui Benschop (and her lovely family) and Eve Pleydell for giving me shelter and so many other things over the last years. Katie Owen for cheering me up over a coffee so many times. Petra Stübben, Jens Schwarz and Frederik for shared moments far away from the homeland and those early morning get-togethers watching the German soccer team play in championships. Thanks to all my other friends and family, most far away, for being there for me. To my parents-in-laws for all their support in our endeavours and visiting us in the final months of the thesis, providing me with time to fully immerse myself in those final steps. To my aunt and uncle Trude and Hartmut for caring so much about me, I couldn't wish for better god-parents to my children. To my cousin Ellen for the time we shared, with a promise to reply more often to my emails once this thesis is submitted. To my father, who did not live long enough, to meet my girls and share my happiness with me. Carina, Franzi, Kathrin, Soni, Brigitte, Susi, Kirsten and Janne it is great to have friends like you, even if we now live at different ends of this planet.

Most importantly thanks to my husband Uli for his support and love and keeping me from taking myself and life too seriously. You truly are the best thing that ever happened to me. And finally to my girls Lea and Stella, to the moon and back this is how much I love you!

Nomenclature

AFLP	Amplified Fragment Length Polymorphism
AMOVA	Analysis of molecular variance
BA	Blood agar
BPW	Buffered Petone Water
CAC	Codex Alimentarius Commission
CI	Confidence interval
CC	Clonal complex
cfu	colony forming unit
CrI	(Bayesian) credible interval
CRISPR	Clustered regularly interspaced short palindromic repeats
DNA	Desoxyribonucleic acid
ELISA	Enzyme-linked immunosorbent assay
ERL	Enteric Reference Laboratory
ERIC	Enterobacterial repetitive intergenic consensus
ESR	Environmental Science and Research Ltd.
EU	European Union
EZDRG	Enteric Zoonosis Disease Research Group
FAO	World Agricultural Organisation
FAME	Fatty acid methyl ester
HACCP	Hazard Analysis and Critical Control Points
HL	Heat-labile (antigen)
HS	Heat-stable (antigen)
mCCDA	Modified Cefoperazone Charcoal Desoxycholate agar
MCMC	Markov Chain Monte Carlo
MEE	Multi locus enzyme electrophoresis
MLST	Multi locus sequence typing

MPRM	Modular process risk model
MRA	Microbial risk assessment
MST	Microbial source tracking
MST	Minimum spanning trees
NMD	National Microbiological Database
NZ	New Zealand
NZFSA	New Zealand Food Safety Authority
OIE	World Organisation for Animal Health
OR	Odds ratio
PCR	Polymerase chain reaction
PFGE	Pulsed field gel electrophoresis
PHS	Public Health Service
PSI	Proportional similarity index
QRA	Quantitative risk assessment
RA	Risk assessment
RAPD	Randomly amplified polymorphic DNA
rDNA	Ribosomal DNA
RE	Restriction enzyme
REA	Restriction endonuclease analysis
RFLP	Restriction fragment polymorphism
SD	Standard deviation
spp.	Species (multiple)
SPS	Sanitary and phytosanitary (agreement)
ST	Sequence type
SVR	Short variable region
US-NAS	United States National Academy of Science
WHO	World Health Organisation
WTO	World Trade Organisation

Publications

Mullner, P., Spencer S., Wilson D., Jones G., Noble A. D., Midwinter, A., Collins-Emerson, J., Carter, P., Hathaway, S., French, N. P. Assigning the source of human campylobacteriosis in New Zealand: A comparative genetic and epidemiological approach. *Infection, Genetics and Evolution* (9): 1311-1319 2009.

Mullner, P., G. Jones, A. Noble, S. Spencer, S. Hathaway, and N. French. Source Attribution of Food Borne Zoonoses in New Zealand: A Modified Hald Model. *Risk Analysis*(7): 970-984 Jul 2009.

Mullner P., Spencer S., Jones G., Noble A. D., Wilson D., Collins-Emerson, J., Midwinter, A., Carter, P., Hathaway, S., French, N. P. Attribution of campylobacteriosis in New Zealand: Approaches, results and outlook. *Proceedings of the 12th Symposium International Society for Veterinary Epidemiology and Economics*, Durban, South Africa: ISVEE 12, 432, 2009.

Mullner, P., Benschop, J. Interdisciplinary research at the human and animal health interface: A personal perspective. *Proceedings of the 12th Symposium International Society for Veterinary Epidemiology and Economics*, Durban, South Africa: ISVEE 12, 125, 2009.

Mullner, P., Collins-Emerson, J., Midwinter, A., Carter, P., Hathaway, S., French, N. P. Structured parallel stuides of the molecular epidemiology of *Campylobacter jejuni* in environmental and food sources reveal insight into the importance and interconnection of different reservoirs. *Proceedings of the 12th Symposium International Society for Veterinary Epidemiology and Economics*, Durban, South Africa: ISVEE 12, 431, 2009.

Mullner, P., Shadbolt, T., Spencer, S., Jones, G., Noble, A. D., Collins-Emerson, J., Midwinter, A., Carter, P., Campbell, D. M., Van der Logt, P., Lee, J., Hathaway, S., French, N. P. Enhanced surveillance to identify the origin of human campylobacteriosis in New Zealand: Conclusions drawn from an interdisciplinary sentinell surveillance site. *Proceedings of the 12th Symposium International Society for Veterinary Epidemiology and Economics*, Durban, South Africa: ISVEE 12, 379, 2009.

French, N.P., **Mullner P.**, Spencer S. Controlling campylobacteriosis in New Zealand: combining epidemiology and population genetics to trace the origin of human infections. *New Zealand Microbiological Society, Annual Conference 2008*, Christchurch, NZ.

G. McBride, P. van der Logt, R. Lake, N. French, A. Ball, **P. Mullner**, A. Elliott, A. Hunt and L. Oakley. Modelling initiatives for campylobacteriosis. *New Zealand Microbiological Society, Annual Conference 2008*, Christchurch, NZ.

French, N.P., **Mullner P**. Risk-based tools to inform decision making in food safety and public health - applications in Europe and New Zealand. *FOOD FRENZ, Annual Conference 2008*, Budapest, Hungary.

Mullner P., Collins-Emerson, J., Midwinter, A., Spencer, S., French, N. Molecular and Modelling tools for *Campylobacter* Source Attribution. *Proceedings of the Epidemiology & Animal Health Management Branch of the NZVA*, 2008, Wellington, NZ.

Mullner P., Pleydell, E., Shadbolt, T., Collins-Emerson, J., Midwinter, A., Spencer, S., French, N. Working at the human and animal health interface - a multidisciplinary team approach towards reducing New Zealand's campylobacteriosis problem. *Proceedings of the Epidemiology & Animal Health Management Branch of the NZVA, 2008*, Wellington, NZ.

French, N.P., Carter P., Collins-Emerson J., Midwinter A., **Mullner P.**, Wilson, D. Comparing 'source attribution' models for human campylobacteriosis. *Society for Veterinary Epidemiology and Preventive Medicine, Annual Conference 2008*, Liverpool, UK.

Lake, R., van der Logt, P., McBride, G., French, N., **Mullner, P.**, Andrew Elliott, Andrew Ball. Campylobacteriosis in New Zealand: modelling as a way forward. *NZ Science Review* 2007 Volume 64 (2) pp 37-41.

Mullner P. Source attribution of *C. jejuni* using molecular epidemiology. *New Zealand Institute of Food Science and Technology, Annual conference 2007*, Wellington, NZ.

Mullner P, Carter P, Midwinter A, Collins-Emerson J, Pattison R, French NP. Source attribution of *Campylobacter jejuni* using molecular epidemiology. *Proceedings of the Epidemiology & Animal Health Management Branch of the NZVA, 69, 2007.*

Mullner, P., Carter, P., Cook, R., Tebje-Kelly, J., French, N. P. Strain Typing As A Valuable Tool For Quantitative Microbial Risk Assessment. *Proceedings 11th International Symposium ISVEE*, August 6-11, 2006, Cairns, Australia.

Mullner, P., Carter, P., Midwinter, A., Collins-Emerson, J., French, N. P. Molecular epidemiology of *Campylobacter jejuni*: comparing Multilocus Sequence Types from humans and poultry. *Proceedings 11th International Symposium ISVEE*, August 6-11, 2006, Cairns, Australia.

Terry Ryan, Stuart MacDiarmid, Steve Hathaway, **Petra Mullner**, Neil Cox, Andrew Hill. A Quantitative Analysis of Foodborne Salmonellosis Exposure Pathways in New Zealand. *Proceedings of the International Conference on 'Priority Setting of Foodborne and Zoonotic Pathogens'*, MedVetNet and the US Food Safety Research Consortium, 19-21 July 2006, Berlin, Germany.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Introduction

Campylobacteriosis is the leading enteric zoonosis in the developed world and a major public health concern worldwide (Oberhelman and Taylor 2000). The majority of human of cases are caused by *Campylobacter jejuni* (Gormley, MacRae et al. 2008) which differs in its epidemiology from other *Campylobacter* species (Gillespie, O'Brien et al. 2002). Most *C. jejuni* infections are believed to result from the ingestion of contaminated food (Eberhart-Phillips, Walker et al. 1997), although the role of other, non-food exposures in the epidemiology of sporadic campylobacteriosis is still unknown (Cowden 1992). The pathogen has been isolated from a range of food sources including poultry (Harris, Weiss et al. 1986; Eberhart-Phillips, Walker et al. 1997; Kramer, Frost et al. 2000; Pearson, Greenwood et al. 2000), red meat and milk (Fahey, Morgan et al. 1995; Gillespie, Adak et al. 2003). The primary source of contamination is believed to be animal faeces and this is consistent with both the high carriage rates in poultry, pigs and cattle (Kramer, Frost et al. 2000; Savill, Hudson et al. 2003) and with molecular evidence showing similar genotypes in farm animals and humans (Dingle, Colles et al. 2001; Fitzgerald, Stanley et al. 2001; Savill, Hudson et al. 2003; French, Barrigas et al. 2005). Contamination of the environment by domestic and wild animal faeces presents alternative exposure pathways for human infection; for example via drinking water (Duke, Breathnach et al. 1996; Savill, Hudson et al. 2001; Champion, Best et al. 2002; Said, Wright et al. 2003) and recreational water contact (Adak, Cowden et al. 1995; Savill, Hudson et al. 2001).

New Zealand has among the highest enteric infectious diseases rates in industrialised countries (Lake, Baker et al. 2000) with the high ratio of domestic production animals to humans and the frequent use of rural water supplies in New Zealand having been discussed as underlying causes (Crump, Murdoch et al. 2001). The Acute Gastrointestinal Illness (AGI) Study estimates that more than five million working days are lost each year in New Zealand due to gastrointestinal illness (Lake, Adlam et al. 2007) with the number of campylobacteriosis notifications increasing dramatically

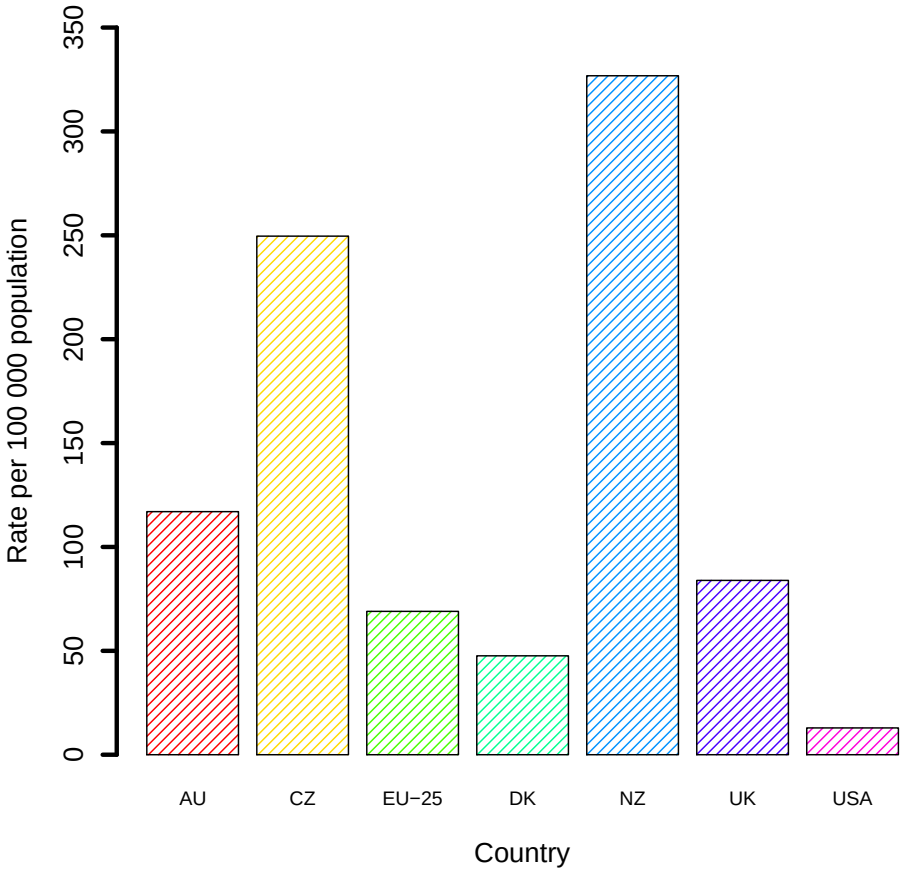

Figure 1.1: Campylobacteriosis rates in different countries (2004).

in the last decade costing the economy an estimated $NZ 77 million annually (Scott, Scott et al. 2000). In international comparison notification rates were high (Figure 1.1). Notifications exceeded 300 cases per 100,000 people per annum in 2005 and 2006 (Institute of Environmental Science and Research Limited (ESR)), equalling an annual total of 13,836 and 15,873 cases (respectively) nationwide. The situation was labelled a country-wide epidemic and there was a call for urgent action to reduce the human burden of disease (Baker, Wilson et al. 2006). The topic received an immense amount of media attention nationwide.

Salmonellosis is among the most common causes of foodborne diarrhoeal disease worldwide (Hald, Wegener et al. 2005). In industrialised countries the main sources of infection are animal-derived products, notably fresh meat and poultry products, whereas in developing countries contaminated water, vegetables and human contacts are believed to play a major role in disease transmission (Hald, Wegener et al. 2005). Salmonellosis is, after campylobacteriosis, the second most frequent zoonosis in New

Zealand. The rate of salmonellosis notifications has ranged from 26.5 to 62.3 cases per 100,000 population between 2000 and 2008 (Institute of Environmental Science and Research Limited (ESR)).

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 have been used worldwide (Batz, Doyle et al. 2005). A key part to successful disease control is the quantification of the contribution of different disease sources to the human burden of disease as such data can inform targeted surveillance, research activities and risk mitigation strategies. A recently developed source attribution model for salmonellosis (Hald, Vose et al. 2004) has shown many advantages over traditional epidemiological methods (Mullner, Carter et al. 2007) and this approach has now been used to model sources of salmonellosis in several countries (Pires, Nichols et al. 2008; Wahlstroem, Anderson et al. 2008). However while there is a strong need for such models for several zoonotic pathogens this method has not previously been extended to other diseases. Besides the application of epidemiological models to apportion cases of human disease, new opportunities for source attribution have recently emerged using the population genetics of pathogens (Wilson, Gabriel et al. 2008; Sheppard, Dallas et al. 2009).

Although common typing methods, such as serotyping (Nielson 2004), have routinely been applied to *Salmonella* and provided a valuable tool for the characterisation of this pathogen, the successful application of a variety of typing tools to *Campylobacter* has been hindered by its high genetic diversity. Historically, isolates of *Campylobacter* have been routinely characterised by different methods including serotyping and phage typing. However the typing schemes available were unable to rationalise the diversity present within *Campylobacter* populations and therefore could not be used to associate groups of isolates with related genotypes. Furthermore, phenotyping data could only provide limited information on the relatedness of isolates. In order to overcome the problems associated with phenotyping methods, a number of genotyping methods have been applied to *Campylobacter* spp. However, these genotyping schemes have also met with varying success, due to the high rate of exchange of genetic material that has been observed for this pathogen (Sheppard, McCarthy et al. 2008). Of the available genotyping methods Pulsed Field Gel Electrophoresis (PFGE) has proved to be a valuable method of strain typing and this method will continue to be an important tool allowing both for national and international comparison of isolates. However, this method has a number of disadvantages related to the difficulty of standardising the laboratory methods and the comparison of banding patterns. In light of these difficulties the development of multilocus sequence typing (MLST) as an alternative to whole genome sequencing has facilitated epidemiological investigations of pathogens of major public health importance and has been sucessfully applied to highly genetic diverse pathogens

such as *C. jejuni* (Dingle, Colles et al. 2001). For these reasons sequence typing by MLST is now emerging as the new gold standard for national and international epidemiological characterization of *C. jejuni* (Dingle, Colles et al. 2001; Dingle, Colles et al. 2002). Nucleic acid sequence-based methods such as MLST have many benefits for strain characterisation including reproducibility, standardization, auditablity and electronic portability. Sequence typing data are a universal currency for national and international, spatial and temporal comparisons, thus allowing robust phylogenetic analyses of organisms.

1.2 The Manawatu Sentinel Surveillance Site

Sentinel surveillance systems involve selected reporting site(s) from which information collected may be extended to the general population. Concentrating resources in the defined site(s) provides a wealth of information, thus enabling more accurate final estimates than those normally available from broader national surveillance programmes. More comprehensive than traditional passive approaches, sentinel surveillance includes standardised microbiological and epidemiological methods, providing useful information for risk assessments and the investigation of diseases.

The Manawatu study consisted of isolates from human cases of campylobacteriosis as well as environmental and food sources across a defined geographical area of New Zealand and extending over a three year period. At its core were two structured and parallel studies of the reported human cases of campylobacteriosis in a geographically defined region, as well as the potential sources of *C. jejuni* within the same region. The intention was to sample disease cases and exposure sources at the same time and in the same place as part of an effort to assess the contributions of different sources to the burden of human disease. A total of 2,348 samples were collected during the study period, including samples from 773 human cases that originated in the study area. The total number of samples included human cases, faecal samples collected from sheep and cattle on farms, environmental water samples and retail chicken and red meat samples purchased in the region. Almost all samples underwent further molecular analysis, applying PCR based speciation techniques and MLST analysis. The project was interdisciplinary with a wide range of speciality fields involved. The collection of human samples from the regional diagnostic laboratory was supported by a collaboration between the Hopkirk Research Institute, the regional diagnostic laboratory MedLab Central and the MidCentral Health, Public Health Service (PHS). The laboratory team provided advice and expertise on all molecular biological and genetic aspects of the study, and the data were analysed by combining epidemiological and statistical expertise.

The increasing availability of tools such as MLST is enabling researchers to investigate transmission patterns within populations at the strain level and to evaluate

host- and strain-specific risk factors, which is changing the way epidemiologists study infectious diseases (Murray 2002). By applying novel typing methods and combining them with a comparative modelling approach, this thesis aims to provide a better understanding of the complex epidemiology of *C. jejuni*. Although no gold standard was available the development of a set of models and approaches allowed them to be used as cross-references. Consistent results from different models, with different underlying assumptions, strengthened our understanding and confidence in the results obtained.

This study was embedded within a larger multidisciplinary initiative to control foodborne zoonoses in New Zealand and it has informed policy making. After the peak of cases in 2006 the New Zealand Food Safety Authority (NZFSA), supported by preliminary findings from the Manawatu Sentinel Surveillance Site, introduced its *Campylobacter* in Poultry Risk Management Strategy [1] . As a likely result of interventions introduced in the poultry production sector, human campylobacteriosis notifications dropped to a 16-year low of 6,689 cases in 2008 (Figure 1.2)

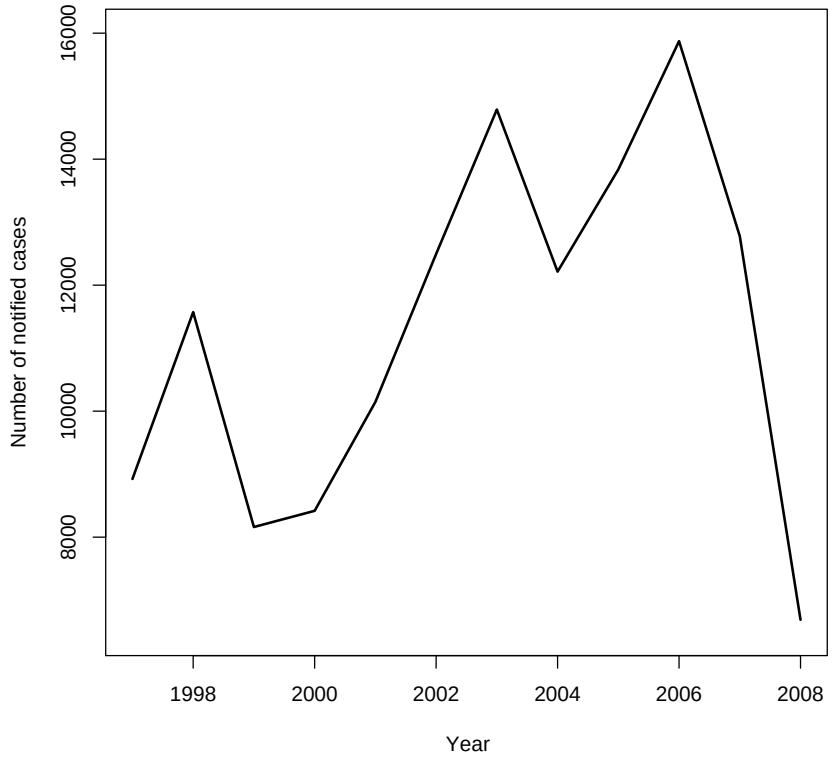

Figure 1.2: Number of notified campylobacteriosis cases in New Zealand 1997 - 2008.

[1]Full document available at: http://www.nzfsa.govt.nz/consumers/food-safety-topics/foodborne-illnesses/campylobacter/strategy/index.htm

1.3 The structure of this thesis

The aim of this thesis was to estimate the contribution of different food sources to the burden of human campylobacteriosis due to *C. jejuni* by combining molecular, epidemiological and modelling tools. Data were sourced from the Manawatu sentinel surveillance site from 2005 to 2008. In addition, based on routine surveillance data, cases of salmonellosis were apportioned and the epidemiology of this pathogen in human cases and food sources was described.

All chapters of this thesis, except for the introductory literature review and description and analysis of the *Salmonella* data, are presented in the form of manuscripts for peer-reviewed publication. As a consequence of this style of thesis presentation, there is some repetition between chapters. Furthermore, the process of writing for publications includes responding to co-authors', reviewers' and editors' requests and requires a substantial amount of distilling of material and presentation of only the most pertinent results.

This thesis begins with a review of typing techniques for typing *Campylobacter spp.* and techniques for microbial risk assessment, including source attribution (Chapter 2). The combination of the two approaches to develop risk (attribution) models is discussed, along with which typing techniques are most suitable for epidemiological investigations. The chapter also discusses the impact of ignoring strain variation in microbial risk assessment.

Chapters 3 and 4 describe in detail the findings from molecular epidemiological studies of *C. jejuni* in humans and poultry. Inferences are made about the importance of poultry sources based on the prevalence of *Campylobacter* and the similarity of *Campylobacter* subtypes isolated from poultry and human cases. By combining human surveillance data with typing information, risk factors for infection are identified.

Chapters 5 and 7 of this thesis aim to quantify the contribution of different food sources to the human disease burden of enteric zoonotic diseases. While Chapter 5 describes in technical detail the adaptation of the Hald model to New Zealand for both campylobacteriosis and salmonellosis, in Chapter 7, different methodologies are combined with the aim of allocating cases of human disease to a source within a novel risk attribution framework for campylobacteriosis. Chapter 6 provides a descriptive analysis of the salmonellosis dataset.

This thesis concludes with a discussion of the findings of these studies and their implications for the control and prevention of zoonotic pathogens (Chapter 8). I present an outlook on future applications for the methodologies described and discuss the advantages and shortcomings of the key elements of this thesis.

Chapter 2

A review of the role of strain variation in microbial risk assessment and source attribution with particular reference to typing approaches for *Campylobacter* spp.

2.1 Summary

Campylobacteriosis is the most commonly notified disease in New Zealand and a major public health problem worldwide. Identifying the sources and pathways of human infection is an important step towards prioritising interventions and the development of control strategies. Microbiological risk assessments (MRAs) are increasingly used to understand how pathogens are propagated along the food chain and to improve the quality and safety of food. A number of different model-based approaches have evolved and have been applied in microbial food safety. Strain typing is an essential part of recently developed source (or risk) attribution models, which use the distribution and genetic relatedness of strains found in (food) sources and human isolates to quantify the relative contribution of those sources to the burden of human disease and thereby constitute an important tool for priority setting. However in general, few MRAs consider the effects of strain variation on survival and growth of the pathogen in for example the food chain, or on the dose-response relationship. Using the example of campylobacteriosis, we describe the implications of ignoring strain variation in MRA, highlight the importance of strain typing in the epidemiologic investigations of this pathogen and evaluate the range of methods of sub typing *Campylobacter* spp. with particular reference to their usefulness in risk research for *C. jejuni*. In particular we compare serotyping with restriction fragment length polymorphism (RFLP) methods and multi

locus sequence typing (MLST) using a three point ranking system, which takes into account performance criteria such as discriminatory power and inter-laboratory stability. We conclude that, while no single method performs well in all criteria, MLST is currently the method that has the most desirable characteristics for use in *Campylobacter* spp. risk assessment and epidemiological research, in particular as it can be used to investigate the pathogen populations at a level of discrimination, which allows for identifying sources of human infection.

2.2 Introduction

Campylobacteriosis is the most commonly notified disease in New Zealand and a major public health problem worldwide (Baker, Wilson et al. 2006). In New Zealand the number of notifications has increased steadily in the last decade, peaking at 379 per 100,000 population in 2006 costing the economy an estimated NZ $77 million annually (Scott, Scott et al. 2000). *Campylobacter* enteritis is an acute diarrhoeal disease with clinical manifestations similar to those of other acute bacterial infections of the intestinal tract, for example *Salmonella* or *Shigella* infections. The clinical consequences depend on the virulence of the infecting strain, the challenge dose, and the susceptibility of the patient. The estimated mean incubation period is 3.2 days, ranging from eight hours to eight days. Campylobacteriosis can lead to late-onset complication such as Guillain-Barre Syndrome, an autoimmune-mediated disorder of the peripheral nervous system (Skirrow and Blaser 2000). The majority of reported campylobacteriosis cases are caused by the species *C. jejuni* followed by *C. coli* and *C. lari*. Thermotolerant *Campylobacter* belong to the family Campylobacteriaceae and are Gram-negative oxidase-positive motile bacteria. They are highly infectious pathogens and as little as 500 cells can cause infection (Donnison and Ross 2004).

In common with the rest of the world, most infections in New Zealand are believed to result from the ingestion of contaminated food (Eberhart-Phillips, Walker et al. 1997), although the role of other, non-food exposures in the epidemiology of sporadic campylobacteriosis is still unknown (Cowden 1992). Poultry are the most commonly implicated source of human infection from epidemiological and microbiological studies; however until recently the evidence for this association was often weak and circumstantial. An improved understanding of the relative contribution of different sources of human exposure to *Campylobacter* spp. could inform targeted surveillance, research activities and risk mitigation strategies.

Isolates of *Campylobacter* spp. have been routinely characterised by phenotypic methods including serotyping and phage typing. These methods have failed to provide the necessary discrimination and 'typeability' to inform the epidemiology of *Campylobacter* infections. Using current schemes, a significant number of isolates are non-typable leading to ambiguous data. In addition, phenotyping data provide limited

information on the relatedness of isolates. In order to overcome the problems associated with phenotypic methods a number of genotyping methods have been applied to *Campylobacter* spp. with varying success, in particular as a result of the high rate of genetic exchange observed for this pathogens (Sheppard, McCarthy et al. 2008).

Several Microbial Risk Assessments (FAO/WHO; Rosenquist, Nielsen et al. 2003; Nauta, van der Fels-Klerx et al. 2005) have been developed for *Campylobacter* spp. and have provided valuable information on transmission pathways, mitigation opportunities and many other aspects of disease control. However to date only few approaches (Sopwith, Ashton et al. 2003) go beyond the species level to investigate this pathogen. Integrating strain effects into risk research could potentially lead to a better understanding of *Campylobacter*'s complex epidemiology which includes a variety of potential sources of infection. In addition the ability to differentiate pathogenic from non-pathogenic strains could greatly improve the utility of risk assessment models and more closely define the sources and routes of transmission for human infection. Specific host-pathogen combinations could be identified using this approach and variation in virulence, or the dose-response relationship, could be taken into account. This is of great importance as the identification of effective food safety intervention requires a thorough understanding of the relationship between food and pathogen from farm to consumption.

New molecular epidemiological risk research tools are emerging, which enable researchers to make inferences about the contribution of individual food sources to the human disease burden. Essentially they are either based on the relative frequency of bacterial subtypes in different populations (Rosef, Kapperud et al. 1985; Hald 2002) or the population genetics of the pathogen (Excoffier, Smouse et al. 1992; Champion, Gaunt et al. 2005; Didelot and Falush 2007). In combination with the recent advancements in genotyping they provide an opportunity to gain a better insight into the epidemiology of *Campylobacter* spp. This study assesses current typing techniques of *Campylobacter* spp. as well as methods of risk assessment and recent advances in source attribution research. In particular the advantages of using strain typing data in risk assessment and research are discussed.

2.3 Review of typing methods for *Campylobacter* spp.

Molecular subtyping methods are applied to identify common strains of human cases and possible sources of infection during the investigation of outbreaks of disease (short-term or epidemic epidemiology). In addition, they are used for the investigation of the long-term (global) epidemiology of diseases, which is important to, for example, identify risk factors for infection or for recognising emerging or re-emerging strains of zoonotic pathogens (Sails, Swaminathan et al. 2003). Many typing methods have been developed over recent years, and two major systems of typing approaches are available,

phenotyping and genotyping. Each system has its own advantages and disadvantages and no single typing method has yet been found to be universally applicable (Sails, Swaminathan et al. 2003). The sheer number of methods available adds to the complexity of the epidemiology of *Campylobacter* infections in humans and food animals (Advisory Committee on the Microbiological Safety of Food (ACMSF) 2005).

The validation of typing methods includes an evaluation of their performance (Nielsen, Engberg et al. 2000). According to The European Study Group on Epidemiological Markers (1996) criteria for the evaluation of typing systems can be divided into two categories: performance (efficacy) and convenience (efficiency) criteria. Major performance criteria are typeability, reproducibility, stability and discriminatory power (see Appendix A). Convenience criteria on the other hand may be important for the selection of appropriate typing systems depending on a number of factors, including the scale of the epidemiologic investigation, the timeliness of information and the financial and technical resources needed. The following may be considered: flexibility, rapidity, accessibility and ease of use (Struelens and Members of the European Study Group on Epidemiological Markers (ESGEM) of the European Society for Clinical Microbiology and Infectious Diseases (ESCMID) 1996).

Since different investigations require a different level of efficacy and efficiency, an ideal typing system for universal use does not exist (Struelens and Members of the European Study Group on Epidemiological Markers (ESGEM) of the European Society for Clinical Microbiology and Infectious Diseases (ESCMID) 1996). As a consequence the quality of the microbial type assessment depends on the choice of technology and the interpretation of the experimental data within a sound theoretical framework (van Belkum 2003).

2.3.1 Phenotyping

Phenotyping methods such as biotyping, serotyping and phage typing are useful tools for strain characterisation and are still in widespread use despite their limitations (Fitzgerald, Helsel et al. 2001). Traditionally infectious agents have been differentiated according to the following phenotypic properties: growth and morphology, biochemical, serologic, and functional or physiologic characteristics (Riley 2004). Strain characterisation by existing phenotyping methods for *Campylobacter* is sometimes ambiguous due to the unavailability of standardised reagents, cross-reactivity between certain antigens and the high proportion of non-typeable strains (Jackson, Fox et al. 1996). In addition, serotyping and phage typing require the production and maintenance of panels of reagents, which are labour-intensive and costly, making these methods impractical for most clinical laboratories (Sails, Swaminathan et al. 2003). According to Duim et al. (1999) phenotyping methods are not generally available, due to lack of specific reagents, restricted differentiation power and the presence of a high proportion of non-typeable

strains. Whilst current phenotyping methods may provide adequate discrimination for outbreak investigations, according to Jackson (1996) the epidemiology of sporadic *Campylobacter* spp. infection often remains ill-defined when using this approach.

Serotyping

Penner (heat-stable) and Lior (heat-labile) serotyping have commonly been used to classify *C. jejuni* strains (Leonard, Takata et al. 2003). Penner serotyping identifies soluble heat-stable (HS) antigens by passive hemagglutination (Frost, Kramer et al. 1999), while the Lior serotyping scheme detects variation in heat-labile (HL) antigens (Newell, Frost et al. 2000). These methods, of which the method developed by Penner has gained more recognition internationally (Klena 2001), have contributed significantly to the development of successful surveillance programs for *C. jejuni* (van Belkum 2003). The principle behind each technique is the same; *Campylobacter* cells produce surface molecules that are recognised by immune systems, and antibodies generated to these different cell surface components may therefore be used to discriminate between isolates (Klena 2001). Serotyping is practical and valid but laborious, time consuming and costly as a panel of antisera is needed. It can produce ambiguous results due to low typeability, reproducibility and stability caused by non-typeable strains, transient antigen expression and cross-reactivity between antigens (Fitzgerald, Helsel et al. 2001). Furthermore, besides lacking reproducibility among laboratories the method is deficient in discriminatory power (Colles, Jones et al. 2003). Additional phenotyping techniques (for example phage typing (Newell, Frost et al. 2000)) have been applied to improve the quality of epidemiological analysis by serotyping (Jackson, Fox et al. 1996).

Phage typing

Campylobacter cells produce a wide assortment of surface molecules, which are used by viruses that attack bacteria (bacteriophages) as receptors for attachment to the bacterial cell. The expression of different surface structures determines whether a phage will be able to infect, and thus influence the phage type (Klena 2001). Phage typing provides good resolution for epidemiological typing, but is compromised by the instability and lack of expression of phage receptors (Jackson, Fox et al. 1996). In terms of stand-alone typing methods, phage typing therefore performs poorly on three criteria: typeability, reproducibility and discrimination (Klena 2001).

Multilocus enzyme electrophoresis (MEE)

MEE is a band-based molecular method for subtyping of *Campylobacter* spp., that indexes changes affecting the electrophoretic mobility of the gene product in highly conserved housekeeping genes (Sails, Swaminathan et al. 2003). It is a valuable typing tool that is relatively easy to perform but slow in through-put, labour-intensive and

not applicable for use in a routine laboratory due to equipment requirements (Stanley, Wareing et al. 2004). However this phenotypic analysis of electrophoretic variants of a set of housekeeping genes scores allelic variation of multiple genes and thus is also an indirect but well-validated genotyping method (Struelens and Members of the European Study Group on Epidemiological Markers (ESGEM) of the European Society for Clinical Microbiology and Infectious Diseases (ESCMID) 1996).

Biotyping

Methods assessing an organism's ability to i) produce or utilise biochemical substrates and ii) replicate on media containing potentially inhibitory compounds or under potentially adverse environmental conditions, are collectively known as biotyping (Klena 2001). The materials necessary for determining an isolate's biotype are available to most clinical and research laboratories (Klena 2001) and the method is easy to use, score and interpret, even in small laboratories (Struelens and Members of the European Study Group on Epidemiological Markers (ESGEM) of the European Society for Clinical Microbiology and Infectious Diseases (ESCMID) 1996). However the information gathered is, by itself, too general to subtype strains of *Campylobacter* and therefore lacks discrimination (Struelens and Members of the European Study Group on Epidemiological Markers (ESGEM) of the European Society for Clinical Microbiology and Infectious Diseases (ESCMID) 1996; Klena 2001). Due to its low discriminatory power biotyping is often used in combination with other typing systems, such as serotyping (Steele, McNab et al. 1998).

Fatty Acid Profile Typing

The Microbial Identification System utilises gas chromatography to analyse fatty acid methyl esters (FAMEs) generated after acid hydrolysis of whole bacterial cells. This system was designed for bacterial species identification and has shown potential for epidemiological typing of bacterial isolates, but proved not to be discriminatory enough for *Campylobacter* spp. (Steele, McNab et al. 1998; Klena 2001).

2.3.2 Genotyping

Genotyping methods, such as pulsed field gel electrophoresis (PFGE) (Schwartz D. C. and R. 1984) or amplified fragment length polymorphism (AFLP)(Vos, Hogers et al. 1995), have become increasingly popular. These methods produce 'genetic fingerprints', each requiring a different experimental approach, and as a consequence vary in their level of discrimination (De Boer, Duim et al. 2000). Genotyping allows for computer assisted data analysis in a numeric manner, catering for data sharing, and eases the processing of large numbers of samples (De Boer, Duim et al. 2000).

Various generations of genotyping technologies can be identified: plasmid fingerprinting was succeeded by southern blotting procedures, PFGE, polymerase chain reaction (PCR) approaches and, ultimately, deoxyribonucleic acid (DNA) sequencing of limited regions in microbial genomes (van Belkum 2003). According to Duim et al. (1999), these methods often have lacked discriminatory power and had poor reproducibility and therefore highly sensitive and reliable genomic typing methods are needed. However immense progress has been made in the last decade through the development and evaluation of new techniques (De Boer, Duim et al. 2000; Newell, Frost et al. 2000; Fitzgerald, Helsel et al. 2001; Leonard, Takata et al. 2003).

All genotyping procedures rely on variants of a small number of basic approaches, often used in combination, either applied to particular parts of the genome or the whole genome itself. The ultimate method for detecting DNA sequence variation is to compare nucleotide sequences directly, however this has not generally been a practical approach, although new technology is rapidly changing this scenario (Thompson 2000). Recently comparative phylogenomics using DNA micro arrays for whole genome comparisons (genomotyping) has been found to be a tool for strain discrimination (Champion, Gaunt et al. 2005). Genotyping methods have not been standardized or broadly accepted according to Dingle et al (Dingle, Colles et al. 2001). However this is now changing with two databases accessible through the World Wide Web: PulseNet (for PFGE; http://www.cdc.gov/pulsenet/) and PubMLST (for MLST; http://pubmlst.org/) (Fitzgerald, Helsel et al. 2001; Dingle, Colles et al. 2005).

Pulsed Field Gel Electrophoresis (PFGE)

This method is a modification of the more traditional restriction enzyme analysis (REA), which has been used for the subtyping of bacterial strains for several years (Newell, Frost et al. 2000). Macro restriction profiling uses restriction enzymes (RE) that digest the genomic DNA infrequently. This generates a profile of between four and twenty distinct bands per sample (Klena 2001). PFGE is the currently most accepted technique (De Boer, Duim et al. 2000) for *Campylobacter* spp. which examines polymorphism throughout the genome (Fitzgerald, Helsel et al. 2001). This method delivers high resolution (De Boer, Duim et al. 2000), excellent discriminatory power (Newell, Frost et al. 2000) and has a rapid protocol available (Fitzgerald, Helsel et al. 2001). However it needs special apparatus and is time consuming and laborious and therefore unsuitable for typing large numbers (De Boer, Duim et al. 2000). PFGE patterns can differ due to genetic variability and the low number of bands observed in *Campylobacter* spp. Compared to other organisms can make the interpretation of results difficult due to a lack of stability and reproducibility, (Duim, Wassenaar et al. 1999). Additionally inter laboratory variation makes the comparison of results difficult. It is well established that PFGE profiles of related strains can be altered by a variety

of genetic phenomena, including point mutations in restriction sites and genomic rearrangements (Sails, Swaminathan et al. 2003). In consequence the interpretation of the results can be difficult since genetic instability, even during in vitro culture, can lead to minor or major changes in profiles (Newell, Frost et al. 2000). Furthermore the discriminatory power of PFGE is dependent upon the restriction enzyme used (Nielsen, Engberg et al. 2000). A database and network (PulseNet) is available for this method for the rapid analysis and comparison of banding patterns (Fitzgerald, Helsel et al. 2001).

Multilocus Sequence Typing (MLST)

For this method, fragments of seven housekeeping genes are amplified by PCR, and the nucleotide sequence of these amplicons is determined. Nucleotide sequence extension reaction products are separated and detected on an automated DNA analyser. In a final step allele numbers and sequence types (ST) are assigned using the *Campylobacter* MLST database (Dingle, Colles et al. 2001; Colles, Jones et al. 2003), which is also used as an international data sharing platform, and offers analytical tools such as eBurst (Spratt, Hanage et al. 2004). The system of MLST was developed as an alternative to whole genome sequencing for epidemiological characterisation of pathogens of major public health importance (Dingle, Colles et al. 2001). Sequence typing by MLST is internationally emerging as the new gold standard for national and international epidemiological characterisation of major pathogenic microorganisms such as *C. jejuni* (Urwin and Maiden 2003; Sullivan, Diggle et al. 2005). MLST is an efficient and accurate isolate characterisation that is highly reproducible and its results can be shared via the Internet (Colles, Jones et al. 2003). There are however minor problems with typability, which require repeated analysis of individual genes. This technique has sufficient resolution to accommodate a high level of diversity present in a bacterial pathogen population, but can also rationalise this diversity into groups of isolates with related genotypes (Colles, Jones et al. 2003). It is particularly suited to long-term and global epidemiology as it identifies variation which is accumulating slowly within a population, but can be used to investigate individual outbreaks, especially when combined with other data from other loci such as the *fla*A Short Variable Region (SVR) and porA (MOMP) genes (Dingle, Colles et al. 2001; Dingle, Colles et al. 2005). Data generated can be applied to the investigation of population structure and evolutionary mechanisms (Dingle, Colles et al. 2001), as well as to determine the relative importance of the food chain — and its individual components — as sources of human *Campylobacter* infection (Dingle, Colles et al. 2002). The cost of reagents and sequencing make MLST relatively costly compared to other methods, however Dingle et al (Dingle, Colles et al. 2002) showed that nucleotide sequence-based techniques such as MLST can achieve high throughput and cost effectiveness. *C. coli* shows less diversity than *C. jejuni* at

each of the MLST loci, but diversity of the *fla*A SVR locus can allow closely related strains with the same MLST ST to be distinguished, providing sufficient resolution for the investigation of isolates from human sources and the potential sources of human infection (Dingle, Colles et al. 2005). This approach is related to the phenotyping approach MEE. However the greater number of alleles detected by MLST than by MEE implies that it has the potential to detect more subtypes than the number that can be detected by MEE (Sails, Swaminathan et al. 2003). MLST is less labour-intense, generally faster than MEE and amenable to automation (Sails, Swaminathan et al. 2003). It does not depend on band analysis, which, due to an difficult interpretation, makes results subject to misclassification (Newell, Frost et al. 2000).

Amplified Fragment Length Polymorphism (AFLP)

AFLP analysis is based on the selective amplification of restriction fragments generated from total genomic DNA (Duim, Godschalk et al. 2003). It produces multiple bands derived from all over the genome, which prevents over- or misinterpretation due to point mutations or single locus recombination (Duim, Wassenaar et al. 1999). Data generated can be analysed via an automated sequencer and can therefore be compiled, compared and exchanged (Duim, Wassenaar et al. 1999). To fulfil the needs for a method that combines high resolution, high throughput and simple reliable data analysis AFLP has been adjusted for *Campylobacter* spp. (De Boer, Duim et al. 2000). It has the potential to distinguish genetically unrelated strains from genetically related strains, (related strains can be distinguished by using other genotyping methods) and can be used as primary method for subtyping large numbers of *Campylobacter* strains and therefore can be applied in epidemiological investigations (Duim, Wassenaar et al. 1999). AFLP can easily be automated, which allows for standardisation and high-throughput for epidemiological investigations and provided a high degree of discrimination and identification of *C. jejuni* and *coli* strains (Duim, Wassenaar et al. 1999). However this method is relatively elaborate and requires purification or intact double stranded DNA and specialised equipment and software (Duim, Wassenaar et al. 1999). It is band based and shows minor differences in fingerprints of identical samples despite standardisation, which affect its reproducibility (Duim, Wassenaar et al. 1999). At the moment no international database for the comparison of patterns is available.

*fla*A and *fla*B RFLP typing

fla typing is based on only one locus of the genome (De Boer, Duim et al. 2000), the flagellin subunit of the flagellar filament, which is encoded by two highly homogenous tandem genes, *fla*A and *fla*B (Harrington, Thomson-Carter et al. 1997). The suitability of the highly conserved *fla* genes for PCR-RFLP analysis has been recognised by several groups (Newell, Frost et al. 2000). It is an inexpensive, quick, and

easy to perform method (De Boer, Duim et al. 2000; Fitzgerald, Helsel et al. 2001), which is suitable for the initial grouping of isolates in a surveillance situation (Petersen and On 2000). However the use of a single genetic locus as a tool for epidemiological investigations requires caution (Fitzgerald, Helsel et al. 2001). Strong evidence was found supporting recombination between *fla*A genes of different strains (intergenomic recombination) as well as with *fla*B (intragenomic recombination) (Harrington, Thomson-Carter et al. 1997; Dingle, Colles et al. 2005). This recombination alters the flagellin sequence while leaving the pathogen otherwise genotypically unaltered. Thus in the long term, relationships determined by flagellin gene typing will not accurately reflect true clonal relationships, as determined by overall genotypic similarity methods (Harrington, Thomson-Carter et al. 1997), and therefore it can not be considered a stable method for long-term monitoring of pathogenic *Campylobacter* populations (Harrington, Thomson-Carter et al. 1997). *fla*A amino acid sequence shows poor correlation with MLST clonal complexes, further indicating its instability for epidemiological typing in the absence of other data (Duim, Godschalk et al. 2003). *fla*A PCR-RFLPs are useful in terms of analysing isolates involved in an outbreak situation, especially if nucleotide sequence data is generated, however at a minimum this method should be coupled with other genetic markers to confirm results (Klena 2001). Newell et al. conclude that because of its above mentioned instability for accurate analysis *fla*A needs to be supported by alternative phenotypic or genotypic results (2000). The PubMLST website now includes a database for *fla*A (and *por*A). According to Mellmann et al. (Mellmann, Mosters et al. 2004) *fla*B typing is a rapid, reproducible, discriminatory and stable screening tool. It is more stable than *fla*A, probably due to less selective pressure (Mellmann, Mosters et al. 2004).

Sequence based subtyping of short variable region (SVR)

To overcome problems with the standardisation of patterns of *fla*A PCR-RFLP a relatively simple DNA sequence analysis of the short variable region (SVR), a sequence near the end of *fla*A, was developed (Newell, Frost et al. 2000). This approach provides a level of discrimination similar to the entire *fla*A sequence (Fitzgerald, Helsel et al. 2001). It is useful for rapid, preliminary strain characterisation to establish an epidemiological link in a well defined setting and could correctly differentiate an outbreak and a sporadic strain (Fitzgerald, Helsel et al. 2001). The diversity of this locus can allow closely related strains with the same MLST sequence type to be distinguished (Dingle, Colles et al. 2005). However it is less applicable for smaller laboratories due to the costs of the sequencer (Fitzgerald, Helsel et al. 2001). Additionally this approach cannot distinguish between *C. jejuni* and *C. coli* because of a common gene pool for *fla*A shared by theses species, due to frequent interspecies recombination. Therefore its use in epidemiological studies can be questioned and it is unsuitable, when used alone,

as a marker for the molecular epidemiology of *Campylobacter* spp. (Dingle, Colles et al. 2005).

Ribotyping

The detection of ribosomal DNA (rDNA) cistrons, using a labelled probe, within restriction endonuclease profiles of test DNA isolates is known as ribotyping and the resultant profile is considered a ribopattern (Klena 2001). It is an automated high throughput process, adapted for data exchange as it uses standardised material, and method and results therefore are comparable (De Boer, Duim et al. 2000). Disadvantages of this application are the high costs, low discriminatory power and that it does only allow with difficulties to interfere with its automated settings and interferences. Additionally the Riboprinter used is not capable of cluster analysis (De Boer, Duim et al. 2000). The substantial equipment and consumable cost and small sample throughput is likely to restrict the use of this technology (Newell, Frost et al. 2000).

Randomly amplified polymorphic DNA (RAPD)

RAPD is a rapid typing method based upon randomly amplified polymorphic DNA (RAPD), with the entire genome potentially accessible to priming and amplification (Hilton, Mortiboy et al. 1997). This method is similar to AFLP, which selectively amplifies restriction fragments generated from total genomic DNA. It has a high discriminatory power (diversity index of 0.999 indicating that if two strains were sampled randomly from the *Campylobacter* population, on 99.9% of occasions they would fall into different RAPD types (Hilton, Mortiboy et al. 1997)). When used on *C. coli* and *C. jejuni* the method was highly, but not excessively discriminatory; identical types were detected and all isolates were typeable (Madden, Moran et al. 1996). Excellent typeability and reproducibility coupled with relative technical ease as reported by Hilton et al. (1997) make it a valuable tool for classification of *Campylobacter* isolates suitable for epidemiological investigation. But there have also been concerns about this approach. According to Newell et al. (2000) significant reproducibility problems have largely limited its widespread use. Additionally RAPD may be too sensitive for the study of *C. jejuni* isolates under some circumstances, and therefore it is recommended to use it only in conjunction with other genotyping methods (Newell, Frost et al. 2000).

DNA microarray

In DNA-DNA hybridisation methods, which measure the overall genetic relatedness of two isolates, DNA from the isolate under investigation is denatured and hybridised to denatured DNA isolates from a set of known standard and reference strains (Klena 2001). DNA microarrays are a highly specific epidemiological tool for the analysis of *C. jejuni* isolates and reveal both divergent and highly conserved gene classes among

isolates (Leonard, Takata et al. 2003). It is a method of comparison that is more gene specific than other existing genotyping methods and therefore can reveal genomic differences between isolates that, for example, RAPD analysis may not identify (Leonard, Takata et al. 2003). This approach provides a highly detailed, whole-genome fingerprint of *C. jejuni* isolates and is, thus, a powerful addition to the repertoire of genotyping tools available for this organism (Leonard, Takata et al. 2003). Recently genomotyping, which uses DNA-microarrays for whole genome comparisons was applied to model the phylogeny of *C. jejuni* (Champion, Gaunt et al. 2005). However this technique cannot identify differences between isolates that are due to small nucleotide changes, horizontal acquisition or loss of genetic elements distinct from those in the strain used to generate the microarray (Leonard, Takata et al. 2003). In addition it can be classified as a specialist method, that requires a relative great deal of time and labour when compared to other methods, and is not yet optimised as a high throughput routine technique (Klena 2001).

ERIC fingerprint analysis

This fingerprinting technique is based on the amplification of enterobacterial repetitive intergenic consensus (ERIC) DNA sequences (Moser, Lentzsch et al. 2002). Repetitive ERIC sequences provide a useful target for epidemiological analysis, since they are present in multiple copies at different positions in the genome and therefore permit simultaneous scanning for DNA polymorphisms at multiple genome loci (Moser, Lentzsch et al. 2002). When applied to *Campylobacter* spp. by Moser et al. (2002) this technique was the most discriminatory method to differentiate between 'human' and 'non-human' clusters when used in comparison with PFGE and Penner typing (Moser, Lentzsch et al. 2002). Although less reproducible than PFGE typing ERIC primed PCR can be used as a simple and rapid tool to discriminate different strains of *C. jejuni* (Shi, Liu et al. 1996).

CRISPRs sequencing

Sequence analysis of a genomic region with short tandem repeats designated clustered regularly interspaced short palindromic repeats (CRISPRs) may be useful for subtyping of strains with common AFLP or MLST profiles, but was not considered the method of choice as a large number of isolates was found to be non-typeable (Schouls, Reulen et al. 2003). However this approach has recently been successfully applied to typing and comparative analyses of streptococcus strains (Horvath, Romero et al. 2008) and CRISPRs have been identified as a system of acquired phage resistance with widespread presence and extreme diversity in prokaryotes which has been applied to type several pathogens (Sorek, Kunin et al. 2008).

PCR assays

PCR multiplex assays can be used to differentiate between *Campylobacter* species (Asakura, Samosornsuk et al. 2008) and offer a more sophisticated approach of simultaneous identification of *Campylobacter* species than the separate identification of each species (Yamazaki-Matsune, Taguchi et al. 2007). Using multiple PCR assays on potential pathogenicity and virulence markers may provide an opportunity to further subdivide *Campylobacter* isolates based on their ability to cause disease. A PCR-based approach to binary genomotyping, assessing gene complement by scoring pre-selected LOS, FLAG and CPS genes as either PCR positive ('present') or PCR negative ('absent') was recently developed at the Hopkirk Research Institute (French 2009). However at this stage this is not a routinely applied method which can be used for surveillance or screening but rather a specialised tool to investigate population genetics, and in particular host adaptation, of *Campylobacter* spp.

Summary of genotyping methods

Campylobacter spp. have a high genetic and antigenetic diversity (Colles, Jones et al. 2003) and there is so far no globally accepted gold standard to overcome the associated problems when typing this pathogen. Recombination is a source of genetic diversity in pathogens, and perhaps even one of the major mechanisms. Substantial evidence for recombination in *Campylobacter* spp. was found and estimated to be of similar magnitude to the rate of mutation (Fearnhead, Smith et al. 2005; Wilson, Gabriel et al. 2009).

A wide variety of typing techniques is available and examples were presented earlier in this chapter. Generally speaking genotyping methods vary between their main targets. Some of the methods focus on single loci (for example *fla*A-typing) while others look at the whole genome of the isolate under investigation (for example PFGE). The investigation of nucleotide sequences of genes under diversifying selection are useful when discriminating between related isolates, for example in distinguishing outbreak strains (Dingle, Colles et al. 2002) and several single genes have been used to type *Campylobacter* strains. When assessing these methods it should be taken into account that in general genes that encode outer membrane proteins show greater sequence diversity than those that encode housekeeping genes. This suggests that those genes are under selective pressure to undergo frequent changes (Riley 2004) and are therefore less suitable for long-term epidemiologic investigations. Therefore methods based on a single locus (for example *fla*A) may provide inadequate results due to recombination, whilst methods producing complete genomic fingerprints tend to be more reliable (Duim, Wassenaar et al. 1999).

Genotyping methods can further be divided into those based on sequencing loci of interest (for example MLST) or those using band-based methods such as DNA finger-

printing techniques (for example PFGE, AFLP). The detection of RFLP is a relatively simple and reproducible technique which has been useful in the epidemiological characterisation of many pathogenic microorganisms (Jackson, Fox et al. 1996). However techniques such as RAPD analysis or PFGE, although more accurate than other typing techniques, are based on band patterns of restriction endonuclease fragments and thus provide limited information about the whole genome (Leonard, Takata et al. 2003). Another drawback of band-based method is the possibility of misclassification due to the sometimes ambiguous interpretation of the banding patterns (Newell, Frost et al. 2000). Recently the increasing availability of automated DNA sequencing technology has led to the development of molecular subtyping methods based on the sequencing of one or more gene loci (Sails, Swaminathan et al. 2003). As this approach is not based on band matching, sequencing is a more precise measure of genetic variability (Fitzgerald, Helsel et al. 2001). Additionally, DNA sequencing has the advantage of minimal experimental variation in the results, which facilitates direct inter-laboratory comparisons (Sails, Swaminathan et al. 2003).

Many studies have compared the capabilities of different methods using them on the same or similar isolates, and a variety of conclusions have been made. As an example Mellmann et al. (2004) concluded that combining methods gives better discrimination and identification and, considering the influence of possible recombination events, the use of multiple genotyping methods is recommended. PFGE was the most discriminatory technique in their study when compared to *fla*A, *fla*B typing and MLST. In this study MLST was less concordant to PFGE than *fla*A and *fla*B typing, but still with a high overall concordance of 94%. *fla*A typing was over discriminatory when compared to the other techniques (Mellmann, Mosters et al. 2004). AFLP is supported for optimal epidemiological typing of *Campylobacter* by de Boer et al. and its discriminatory power was greatest when using *fla*A typing and ribotyping on the same isolates (De Boer, Duim et al. 2000). When comparing PFGE, Ribotyping, MLST and AFLP they show the same value for epidemiological typing and similar grouping were obtained by all methods. MLST and AFLP appear to have the highest degree of congruency (Duim, Godschalk et al. 2003). Genomic RAPD and PFGE analysis, together with serotyping techniques, have been used to subtype *C. jejuni* isolates in clinical outbreaks. One limitation of these techniques is that only a very small part of the entire genome is analysed (Leonard, Takata et al. 2003). A combination of MLST and SVR sequencing of the *fla*A flagellin gene allows resolution equivalent for that of PFGE for outbreak investigation was concluded by Dingle et al. (Dingle, Colles et al. 2005). PFGE was found to be the most discriminatory technique when compared to somatic O serotyping, RFLP, *fla*A analysis and DNA sequence analysis of *fla*A including SVR (short variable region) and the entire *fla*A gene. In the same study sequence based *fla*A SVR subtyping was found to be more discriminatory than RFLP (Fitzger-

ald, Helsel et al. 2001). Nielsen et al. concluded that PFGE and RAPD are highly discriminatory methods, based on the whole genome, and these methods are therefore useful for ensuring genotypic similarity in cases of outbreaks (Nielsen, Engberg et al. 2000). In a New Zealand study indistinguishable PFGE patterns were found to have different MLST STs; however PFGE pattern was found to be was good indicator of MLST clonal complex (CC) (Carter 2007). What emerges from these different studies is, that no single technique is universally applicable, and that the choice of typing tool is related to the question posed and the convenience of the technique (Isenberg 1998).

2.3.3 A comparison of phenotyping and genotyping

Genotyping methods have important advantages over phenotypic methods. They measure relatively stable chromosomal differences, whereas phenotyping methods measure characteristics such as enzyme or antigens which may not be stably expressed (De Boer, Duim et al. 2000; Klena 2001). Phenotypic systems are less directly correlated with clonal descent because of the possibility that they are skewed by variable marker expression and evolutionary convergence (Struelens and Members of the European Study Group on Epidemiological Markers (ESGEM) of the European Society for Clinical Microbiology and Infectious Diseases (ESCMID) 1996). Likewise typing systems exploring polymorphisms at multiple sites of the whole genome are more representative than typing systems exploring variation at a single gene locus (De Boer, Duim et al. 2000). In general analysis of variation of cell surface components (for example heat-stable HS serotyping antigen, flagella protein *fla*A short variable region (SVR)) does not provide epidemiological information in *Campylobacter* spp. of the same quality as in, for example, Salmonella spp. due to high frequency of genetic exchange among isolates (Dingle, Colles et al. 2002). A second advantage provided by some, but not all, genotyping methods is that the need for time-consuming culturing can be circumvented (Klena 2001).

However, phenotypic methods, in particular serotyping systems, are used in laboratories worldwide, for example for surveillance of large numbers of isolates (Nielsen, Engberg et al. 2000) and are often still more available and applicable than genotyping methods (Duim, Wassenaar et al. 1999). For improved discrimination of isolates one or more of the genotypic methods are usually selected (Nielsen, Engberg et al. 2000). These molecular biological methods offer greater potential for differentiating between strains, even among isolates who share the same surface markers (for example serogroup or phage type determinants) (Jackson, Fox et al. 1996). For example the development of genotypic fingerprinting methods for *Campylobacter* epidemiology has increased discrimination as well as overcome problems (particularly non-typeability) associated with established phenotyping methods (Jackson, Fox et al. 1996). Similarly phenotypic traits alone can mark stable cell lineages or clonality, but can also obscure

the true genetic relationships between strains, particularly where surface determinants are labile, as with somatic and flagellar antigens, and phage absorption sites (Jackson, Fox et al. 1996). In contrast to the uncertainties of characterisation based on variable phenotypic characteristics, the application of molecular typing methods can provide a stable and highly discriminatory analysis of bacterial isolates (Hilton, Mortiboy et al. 1997). Given the sheer number of *Campylobacter* strains encountered by some reference laboratories, cost and ease of use are of prime importance. In this case, phenotypic methods used to be the methods of choice, while genetically based methods were mainly used for research purposes (Newell, Frost et al. 2000).

In summary genotyping offers greater capacity for differentiating strains than phenotyping and can be useful in making phylogenetic as well as epidemiological inferences (Fitzgerald, Helsel et al. 2001). In addition these techniques are becoming more and more available even for routine diagnosis. The problems encountered with untypeable reactions in phenotyping methods make genetic methods much more appealing as candidates for the production and standardisation of data which are comparable around the world (Wareing, Bolton et al. 2002).

2.4 MRA approaches

2.4.1 Overview

Risk assessment (RA) is the science-based component of risk analysis, which also includes risk management and risk communication as further steps (FAO/WHO 2002) (Figure 2.1). It provides a framework for organizing data and information in a structured manner and thereby allows us to understand better the interaction between hazards, foods and human illness (Cahill 2005). Microbiological risk assessment (MRA) is commonly concerned with food safety risks, and involves estimation of the magnitude of microbial exposure at various stages in the food production chain to estimate the risk of food borne infections (Thrusfield 2005). It has become a key feature of food safety management and is regularly used to guide policy (Kelly, Hartnett et al. 2003). MRA is evolving within both food safety regulatory agencies and academia in an ever increasing number of countries and is recognised as a resource-intensive task requiring a multidisciplinary approach (Cahill 2005).

The FAO/WHO Codex Alimentarius Commission (CAC) defines risk assessment as a scientifically process based on four steps (Codex Alimentarius Commission (CAC) 1999):

- **Hazard** identification, which is the identification of the biological agent that may be present in a particular food or group of foods and capable of causing adverse health effects.

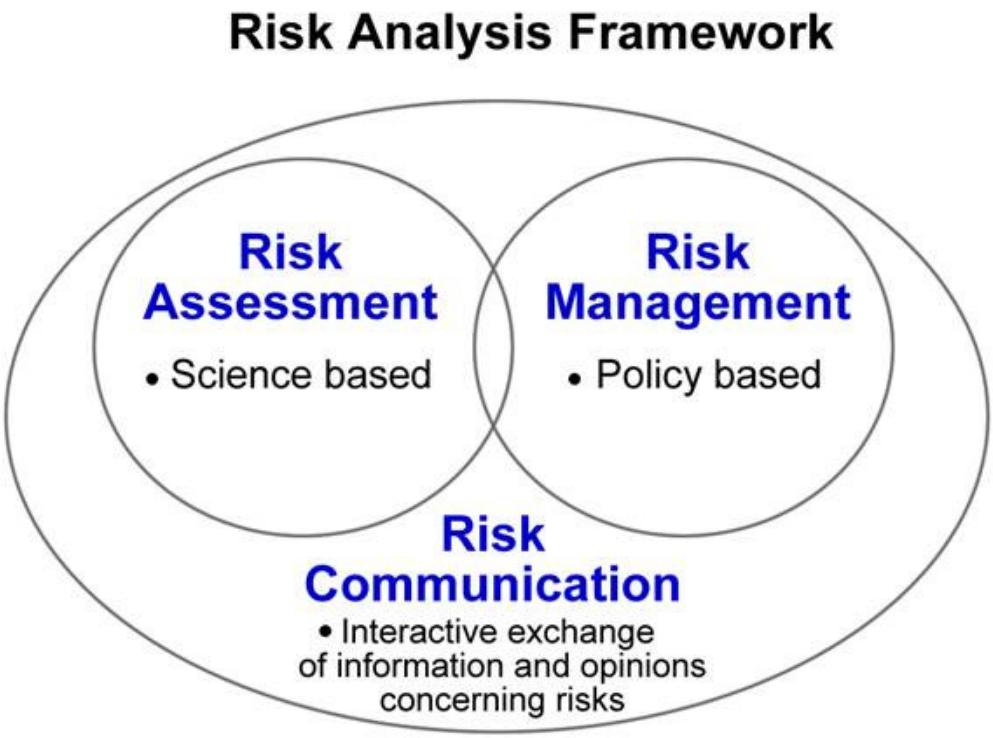

Figure 2.1: The Risk Analysis Framework (World Health Organisation 2007)

- **Hazard characterization**, which is the qualitative or quantitative, or both, evaluation of the nature of the adverse health effects associated with the biological agents that may be present in food, and in such cases a dose-response assessment should be performed if the data are obtainable.

- **Exposure assessment**, which is the qualitative or quantitative, or both, evaluation of the likely intake of the biological agent through food, as well as through exposure from other sources, if relevant.

- **Risk characterisation**, namely the qualitative or quantitative, or both, estimation, including attendant uncertainties, of the probability of occurrence and severity of known or potential adverse health effects in a given population based on hazard identification, hazard characterization and exposure assessment.

Conceptually the risk assessment approach starts from the dynamics of the hazard in the food chain and uses predictive models to estimate the outcome in terms of public health. It thereby can provide valuable information on the complex dynamics of pathogens during food processing (Havelaar, Braeunig et al. 2007).

MRAs are increasingly used to understand how pathogens are propagated along the food chain and to improve the quality and safety of food. With clear documentation

MRA provides several advantages: scientific justification for actions, a means of demonstrating equivalence as well as an effective communication tool (Cahill 2005). This approach is particularly useful as by describing the system in mathematical models the changes in model output when certain interventions are implemented can be examined (Havelaar, Braeunig et al. 2007). In addition the increased insight in the propagation of hazardous microorganism in the food chain helps to identify more targeted control strategies. Finally, MRAs help to identify data gaps, and thereby facilitate the identification of research needs, the establishment of research priorities and the design of commissioned studies (Cahill 2005). If exposure estimates and dose-response functions are sufficiently accurate, risk assessments may provide excellent estimates of the true impact of illness (Batz, Doyle et al. 2005).

To date risk assessments on several food pathogen combinations have been successfully completed and support decision making in food safety. However while MRA is becoming an important tool for assessing risk from food borne pathogens, often it is not within the capacity of individual countries to complete a quantitative MRA as it is a resource-intensive and multidisciplinary task (FAO/WHO 2002). Because of this MRAs have been undertaken for only a limited number of pathogen-food combinations to date (Batz, Doyle et al. 2005). However MRAs are increasingly used for animal food products in particular as a result of the adoption of Sanitary and Phytosanitary (SPS) agreement by WTO member states in 1995. This agreement requires all member states to base all food safety regulatory measures on sound scientific risk assessment (Kelly, Hartnett et al. 2003).

Most quantitative RA models deal with one pathogen occurring in a single food commodity (Havelaar, Braeunig et al. 2007), for example *Campylobacter* spp in chicken (Hartnett, Kelly et al. 2001; Rosenquist, Nielsen et al. 2003), and are often aimed to identify options for prevention, intervention and control (Havelaar, Braeunig et al. 2007). Amongst others the risk of infection from Listeria in soft cheese (Bemrah, Sanaa et al. 1998), Salmonella in eggs and broiler chicken (FAO/WHO 2002) and *E. coli* in steak tartar (Nauta, Evers et al. 2001) has been analysed. A number of different model-based approaches have evolved and been applied in microbial food safety. Factors such as scope, available time, data resources, and the skill and expertise of the risk assessor will all have an impact on the approach taken (Cahill 2005). Table 2.1 gives a schematic overview of different approaches to MRA.

Source attribution has recently evolved as a novel approach to microbial risk assessment and a variety of tools and methods are becoming available. Although included in this overview of approaches to MRA these techniques are in more detail described and discussed in section 2.4.4 of this chapter.

Common to all risk assessment is the need for representative data. Therefore in general most risk assessment approaches are resource-intensive and in the face of large

data gaps may not be practical or possible at all. In particular source attribution methods, that rely on pathogen subtyping results, are dependent on the availability of good quality data on a high level of resolution. As in all epidemiological studies, bias is of major concern for the validity of results (Rothman and Greenland 1998) and should carefully be considered.

2.4.2 Different approaches to MRA

Risk profiles

A risk profile is a document that provides a summary of relevant information on a specific food safety issue. The purpose of a risk profile is to provide contextual and background information relevant to a food/hazard combination, so that risk managers can make decisions and, if necessary, take further actions (for example commission a full risk assessment).

Covello Merkhofer and Codex Alimentarius Commission approaches

The Covello-Merkhofer system builds the base for risk assessments within an World Organisation for Animal Health (OIE) International Animal Health Code risk analysis framework (Vose, Acar et al. 2001). Based on this approach a risk assessment model consists of four interrelated but conceptually distinct steps developed by Covello and Merkhofer in 1993 (1993). These steps include: Release assessment, Exposure assessment, Consequence assessment and Risk estimation, which are preceded by a hazard identification process (Covello and Merkhofer 1993). The OIE Terrestrial Animal Health code (Office Internationale des Epizooties (OIE) 2005) (http://www.oie.int/eng/publicat/encode.htm) describes the individual steps as follows:

- **Release assessment** consists of describing the biological pathways necessary for an importation activity to release pathogenic agents into a particular environment, and estimating the probability of that complete process occurring, either qualitatively or quantitatively.
- **Exposure assessment** consists of describing the biological pathways necessary for exposure of animals and humans in the importing country to the hazard released from a given risk source, and estimating the probability of the exposures occurring, either qualitatively or quantitatively.
- **Consequence assessment** is defined to consist of describing the relationship between specified exposures to a biological agent and the consequences of those exposures.

Table 2.1: Systematic classification of different approaches to MRA.

Categorized by	Types of approaches	
Scope	Risk profile	(Full) risk assessment
Legislative context	Codex Alimentarius Commission	Covello Merkhofer
Statistical approach	Qualitative RA	Quantitative RA
Statistical approach	Deterministic RA	Probabilistic RA
Statistical approach	Frequentist	Bayesian
Level of detail	Simple Exposure model	Food chain mechanistic model
Step in food production	Primary production	Processing and retail
		Consumption
Purpose of RA	Farm to fork RA	Source attribution models

- The final **risk estimation step**, according to these guidelines, consists of integrating the results from the previous steps to produce overall measures of risk associated with the hazards identified at the outset.

The Codex Alimentarius Commission (CAC, http://www.codexalimentarius.net/) approach has been developed for food safety issues and is based on the United States National Academy of Science (US-NAS) system (Vose, Acar et al. 2001). Principles and Guidelines for the Conduct of Microbiological Risk Assessment have been developed by this group (Codex Alimentarius Commission (CAC) 1999) to assist in the management of food safety hazards while minimising trade disruptions (Murray 2002).The risk assessment part of this risk analysis approach can be divided into four major steps: **hazard identification**, **exposure assessment**, **hazard characterisation** and **risk characterisation** (Murray 2002), which are further defined by the CAC Principles and Guidelines as follows (Codex Alimentarius Commission (CAC) 1999) and have been presented earlier on in this chapter. For microbial agents, the purpose of hazard identification is to identify the microorganisms or the microbial toxins of concern with food. Exposure assessment includes an assessment of the extent of actual or anticipated human exposure. The hazard characterisation step provides a qualitative or quantitative description of the severity and duration of adverse effects that may result from the ingestion of a microorganism or its toxin in food. A dose-response assessment should be performed if the data are obtainable. Risk characterization represents the integration of the hazard identification, exposure assessment and hazard characterization, determinations to obtain a risk estimate; providing a qualitative or quantitative estimate of the likelihood and severity of the adverse effects which could occur in a given population, including a description of the uncertainties associated with these estimates. Vose et al. (2001), provide a more detailed description of the differences between the Covello Merkhofer and CAC methods.

Qualitative and quantitative approaches

Qualitative risk assessment is a descriptive treatment of information and the risk estimate generated will generally be of a categorical nature such as for example high, medium, low (Cahill 2005). It is particularly useful when resources are limited and can be used as a first level of evaluation. This technique is suitable for the majority of risk assessments and is, in fact, the most common type of assessment undertaken to support routine decision-making in an import risk analysis context (Murray 2002). Completed models in food safety include an exposure assessment of *E. coli* O157 in pasteurized milk (Clough, Clancy et al. 2006) and an assessment of the human health impact of Salmonella in imported soybean products (Hald, Wingstrand et al. 2006). **Quantitative risk assessment** is a mathematical analysis of numerical data and can either be deterministic or probabilistic (Cahill 2005). Quantification involves devel-

oping a mathematical model to link various aspects of the epidemiology of a disease, with inputs and outputs being expressed numerically (Murray 2002; Clough, Clancy et al. 2009). For quantitative models in MRA, exposure assessment requires data for pathogen occurrence and distribution in foods and live animals, parameters for growth and decline and consumption (Coleman and Marks 1999). In some circumstances it might be more desirable to undertake a quantitative risk assessment, as this allows for further insight into a particular problem and helps to identify critical steps or to compare sanitary measures (Murray 2002). However constraints in data quality, time, personal or resources may not permit this (Coleman and Marks 1999).

Deterministic and probabilistic approaches

Single point or **deterministic** models use a single "best guess" estimate of each variable within a model to determine the model's outcome(s) (Vose 2000). This approach suffers from important drawbacks. If as an example three points estimates for a variable are used: minimum, best guess and maximum value, no recognition is being given to the fact that that for example the best guess value is much more likely to occur than the minimum or maximum values. Additionally only these values are being used, when in fact the variable could take on any value (Vose 2000). The structure of a **probabilistic** model is very similar to a deterministic model, with all the multiplications, additions, etc. that link the variables together, except that each variable is represented by a probability function instead of a single value (Vose 2000). We can then calculate the combined impact of variation in each of the model's input distributions to determine a probability distribution of the possible model outcomes, which can be done by using a simulation (for example Monte Carlo Simulation) (Murray 2002). Examples of deterministic risk assessments can be found more readily in the area of chemical risk or safety assessment while most of the best known examples in food safety are probabilistic (Cahill 2005).

Classical (frequentist) and Bayesian methods

A number of traditional statistical frequentist techniques are available for quantifying parameters under certain assumptions, which are often considered to be exact techniques, given the assumptions made are correct (Vose 2000). According to classical theory, probabilities are numbers associated with events and the probability of an event occurring is the value to which the long-run frequency would converge as the number of experiments increases towards infinity (Covello and Merkhofer 1993). Many major risk assessments have successfully applied this technique (Cassin, Lammerding et al. 1998; Hartnett, Kelly et al. 2001; Nauta, Evers et al. 2001; FAO/WHO 2002). On the other hand from a Bayesian perspective probability is not only a function of the event, but of the state of information (Covello and Merkhofer 1993) and mathemat-

ically describes the learning process (Vose 2000). Bayesian inference is an extremely powerful technique, based on Bayes' theorem, for using data to improve one's estimate of a parameter (Vose 2000). Bayesian risk assessment methods tend to make explicit use of expert judgement (Covello and Merkhofer 1993) and the approach can be seen as a means of transforming prior into posterior opinions, rather than producing the posterior distribution (Spiegelhalter, Abrams et al. 2004). Barker et al. (2002) illustrate the benefits associated with a Bayesian view of food safety assessment by a Bayesian belief network approach for *Clostridium botulinum*. Additional example are a Bayesian approach to estimate the risk of human Salmonellosis from several food sources (Hald, Vose et al. 2004) and the use of this technique for a risk assessment of *Listeria monocytogenes* in a salmon product (Delignette-Muller, Cornu et al. 2006). Although the classical and Bayesian perspective represent polar points of view, the needs and constraints of applications often require a perspective that lies somewhere between both extremes (Covello and Merkhofer 1993). To date most risk assessments are based on a frequentist approach, however this is rapidly changing with Bayesian techniques becoming more widely available. Please find a detailed description of Bayesian approaches in Appendix D)

Food chain mechanistic and simple exposure models

Food chain mechanistic models, such as the modular process risk model (MPRM) developed by Nauta et al. (2001; 2001; 2005), allow an assessment of current health risks, but also offer the opportunity to compare the effects of risk management interventions proposed to reduce those risks. MPRM models provide a farm-to-fork risk assessment, which quantitatively describes the transmission of biological hazards through the food pathway (Nauta, Evers et al. 2001). However food chain mechanistic models do not necessarily have to assess the whole farm-to-fork continuum. Alternatively simple exposure models like the New Zealand *Campylobacter* exposure model (McBride, Meleason et al. 2005) can provide a quick and non resource intensive way of assessing the relative importance of different pathways. Since they tend to be rather generalized, specific interventions cannot be modelled using this approach.

Models focused on primary production, processing or consumption

Although the control and management of food safety risks should be based on an integrated approach that addresses all sectors (Sofos 2002), risk models on particular parts of the farm-to-fork continuum, for example consumption, can be a valuable tool in assessing food safety risks. For example measures taken on farm such as the reduction of pathogen incidence should reduce the likelihood of the contamination of an animal food product (Sofos 2002), whilst on the other hand consumer models can help identify valuable mitigation strategies in private households (Luber, Brynestad et al. 2006;

Lindqvist and Lindblad 2008).

Farm to fork models and source attribution models

Farm-to-fork approaches assess the risk form different stages of the food chain system and examine the practices and procedures in place. They model the whole process by which food reaches the consumer's table ('fork') from the primary processing stage ('farm'). These 'holistic' approaches are commonly applied worldwide and epidemiological intelligence is essential to any such approach (Schwabe 1994). In recent decades there has been a growing awareness of an integrated, multidisciplinary approach, considering the whole of the food chain (Slorach, Maijala et al. 2002) and farm-to-fork models are currently being used by all major food safety agencies worldwide. On the other hand the successful control of any zoonotic pathogen requires methods that are able to quantify the contribution of each of the possible sources (Havelaar, Braeunig et al. 2007). This will also support decision makers in identifying and prioritising effective food safety interventions. According to Batz et al. (Batz, Doyle et al. 2005) approaches to source attribution can be grouped into two broad categories, epidemiologic and microbiologic. Epidemiologic approaches are based on public health surveillance and include food borne outbreak data and case-control studies, whereas microbiologic approaches, such as the Hald model (Hald, Vose et al. 2004; Mullner, Jones et al. 2009), microbial source tracking (MST) methods and risk assessments of specific pathogens, rely on data on pathogen samples drawn from human, animals and food sources and include pathogen subtyping. Source attribution models are a valuable tool for priority setting but they can not be used to model interventions.

2.4.3 Major *Campylobacter* risk assessments to date

Risk assessments of *Campylobacter* spp. have recently been undertaken or are underway in several industrialised countries, as well as at the international level by FAO and WHO (Cahill 2005). The below presented risk assessments aim to improve the safety of the domestic food supply, determine appropriate, feasible and effective interventions and risk mitigation strategies as well as try to identify research needs and achieve compliance with international standards with regard to food trade. The risk assessments undertaken in the United Kingdom (Hartnett, Kelly et al. 2001; Hartnett, Kelly et al. 2002), Denmark (Rosenquist, Nielsen et al. 2003), New Zealand (Lake, Hudson et al. 2007) and The Netherlands (Nauta, van der Fels-Klerx et al. 2005) estimate the risk associated with *C. jejuni* in poultry meat from the farm to the point of consumption.

Denmark

The Danish risk model (Rosenquist, Nielsen et al. 2003) places great emphasis on post processing and how consumer behaviour, age and sex impacts on the risk of acquiring

campylobacteriosis from poultry meat.

The Netherlands

This risk assessment is part of the *Campylobacter* Risk Management and Assessment (CARMA) project (http://www.rivm.nl/carma/indexeng.html). The CARMA project started in 2001 and also includes economic analysis and research on stakeholder perceptions of the problem and possible intervention. As part of this project Nauta et al. (2005) have developed a MPRM model for poultry processing.

Switzerland

In Switzerland a PhD project (Wieland, Wittwer et al. 2006) has assessed the significance of different exposure pathways for campylobacteriosis using genotyping data.

Sweden

A recent Swedish risk assessment (Lindqvist and Lindblad 2008) evaluates risk management strategies and the frequency of consumer mishandling, with the probability of illness per handling if the chicken was mishandled as model output.

United Kingdom

The risk assessment by Hartnett (2001; 2002) paid particular attention to the transmission of *Campylobacter* spp. among and within flocks at the farm level and among birds at the transport stage. In addition genetic risk attribution approaches were applied to data from Scotland and Lancashire to apportion cases of human disease (Wilson, Gabriel et al. 2008; Sheppard, Dallas et al. 2009).

Germany

The German risk assessment (Federal Institute for Risk Assessment BFR 2005; Luber, Brynestad et al. 2006) is focused on the consumer part of the farm-to-fork continuum. It assesses the risk from consuming chicken pieces and specifically models cross-contamination during handling in the kitchen.

WHO

By necessity risk assessment undertaken at the international level adopts a broader approach than the country specific assessments (Cahill 2005). The aims of the risk assessment work by FAO/WHO (FAO/WHO; FAO/WHO Global Forum of Food Safety Regulators - GF CRD Denmark-1 2002) were to provide estimates for the risk of exposure and illness and the change in risk likely to occur following different interventions in primary production, processing and food handling, and secondly to generate a toolbox and examples for countries to use in their own risk assessments (Cahill 2005). This

model is a farm to fork approach and modular in nature. Relying heavily on the work that has been generated at the national level, it cannot be considered representative of all production systems. However it attempts to understand how the incidence of human campylobacteriosis is influenced by various factors while trying to enable the consideration of the broadest range of intervention strategies.

New Zealand

The New Zealand Food Safety Authority (NZFSA) has commissioned several risk profiles for *Campylobacter* in poultry, offal and red meat (Lake, Hudson et al. 2003; Lake, Hudson et al. 2007; Lake, Hudson et al. 2007). In addition a document titled "Transmission routes for campylobacteriosis in New Zealand" (Lake 2006) tries to identify the identity and relative importance of different transmission routes for *Campylobacter* infection in New Zealand. Recently a quantitative risk assessment model (Lake, Hudson et al. 2007) was developed to investigate *Campylobacter* spp. contamination in the processing and consumption stages of the New Zealand poultry food chain. The output of the model is intended to describe the exposure of consumers, in terms of probability that an exposure will be contaminated, and if so, the number of bacteria involved. The purpose of this model is to assess the effect of changes in the poultry food chain on that exposure. The modelling group of the Enteric Zoonosis Disease Research Group (EZDRG) has develop a simple exposure model, which estimates the relative importance of four commonly identified sources of human campylobacteriosis (McBride, Meleason et al. 2005). In addition a model commissioned by the Ministry for the Environment estimates the proportion of *Campylobacter* cases that is attributable to freshwater contact recreation (McBride, Till et al. 2002). Cases of human disease could recently be apportioned in this country using a modified Hald model (Mullner, Jones et al. 2009), which is described in Chapter 5 of this thesis.

2.4.4 Source attribution approaches

The capacity to attribute cases of food borne disease to a food vehicle or another source responsible for illness is critical for the identification and prioritization of food safety interventions and a variety of approaches to source attribution are used worldwide (Batz, Doyle et al. 2005). In addition the recognition of effective food safety intervention requires an understanding of the relationship between food and pathogen from farm to consumption. The assessment of the origin of human infection is a major global public health issue and classical epidemiological approaches such as case-control studies (Eberhart-Phillips, Walker et al. 1997) have been conducted to fulfil 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. For example a case-control study conducted by Eberhart-

Phillips et al (Eberhart-Phillips, Walker et al. 1997) showed an association between the risk of campylobacteriosis and recent consumption of raw or undercooked chicken and chicken consumed at restaurants. However recent consumption of baked or roasted chicken was protective in another survey (Adak, Cowden et al. 1995).

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 (Greenland and Rothman 1998). In contrast to an outbreak situation, where the attributable fraction for an identified risk factor would be very high, the source of infection in sporadic cases is more difficult to identify (Riley 2004). In addition different species of bacteria can have a different epidemiology, as illustrated for *C. jejuni* and *C. coli* (Gillespie, O'Brien et al. 2002). 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 and power; seeking to associate statistically very small risks relative to the background, in particular when analysing sporadic cases of endemic diseases, may limit the utility of these studies. This is why the study of endemic disease risks becomes much more difficult, especially when the epidemiology of disease is complex, including a large amount of risk factors and potential sources. According to Haas et al. (1999) the use of epidemiology alone without other scientific fields integrated into the process will not fulfil the needs for a complete risk assessment.

Many recent advances have been made in the field of molecular epidemiology which have been a major driver of the recent developments in the area of source attribution. The combination of knowledge of the molecular epidemiology, for example the relative occurrence of *Campylobacter* subtypes in particular sources, with modelling and statistical tools provides us with an opportunity to gain a better understanding of the relative importance of individual pathways to the burden of human disease. Source attribution is a valuable tool for decision makers and many techniques, ranging from very simple spreadsheet models, like the Dutch model (Hald 2002), to complex models implemented in special software, like the island model (Wilson, Gabriel et al. 2008), are now available. Since each method is based on several assumptions that simplify the complexity of the epidemiology of *Campylobacter*, all methods have their limitations and conclusions need to be carefully discussed.

Several molecular epidemiological models and tools are currently being used to assess the relative contribution of a variety of sources to the burden of human disease. They constitute two different approaches; they are either based on a comparison of the relative frequency and epidemiology of pathogen subtypes in different source, or try to make inferences based on the population genetics of a pathogen. The Hald and the

Dutch model (Hald, Vose et al. 2004) as well as the proportional similarity Index (PSI) (Rosef, Kapperud et al. 1985) are examples of approaches which rely on the differences in the occurrence of bacterial subtypes in different sources. The general assumption behind this is that the more similar the distribution of the bacterial subtypes between two sources (for example human cases and chicken) the more likely transmission occurs between those sources. In addition the Hald model includes epidemiological aspects such as differences in the ability of individual sources to cause disease. On the other hand several new tools, such as Clonal Frame (Didelot and Falush 2007) that were driven by the development of new genotyping typing techniques, have emerged and enable researchers make inferences about risks based on population genetics. The reasoning behind this is that genetic relations between pathogen subtypes can be indicators of host association or transmission pathways. These methods enable the comparison of bacterial populations from different sources based on the occurrence of bacterial subtypes, but also take into account how closely these subtypes are related and how they could be evolved from each other. An understanding of relatedness between strains can add on to our knowledge of how different sources contribute to the disease burden. The recent application of the MLST typing scheme to *Campylobacter* (Dingle, Colles et al. 2001) has boosted investigation into this pathogen's population structure and host association by a variety of techniques such as Minimum Spanning Trees (Feil, Li et al. 2004) and STRUCTURE (Pritchard, Stephens et al. 2000). These techniques will be in detail described below.

Some of these methods can directly attribute risks to sources, for example the Island model (Wilson, Gabriel et al. 2008), the Hald model (Hald, Vose et al. 2004) and STRUCTURE (Pritchard, Stephens et al. 2000), others are visualisation or clustering tools, to illustrate the relatedness of bacterial subtypes (for example Minimum Spanning Trees (Spratt, Hanage et al. 2004) or the overlap between frequency distributions (for example the proportional similarity index (Rosef, Kapperud et al. 1985) but cannot provide direct estimates of risks. However these tools still provide an increased insight into the population dynamics of pathogen and can support the conclusion drawn from modelling results as well as improve our understanding of the relative importance of disease sources. Combining several different approaches, while discussing their individual strengths and shortcomings, enables us to gain more confidence in our understanding of the relative importance of the many infection pathways for campylobacteriosis.

Hald model and Dutch model

Recently a Bayesian risk assessment model has been developed (Hald, Vose et al. 2004), that uses the variation of strains in different sources to quantify the contribution of different food sources to the number of human cases of Salmonella in Denmark. The principle behind this approach is to compare the number of human cases caused by dif-

ferent 'types' with their prevalence in different food sources, weighted by the amount of food source consumed. This model is a further development of a frequentist model used in the Netherlands ('Dutch model') (Hald 2002) and requires a heterogeneous distribution of some types among animal and food sources. An important improvement of the Danish model is that while the Dutch model is deterministic and provides risk estimates without uncertainty the Hald model gives Bayesian credibility intervals around the estimated parameters. This approach was initially focused on salmonellosis but was recently extended to campylobacteriosis (Mullner, Jones et al. 2009) and is applied and discussed in Chapter 5 and Chapter 7 of this thesis. Since it is focused on the major food-animal reservoirs, it cannot identify responsible foods at the point of consumption or at other points along the farm-to-fork continuum (Batz, Doyle et al. 2005). Additionally, causation cannot be discerned for cases without distinctive Salmonella types (as they are proportionally attributed) and other sources of infections are not directly included, thereby assuming that the original source of bacterial infection are animal reservoirs (Batz, Doyle et al. 2005).

Proportional similarity index (Czekanowski index)

The proportional similarity index (PSI) or Czekanowski index is an objective and simple estimate of the area of intersection between two frequency distributions (Rosef, Kapperud et al. 1985). The PSI estimates the similarity between the frequency distributions of for example bacterial sub types between different reservoirs. It is calculated by:

$$PS = 1 - 0.5 \sum_i \mid p_1 - q_i \mid = \sum_i min(p_i, q_i) \ ,$$

where p_i and q_i represent the proportion of strains belonging to type i out of all strains typed from species P and Q (Feinsinger, Spears et al. 1981; Rosef, Kapperud et al. 1985). The values for PS range from 1 for the highest possible similarity to 0 for distribution with no common types. Bootstrap confidence intervals for this measure can be estimated based on the approach applied by Garrett et al. (Garrett, Devane et al. 2007). The occurrence of non-typable strains in a dataset requires special attention when applying this method (Rosef, Kapperud et al. 1985; Garrett, Devane et al. 2007). The assumption made by this method is that epidemiological affinity between species is proportional to the similarity between the serotype distributions of the species being compared. This may be incorrect since many animal isolates may not be pathogenic, even if they are identical as determined by the typing method used. If a source also contains a high proportion of non-pathogenic strains its importance as contributor to human cases may be masked (Garrett, Devane et al. 2007). In addition some of the human cases may have originated from non included sources, for example by travel

(Rosef, Kapperud et al. 1985). This method is used in Chapter 4 and 7 of this thesis.

Minimum spanning trees (MSTs)

Minimum spanning trees (MSTs) (Figure 2.2) have been developed as alternative methods to represent the relationships among isolates from bacterial population to overcome problems associated with other commonly used clustering methods (Spratt, Hanage et al. 2004). This approach does not impose a tree-like pattern of descent, which is particular important for investigating species that undergo frequent recombination, like *Campylobacter*. MSTs are a graphical clustering tool that is part of the BioNumerics software (Applied Maths; http://applied-maths.com/bionumerics/bionumerics.htm) and is also available at the PubMLST website (http://pubmlst.org/analysis/). This method links allele designations within an MLST database to a minimum spanning tree, which is calculated by using Prim's algorithm (Wirth, Falush et al. 2006). The algorithm has been modified to choose between otherwise equivalent, alternative sub trees at each step by implementing priority rules that incorporate aspects of the BURST algorithm (Feil, Li et al. 2004; Spratt, Hanage et al. 2004). The BURST algorithm is a clustering algorithm for use on MLST data from bacterial pathogens and differs radically from normal clustering algorithms as it incorporates a simple, but realistic model, of the way in which bacterial clones emerge and diversify to form clonal complexes. According to this model, a genotype occasionally increases in frequency, as a consequence of genetic drift or natural selection, and diversifies by the accumulation of mutations and/or localised recombinational replacements to result in slight variants of the founding genotype. The members of an emerging clone will initially be indistinguishable in their allelic profile but over time they will diversify to variants of the founding ST (Spratt, Hanage et al. 2004). The approach consecutively examines the relationships between very closely related genotypes within clonal complexes, and the key part of the analysis is in the identification of the most likely extant 'ancestral' genotype of each clonal complex, from which the clonal variants may have descended. This ancestral genotype is called the 'consensus' clones and radial spread from this ancestral core is reflected by a series of circles, the size of which represent the number of isolates per ST.

STRUCTURE

This model-based clustering method (Figure 2.3) is used on multilocus genotype data to infer population structure and attempts to assign individuals probabilistically to populations on the basis of their genotypes, while simultaneously estimating population allele frequencies (Pritchard, Stephens et al. 2000). This method is useful in particular for identifying populations and assigning individuals where there is little information about population structure. In addition, a framework has been developed that allows

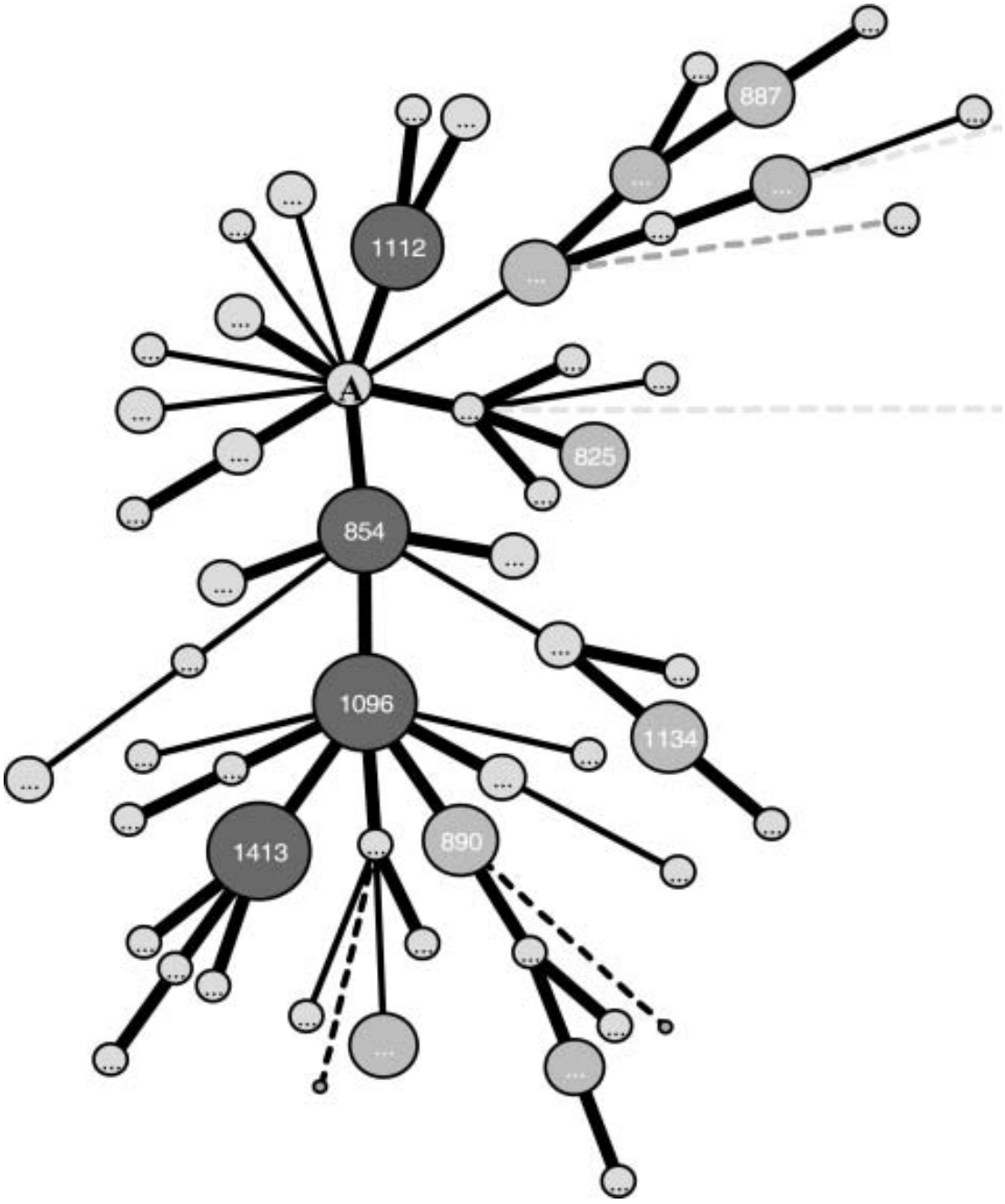

Figure 2.2: Minimum spanning tree constructed using the STs from 151 *C. coli* isolates from two swine production systems (Thakur, Morgan Morrow et al. 2006).

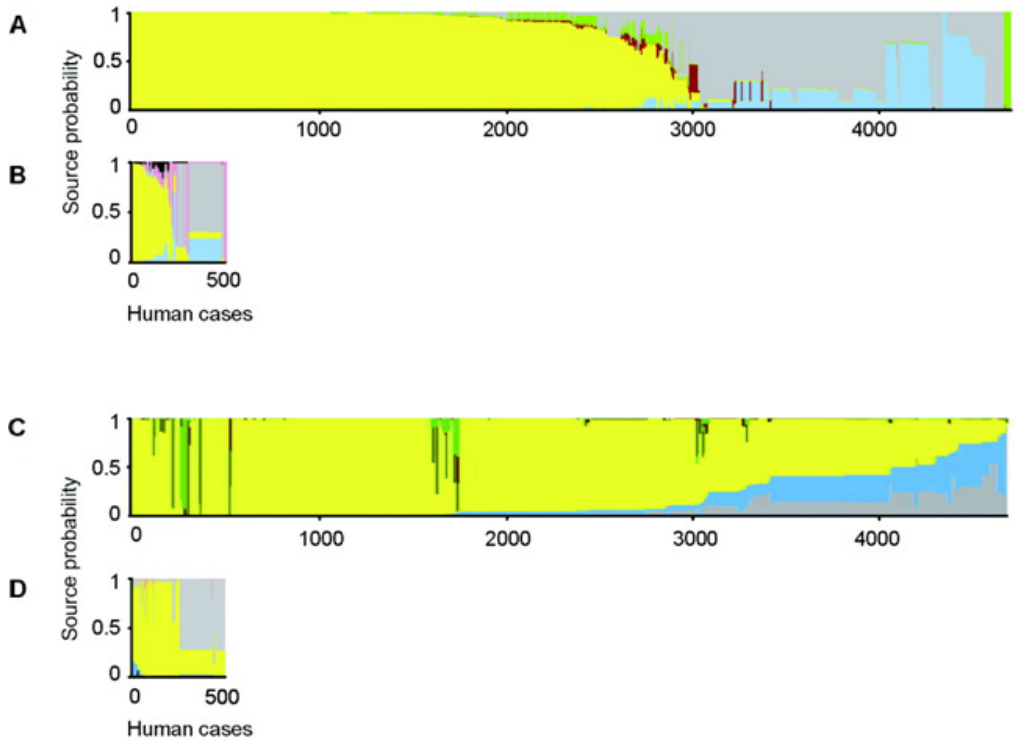

Figure 2.3: Assignment of human clinical cases of campylobacteriosis to source using the Bayesian clustering algorithm STRUCTURE (A and B) and the asymmetric island model (C and D). Each isolate is represented by a vertical bar, showing the estimated probability that it comes from each of the putative sources. Sources for *Campylobacter jejuni* were cattle (*blue*), chickens (*yellow*), wild birds (*brown*), the environment (*green*), and sheep (*light gray*). Sources for *Campylobacter coli* were cattle (*blue*), chickens (*yellow*), sheep (*gray*), swine (*pink*), and turkeys (*black*). Isolates are ordered by attributed source. (Sheppard, Dallas et al. 2009)

for combining genetic information with prior information about the geographic sampling location of individuals (Pritchard, Stephens et al. 2000). The algorithms from this method are implemented in the computer software package STRUCTURE which is available at http://www.stats.ox.ac.uk/ pritch/home.html. Recently the model has been extended (Falush, Stephens et al. 2003) to allow for linkage between loci as well as a new prior model for the allele frequencies has been applied, which allows for the identification of subtle population subdivisions. However there may not be enough information in a particular dataset to estimate all the parameters (Falush, Stephens et al. 2003). In addition an approach has been developed to extend this model to dominant markers, such as AFLP, which will be particularly useful when only limited resources for genotyping are available (Falush, Stephens et al. 2007). This method is applied in Chapter 4 of this thesis to investigate the population structure of *C. jejuni* in the New Zealand poultry industry.

Analysis of molecular variance (AMOVA)

AMOVA creates a distance matrix between samples in order to measure the genetic structure of the population from which the samples are drawn. AMOVA is a method of estimating population differentiation directly from molecular data and testing hypotheses about such differentiation. A variety of molecular data such as molecular marker data (for example, RFLP or AFLP), direct sequence data, or phylogenetic trees based on molecular data, may be analyzed using this method (Excoffier, Smouse et al. 1992). It extends procedures that explicitly use an analysis of variance format to estimate the degree of intra-specific genetic subdivision and permits flexible exploration of a given dataset (Excoffier, Smouse et al. 1992). Certain assumptions are made about the nature of the population, for example, that mating is entirely random and non-assortative, and no inbreeding occurs in addition the effects of selection, are not fully accounted for by this model (Excoffier, Smouse et al. 1992). AMOVA is implemented in Arlequin (Excoffier, Laval et al. 2005) an integrated software package for population genetics data analysis and is freely available on http://cmpg.unibe.ch/software/arlequin3. It provides users with quite a large set of basic methods and statistical tests, in order to extract information on genetic and demographic features of a collection of population samples.

Clonal Frame

The variety of evolutionary mechanisms by which bacteria evolve can pose problems when attempting to infer relationships between strains. This approach developed by Didelot and Falush (Didelot and Falush 2007) for MLST data, does infer the clonal relationship of bacteria by in particular accounting for not only point mutation but also recombination events. The model estimates the extent of the clonal frame for each branch of the genealogy, which is the subset of the genome that has not undergone recombination (Figure 2.4). This method can be used to decide whether a subset of isolates share common ancestry, to estimate the age of the common ancestor and hence to address a variety of epidemiological and ecological questions that hinge on the pattern of bacterial spread (Didelot and Falush 2007). Since the key assumption of the model concerns recombination, and it does not model the origin of genetic imports, the model tends to underestimate the number of recombination compared to mutation events and can infer incorrect subdivisions, particularly if recombination is relatively frequent compared to mutation. The algorithms have been implemented in a computer software package which is freely available at: http://bacteria.stats.ox.ac.uk/.

Island model

This method (Figure 2.3) is based on coalescent models, which are different from classical methods in their explicit considerations of the genealogical history of sampled

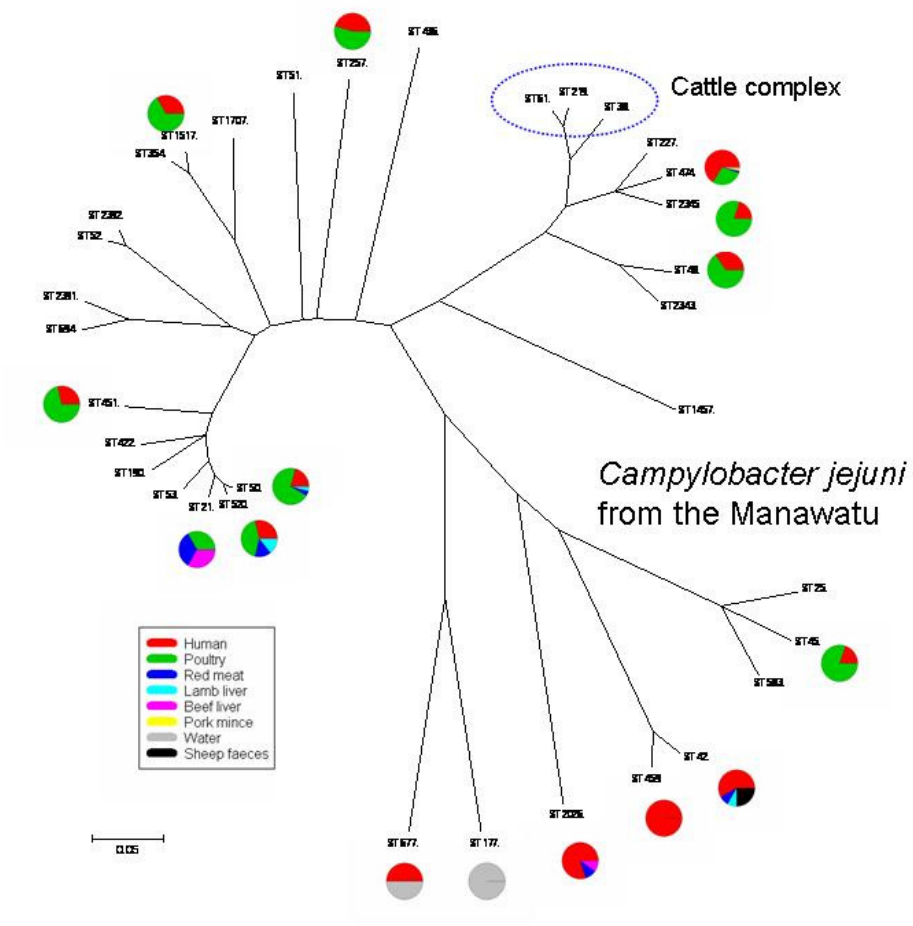

Figure 2.4: Consensus tree of *Campylobacter* isolates collected in the Manawatu region of New Zealand (provided by N. P. French)

alleles, and are one of many tools for addressing problems in population genetics (Hudson 1998). Migration in or out of a population may be responsible for a marked change in allele frequencies, with mobility being the biggest driver of the rate of gene flow between two populations. Island models were first proposed by Wright (Wright 1931) and are models of gene flow derived from population genetics. Recently an island model has been successfully applied to *Campylobacter* MLST data (Wilson, Gabriel et al. 2008; Sheppard, Dallas et al. 2009) which uses the MLST data in animal populations ('islands') to estimate mutation rates, recombination rates and migration rates. Based on these estimates, migration into the human population is estimated for source attribution. This model produces source attribution estimates which include uncertainty and is capable of attributing all human cases to a food source, not only the ones caused by subtypes detected in a source. This process of allocating cases to a source based on genetic relatedness relies on careful epidemiological and population genetic considerations and the assumptions behind this approach are in detail discussed in Wilson and Gabriel (2008). This method is applied in Chapter 3 and 7 of this thesis.

2.4.5 Strain variation and MRA approaches

General considerations

Microbial hazards differ from for example chemical hazards in that they are living entities and therefore their frequency and numbers as well as the source may be among the primary issues to consider in undertaking an MRA (Cahill 2005). Furthermore, microorganisms can increase or decrease in numbers, vary in terms of geno- and phenotype and can be both dynamic and adaptable. A single exposure can be sufficient to cause an adverse health effect and each exposure can be considered an independent non-cumulative event (Cahill 2005).

It is well known that at the bacterial species level there is sufficient genetic diversity to allow identification of different clones, frequently characterised by different virulence (Manfreda and De Cesare 2005). A clone is defined as genetically related isolates that are indistinguishable from each other by a variety of molecular typing methods, or isolates that are so similar, that they are presumed to be derived from a common ancestor (Tenover, Arbeit et al. 1995). Typing of bacterial isolates from different sources is a valuable tool for intervention and infection control and could add immense value to risk assessment studies of sources of human disease.

The ability to characterise pathogens into individual strains or clones by molecular biology techniques reveals characteristics about the transmission of infectious diseases, that would have been extremely difficult, if not impossible, to identify by conventional laboratory methods (Riley 2004). A strain (or clone), characterised by a molecular typing method, that has previously been shown to be linked to a particular vehicle may provide a clue about the sources of infection for sporadic cases of an illness (Riley 2004).

For example a comparison of *Campylobacter* types from food animals and food of animal origin, with isolates from humans, can be used to estimate the number of human cases attributable to certain animal sources (Nielsen, Engberg et al. 2000). According to Whiting and Golden (2002) unless the identity and characteristics of a single strain are known, as when reconstructing a single outbreak, interpretation of the potential hazard of a microbial pathogen must consider strain-to-strain variation. Studies indicate that strain variation plays an important role in assessing the potential exposure of a consumer to a pathogen (Hof and Rocourt 1992; Whiting and Golden 2002; DeJesus and Whiting 2003; Oscar 2004; Cahill 2005). Strain specific case-control studies or strain specific exposure assessment would be a way to address these issues. However when using these approach it has to be acknowledged, that these strain similarity comparisons simplify transmission dynamics by assuming a simple direct transmission from a subtype found in one source to another source (Riley 2004). The consideration of the whole pathway using a multidisciplinary approach allows for a more detailed and mechanistic but more resourceful view on disease transmission (Grenfell, Pybus et al. 2004; Xiao, Clancy et al. 2006).

Pathogenicity and virulence characteristics of individual strains also play an important role, in particular with regard to the dose-response relationship between a pathogen and an outcome in a population. Pathogenicity is defined as the property of an organism that determines the extent to which overt disease is produced in an infected population, or the power of an organism to produce disease (Last 2001). Virulence, on the other hand, is the degree of pathogenicity or the disease-evoking power of a microorganism in a given host (Last 2001). The isolation of bacteria from food items does not, in itself, suggest a true health risk. The pathogenic potency of different isolates has to be taken into account (Hof and Rocourt 1992). There is, for example, variability of pathogenicity within serovars and different virulence in different serovars of *Listeria monocytogenes* (Hof and Rocourt 1992). This is also known for other pathogens, for example epidemiological data indicate that some *Salmonella* strains are particularly pathogenic (Latimer, Jaykus et al. 2001). The influence of strain effects on survival and growth has also been investigated by several authors. Variation among strains of for example *E. coli* O157:H7 (Whiting and Golden 2002) and *Listeria monocytogenes* (DeJesus and Whiting 2003) could be detected and both authors propose to consider these results when undertaking microbial risk assessments. The ability to differentiate pathogenic from non-pathogenic strains would greatly improve the accuracy of risk assessment models and more closely define the important sources and routes of transmission for human infection (Newell, Frost et al. 2000).

Subtyping of human cases and foods can provide a direct link between a public health endpoint and the source of infection and can be used to identify specific food or animal reservoirs (Batz, Doyle et al. 2005). When ignoring the specific epidemiology of

different strains, specific host-pathogen combinations cannot be identified and the overall results combining all strains are subject to major bias. As an example, case-control studies are potentially biased when the different epidemiology of strains is not considered. Gillespie et al. (2002) proposed the use of case-case studies for *Campylobacter* at the species level, using a combined epidemiological and microbiological dataset, linking detailed epidemiologic exposure information with detailed microbiological strain characterisation. Given the known variation in the epidemiology of different pathogen types (Heir, Lindstedt et al. 2004; Kemp, Leatherbarrow et al. 2005), which is well recognised for *Salmonella* spp. on a serotype level (Manfreda and De Cesare 2005), these authors consider it inappropriate to conduct a case-control study when cases are defined at the species level. Sopwith et al. have also addressed this issue by analysing combined serotyping and epidemiological data (Sopwith, Ashton et al. 2003).

Using strain level data is the core feature of the newly used developed attribution tools, for example the Hald model (2004) which utilises the relative frequency of different strains in humans and food to provide risk estimates, and is therefore uniquely dependent upon the availability of strain typing data. This method allows for variation of some of the characteristics of different strains by estimating a "bacteria-related" factor for each strain as an indicator of the pathogen's ability to cause disease and survive processing, On the other hand novel risk research tools based on population genetics (Sheppard, Dallas et al. 2009) are explicitly using the population structure of pathogen populations to draw conclusion on the most likely sources of human infection.

Impact of strain variation on MRA approaches

According to Coleman et al. (2004) current food chain risk models of *Salmonella* and *Campylobacter* spp. do not take into account strain variability. However ignoring potential strain effects might cause misleading estimations of the possible risks as well as lead to substantial uncertainty in risk estimates. This could provide a misleading representation of both the public health implications and the confidence that certain interventions would be effective in reducing risk. The authors present evidence that strains differ in their capability to cause illness and point out the difficulties of modelling risk in the presence of strain diversity. They conclude that epidemiological and biological considerations of the disease triangle should influence the choice of the final model.

Pathogen strain variability is considered likely to have a strong influence upon risk assessment models (Coleman and Marks 1998). Dose-response assessment is a crucial step in the risk assessment process in which the risk assessor seeks to quantitatively characterize the relationship between the dose of the microbial contaminant consumed and the infection/illness response of the exposed subject (Latimer, Jaykus et al. 2001). Experimental dose-response data are often poor and studies cannot assess the inherent

variability in the virulence and pathogenicity of multiple strains of microorganisms (Latimer, Jaykus et al. 2001). For example commonly used clinical data from humans was gathered in a study in which two strains of *Campylobacter* were administered to support the existence of strain effects (Black, Levine et al. 1988; Coleman, Marks et al. 2004). However, for several reasons Coleman proposes that these data are inappropriate for use in a risk assessment (Coleman, Marks et al. 2004). To address differences in strain pathogenicity Latimer et al. developed a weighted composite dose-response model for human salmonellosis, that explicitly models variability in strain virulence and pathogenicity (Latimer, Jaykus et al. 2001). Oscar (2004) developed an approach useful for predicting dose response as a function of pathogen factors and observed variation in virulence among 13 strains of Salmonella in agreement with others (Coleman and Marks 1998; Latimer, Jaykus et al. 2001). Ignoring strain variation when modelling dose-response relationships could therefore result in inaccurate and/or over-precise estimates of risk.

When modelling exposure, for example, using MPRM models (Nauta 2001; Nauta, van der Fels-Klerx et al. 2005), strain differences can constitute an important source of variation in growth and survival of pathogens. The commonly used Gompertz growth model (Walls and Scott 1997; Whiting and Buchanan 1997; Cassin, Lammerding et al. 1998) assumes equal ability of different pathogens to survive and grow under for example processing conditions, however this assumption, enforced by the complexity and dynamic nature of food production, may have limited validity (Hald, Vose et al. 2004). Chan et al. (2001) showed that strains of *Campylobacter* from human clinical isolates showed better survival at refrigeration temperatures than poultry-derived strains. These findings indicate that refrigeration may select poultry-derived strains for cold tolerance which then enter the pool of human clinical isolates. Different survival characteristics in different strains are also documented for other pathogens (Guan, Grenier et al. 2006; Ravva, Sarreal et al. 2006).

Finally, when considering strain effects as a source of variability in risk models, not modelling them separately leads to an increase in the total uncertainty of the model. By doing this and therefore mixing uncertainty and variability we cannot see how much of the total uncertainty in our model comes from variability and how much from uncertainty (Vose 2000). Including this variation will decrease the precision of any estimate provided by models and risk assessments, but reflects the diversity of the microbial world (Whiting and Golden 2002).

The impact of strain variation on the risk estimate

Typing of bacterial isolates may require additional financial cost and is considered a labour and time intensive task. In addition MRAs are, in general, resource intensive to begin with and often suffer from data gaps in several areas. Therefore the benefits of

including strain variation in the risk assessment process need to be carefully considered. To illustrate the possible effects of ignoring strain variation we use a simple simulation model (Figure 2.5). Data on the distribution of counts of *Campylobacter* spp. were derived from a survey of New Zealand chicken carcasses (French and Molecular Epidemiology and Veterinary Public Health Group Hopkirk Institute 2008). Based on this data the numbers of *Campylobacter* spp. on a cooked chicken portion were estimated (McBride, Meleason et al. 2005). In a next step data by Chan et al. (2001) on the observed rate of viability loss for 19 poultry and human derived *Campylobacter* strains was used to model strain specific difference in survival during prolonged refrigeration. Finally, by using the beta poisson dose-response model (Medema, Teunis et al. 1996), the risk of infection when the chilled portion was consumed was calculated.

The mean risk of infection, as estimated by the model, ranged from 0.10 to 0.32 for individual strains with a mean value for all 19 strains of 0.23 (Table 2.2). This indicates that, if strain variation is not considered, the risk of infection can be overestimated when for example survivability is lower than expected. For example, the risk of infection from strain 1 (0.10) differs by a factor of 2.3 from the mean estimate for all strains (0.23). On the other hand this can also result in an underestimate of risk respectively, when a strain has a higher than average survivability. For example the mean risk of infection from Strain 19 (0.32) is 36% higher than the mean risk for all strains (0.23). Furthermore the variability in survival of different strains will effect the precison of estimates of risks. Ignoring this variability will result in too precise estimates that do not reflect the adjunct uncertainty sufficiently.

Table 2.2: Mean risk of infection for different strains of *Campylobacter* after consuming a contaminated chilled chicken portion.

	Number of *C. jejuni* on portion before chilling	Number of *C. jejuni* on portion after chilling	Mean risk of infection
Strain 1	258	13	0.1
Strain 2	258	13	0.11
Strain 3	258	17	0.13
Strain 4	258	18	0.14
Strain 5	258	29	0.17
Strain 6	258	30	0.17
Strain 7	258	31	0.17
Strain 8	258	34	0.19
Strain 9	258	43	0.21
Strain 10	258	48	0.2
Strain 11	258	58	0.22
Strain 12	258	58	0.22
Strain 13	258	71	0.25
Strain 14	258	106	0.27
Strain 15	258	106	0.29
Strain 16	258	109	0.28
Strain 17	258	129	0.29
Strain 18	258	134	0.3
Strain 19	258	173	0.32
Mean strain	258	64	0.23

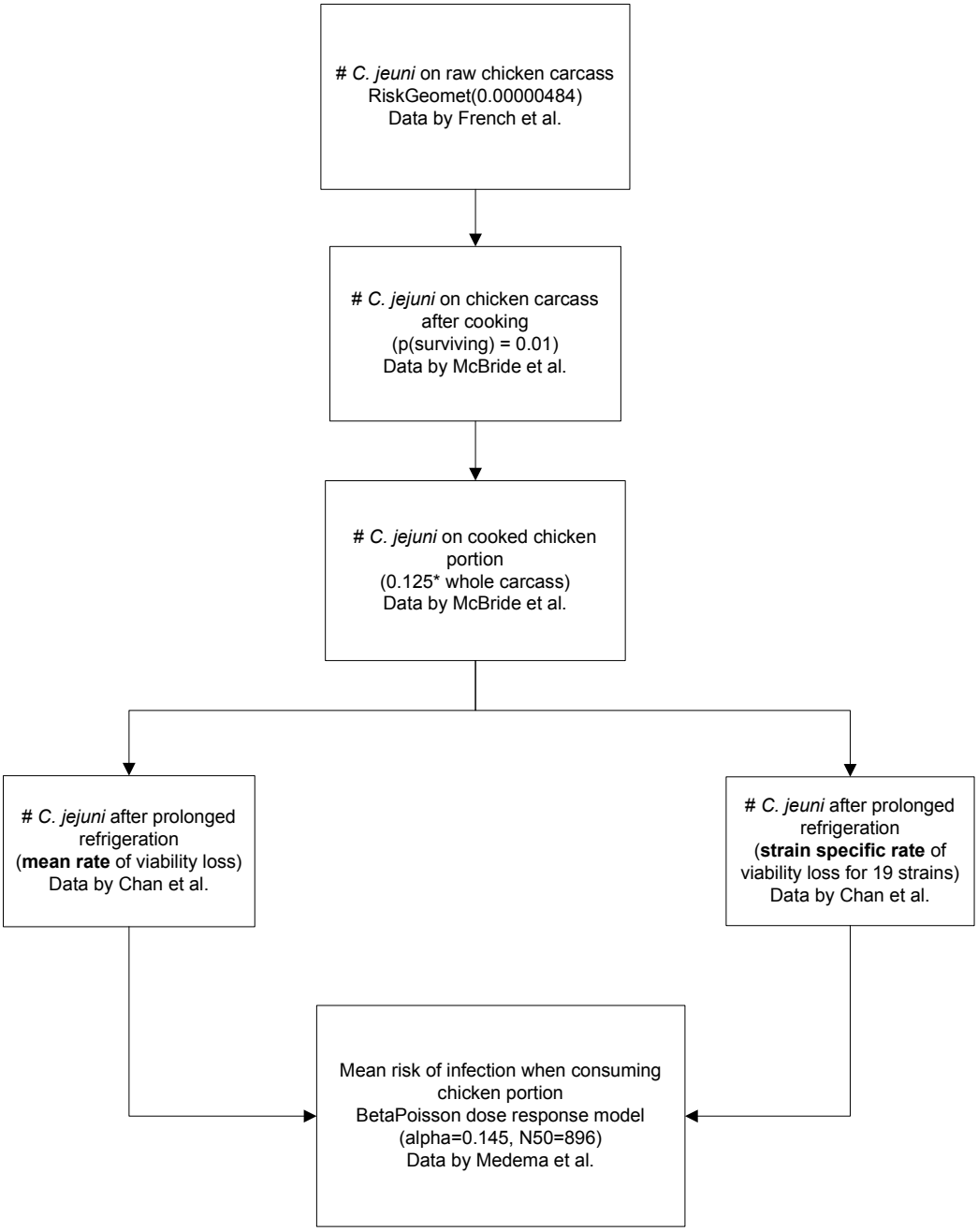

Figure 2.5: Framework for exposure assessment of *Campylobacter* spp. accounting for differences in survival of individual *Campylobacter* strains.

2.4.6 Suitable typing methods for *Campylobacter* risk research

General considerations

The techniques of molecular biology have contributed tremendously to our understanding of the pathogenesis and epidemiology of infectious diseases (Isenberg 1998). There is now a variety of different typing methods available for the characterisation of species and conclusions drawn about most suitable methods are not uniform and differ by pathogen. This is because each method has to be evaluated in context, according to the requirements of different epidemiological investigations. Additionally, there has been an immense progress in the development of advanced subtyping methods in recent years.

Depending on the nature of the bacterial species under investigation, more or less discriminatory methods are suitable for studying the epidemiology of bacteria. Highly discriminatory methods or combinations of methods are necessary for typing of clonal populations, whereas stable and perhaps less discriminatory methods are necessary for typing panmictic populations in order to determine the correct relationships between isolates (Nielsen, Engberg et al. 2000). Riley (2004) illustrated that the utility of a new typing method for epidemiologic purposes does not necessarily correlate with high discriminatory power. With regard to the requirements for epidemiological analysis more discriminatory tools are needed for outbreak investigations than for solving long term epidemiological questions. Genetic markers need to differentiate between related organisms yet provide sufficient commonality as to permit inclusion into already established groupings (Thompson 2000).

Genotypic differences between isolates from different hosts or reservoirs may indicate a decreased likelihood of successful transmission of infection between them. Differentiation into strains which are found in humans as well as in animals, and strains exclusive to animals, could have significant impact on the prevention of human disease (Moser, Lentzsch et al. 2002). As an example genomotyping was recently applied to *C. jejuni* and consecutive phylogenetic analysis revealed genetic markers that were potentially predictive of infection source (Champion, Gaunt et al. 2005), however more recent studies have contradicted results from this study as a likely results of phylogenetic assumptions in this approach (Wilson, Gabriel et al. 2008).

Methods producing a large proportion of untypeable isolates (for example phenotyping) are not well suited for use in risk models, as they result in major methodological challenges. Riley concludes that 'the ultimate test of validity of a molecular typing system is its effectiveness in generating information that leads to improved therapy, prevention and control of an infectious disease problem, that could not have been obtained by traditional laboratory or epidemiological methods' (Riley 2004).

Which methods are currently the most appropriate for *Campylobacter* risk assessment?

For epidemiological studies, a wide range of phenotypic and genotypic typing systems have been developed and applied to subtyping of *C. jejuni* and *C. coli* (Ono, Kurazono et al. 2003). However, there is not yet a universally accepted solution to the question of *Campylobacter* typing (Advisory Committee on the Microbiological Safety of Food (ACMSF) 2005). The Second Report on *Campylobacter* by the ACMSF (2005) summarizes this by describing the features of the main phenotypic and genotypic methods employed for typing in a clear tabular format. Additionally, due to new or further development of typing techniques, there has been continuous progress in the availability of subtyping methodology.

A comparison of *Campylobacter* types from food animals and food of animal origin with isolates from humans makes it possible to produce estimates for the number of human cases attributable to certain animal sources (Nielsen, Engberg et al. 2000). In general strain similarity comparisons deal with the probability that a particular subtyping assignment made by a molecular strain typing method reflects a true epidemiologic relationship of the microorganism in question (Riley 2004). Typing of bacterial isolates from different sources is a prerequisite for intervention and infection control and contributes to risk assessment studies of sources of human campylobacteriosis.

However, *Campylobacter* has a high genetic and antigenetic diversity (Colles, Jones et al. 2003; Sails, Swaminathan et al. 2003; Dingle, Colles et al. 2005), which makes the investigation of this pathogen difficult. Several recently published articles point out that the newly developed approach of MLST is particularly suited to long-term and global epidemiology as it identifies variation which is accumulating slowly within a population (Dingle, Colles et al. 2001). MLST has been successfully applied to investigate the epidemiology of other pathogens and allows for epidemiologically meaningful subtyping (Smith, Green et al. 2005; Zadoks, Schukken et al. 2005). This method has also recently been applied in several countries to examine the diversity and population structure of this pathogen in human and non-human samples (Sopwith, Birtles et al. 2006; Karenlampi, Rautelin et al. 2007; Mickan, Doyle et al. 2007; Sheppard, Dallas et al. 2009).

MEE and MLST both measure variations that occur relatively slowly and without selective pressure and thus are more appropriate for phylogenetic studies. The relative simplicity of the MLST method over MEE has led to an explosion of interest in this new technique (Sails, Swaminathan et al. 2003). The high degree of discrimination possible presents the prospect of employing MLST data that can be used to improve estimates of the contribution of different *Campylobacter* populations to the human disease burden (Colles, Jones et al. 2003). It enables population structure to be established, permits investigations to determine the relative importance of the food

chain and its individual components as a source of human infections, and can be used to investigate a variety of aspects of the molecular epidemiology of this pathogen (Dingle, Colles et al. 2002; Wareing, Bolton et al. 2002; Sheppard, McCarthy et al. 2007; Colles, Jones et al. 2008; French, Midwinter et al. 2008; Gormley, MacRae et al. 2008; McTavish, Pope et al. 2008). There is evidence of association of certain clonal complexes with particular farm animals (Colles, Jones et al. 2003; Kwan, Barrigas et al. 2008; Kwan, Birtles et al. 2008) and this technique permits the confirmation of suspected routes of transmission from the environment and livestock to humans (Dingle, Colles et al. 2001). This approach can assist in resolving issues such as the relationship of environmental to disease isolates at population level as well as the structure of *Campylobacter jejuni* populations (Dingle, Colles et al. 2001). With the ability to group types into Sequence Types (ST) or clonal complexes (CC) this method additionally allows adjustment of the level of discrimination to the needs of the analysis. It was demonstrated that *C. jejuni* is genetically diverse with a weakly clonal population structure, which is thought to consist of clonal complexes or lineages in which the isolates are considered to be derived from common ancestors (Sails, Swaminathan et al. 2003). MLST is specially suitable for the investigation of such bacterial populations and clonal complexes, as defined by MLST, and is an epidemiologically relevant unit for both long and short-term investigations of *C. jejuni* epidemiology (Dingle, Colles et al. 2002). Sails et al. (2003) found that ST - 21, - 45, -48 and - 353 clonal complexes all contained groups of indistinguishable isolates from diverse geographical regions, including five countries in North America, Asia and Europe, and isolates that were isolated from both human and animals. This demonstrates the usefulness of MLST for investigation of the global epidemiology of this important pathogen and illustrates its potential to identify indistinguishable strains or clones in geographically distinct regions (Sails, Swaminathan et al. 2003).

Application of the MLST method to prospective studies of isolates from human patients with gastroenteritis may demonstrate less variability in the loci detected. This apparent lack of diversity among clinical isolates may limit the usefulness of this technique for short-term epidemiological investigations (Sails, Swaminathan et al. 2003). However a combination of MLST and SVR sequencing of the *fla*A flagellin gene allows resolution equivalent for that of PFGE for outbreak investigation (Dingle, Colles et al. 2005). Recently, an extended MLST typing system has been introduced that allows for the differentiation of clinically relevant *Campylobacter* species (Miller, On et al. 2005). Additionally for *C. coli* MLST and *fla*A SVR sequencing could allow accurate assessments of the contribution that different infection sources make to the burden of human disease, as all MLST data are directly comparable (Dingle, Colles et al. 2005).

MLST has a number of technical advantages over other subtyping methods. It is stable method with minimal experimental variation, which facilitates interlaboratory

comparison and provides the potential for the establishment of a standardized nomenclature for DNA sequence data (Duim, Godschalk et al. 2003; Sails, Swaminathan et al. 2003). It also has as high discriminatory power, high reproducibility, and is simple to interpret (Dingle, Colles et al. 2001). Sequence typing by MLST is emerging internationally as the new gold standard for national and epidemiological characterisation of major pathogenic organisms such as *C. jejuni* (Dingle, Colles et al. 2001; Dingle, Colles et al. 2002).

Sequence-based methods, of which MLST is the forerunner, are true typing methods rather than finger printing (Advisory Committee on the Microbiological Safety of Food (ACMSF) 2005). Much of the past microbial occurrence data are developed with very different protocols and monitoring approaches. These data have limited application for quantitative risk assessment and it is now recognised that quantitative databases must be developed, as often this is the major data gap for adequate risk assessment (Haas, Rose et al. 1999). MLST allows for the development of such a database (Dingle, Colles et al. 2001; Dingle, Colles et al. 2002; Dingle, Colles et al. 2005). Heterogeneity in *Campylobacter* MLST types has been observed in other studies (French, Barrigas et al. 2005) and the approach was applied to apportion human case of campylobacteriosis to a disease source in New Zealand (Mullner, Jones et al. 2009), which is in detail described in Chapter 5. In addition several population genetics based models have successfully been applied to MLST data of *Campylobacter* isolates (Pritchard, Stephens et al. 2000; Didelot and Falush 2007; Wilson, Gabriel et al. 2008; Sheppard, Dallas et al. 2009).

Although PFGE is still seen as a gold standard for the global epidemiological investigation of *Campylobacter* species, the ability of this method to identify epidemiological links such as risk factors, or to assess the contribution of different sources to the human burden of campylobacteriosis is being questioned by different authors. Genomic polymorphism analysis, while useful for epidemiological subtyping, requires careful interpretation if used as a measure of actual phylogenetic relationship between isolates (Jackson, Fox et al. 1996). Highly discriminatory methods such as PFGE index variation in the *C. jejuni* genome that may be unstable due to genomic rearrangement, make such methods unsuitable for longitudinal epidemiological studies (Sails, Swaminathan et al. 2003). However an outbreak situation is a rather short-term event, genetic instability is not generally thought to be a problem in this context (Newell, Frost et al. 2000). PFGE has proven itself as a very useful tool for outbreak investigation and has through PulseNet (http://www.cdc.gov/pulsenet/) lead the way for the detection of food borne disease clusters and the early detection of common source outbreaks for several pathogens, including *E. coli*, *Salmonella* and *Campylobacter*.

In a study evaluating phenotypic and genotypic methods for subtyping *C. jejuni* isolates from different sources (Nielsen, Engberg et al. 2000) the grouping of isolates formed by at least four typing systems was used for the evaluation of concordance of

methods and it was found that the highly discriminatory PFGE most often disagreed with the other methods (methods used were Penner serotyping, automated ribotyping, RAPD, PFGE, *fla*A-RFLP and *fla*A-DGGE). Recently the distribution of MLST STs among a collection of New Zealand *C. jejuni* isolates previously characterised using Pulsed Field Gel Electrophoresis (PFGE) and Penner serotyping in the PulseNet Aotearoa New Zealand database was determined (Carter 2007). It could be observed that indistinguishable PFGE patterns did not have identical STs but the PFGE pattern was found to be a good indicator of CC. When comparing MEE, MLST and PFGE some of the clonal complexes identified by MLST contained isolates that were indistinguishable by PFGE; however, overall there was considerable diversity in the PFGE types within each complex. This diversity may be due to the effects of *C. jejuni* genomic instability on PFGE profiles (Sails, Swaminathan et al. 2003). By comparison (with AFLP and *fla*A analysis), for *C. coli*, PFGE appeared to provide less useful information for examining strain-host associations or epidemiological relationships (Leatherbarrow, Hart et al. 2004). In a study by Schouls et al. (Schouls, Reulen et al. 2003) cluster analysis yielded a similar grouping of strains by either AFLP analysis or MLST. This congruence indicates that these methods, both of which are based on characterisation of multiple loci in the genome, are equally suited for typing of *Campylobacter*. AFLP is most discriminatory (when compared with Ribotyping, *fla*A and PFGE), capable of typing large numbers, and best suited for computer assisted analysis (De Boer, Duim et al. 2000). This method is cheaper, faster, and easier to perform than MLST but interlaboratory comparison by AFLP analysis will be difficult because complex banding are PCR based and therefore prone to variation. However further identifying of AFLP bands by sequencing is possible (Godschalk, Bergman et al. 2006). Generally speaking MLST is more expensive, but it does result in solid DNA sequence data that is not subject to experimental variation (Schouls, Reulen et al. 2003).

A three point ranking system for typing systems

To assess the applicability and utility of a typing method within a MRA framework in a more objective way a three point ranking system was developed. Potential methods can be ranked using this system to find the method most useful in the given context. The focus of this simple approach is performance and outcome of the methods rather than factors such as the cost and speed of a typing method. We identified typeability, comparability and portability as well as the characterisation of strains into meaningful groups as the key characteristics for a typing method to inform risk assessments.

Typeability is defined as the proportion of strains that are assigned a type by the typing system (Struelens and Members of the European Study Group on Epidemiological Markers (ESGEM) of the European Society for Clinical Microbiology and Infectious Diseases (ESCMID) 1996). A high percentage of non-typeable strains make

Table 2.3: Ranking of four typing methods for *Campylobacter* spp. using the three point ranking system.

Key characteristic	MLST	PFGE	*fla*A typing	Serotyping
Typeability	2	1	3	4
Comparability and Portability	1	2	3	4
Grouping of strains	1	2	3	4
Rank	1	2	3	4

the use of the technique within risk research difficult as the necessary information will only be available for a part of the isolates processed. As the underlying causes for non-typeability can be multiple, a low typeability will invalidate the inferences made.

Comparability and portability are important features to maximise the benefits of results obtained by typing methods. Results should not only be comparable within the same laboratory but also between laboratories, ideally allowing for comparison on an international scale. Portability allows for an easy exchange and analysis of data and the development of databases for sharing results between individual institutions.

The **characterisation of strains into meaningful groups** postulates that related organisms are grouped together by the method. This allows for inferences about possible host-association of individual strains and is an important feature to attribute risks of disease to individual sources based on information on the molecular level. This characteristic is a function of the desired discriminatory power which depends on the population structure of the pathogen.

Methods under investigation are ranked with respect to these three individual criteria. Ranking scores are summed for the three criteria and the highest ranking method is believed to be the method of choice in the context given. A practical example of this approach is illustrated below on four techniques included in this review (Table 2.3). The three point ranking system was applied to a set of *Campylobacter* typing approaches: serotyping, *fla*A typing using RFLP, PFGE and MLST. The (subjective) ranking scores assigned are based on the above review.

2.5 Conclusions

We propose the use of strain typing in *Campylobacter* risk research as it allows for a more powerful of the epidemiology, in particular through the identification of particular host-pathogen associations. This will inform decision making and will allow for

targeted disease intervention strategies. The results from our risk model show considerable variability between estimates for different strains. This variability should not be ignored and could potentially be enhanced by strain dependent variability in other characteristics like infectivity.

Newly developed source attribution approaches rely on strain typing data and provide a powerful means for the investigation of population structure, host association and transmission pathways, to generate risk estimates. However a suitable typing method has to be chosen, as not all methods are equally capable of meeting the needs of risk research. We developed a simple three point ranking system to support the identification of a suitable typing tool. We are aware that the costs of the extended analysis of the pathogen have to be carefully weighed against the potential benefits gained.

While band based approaches such as PFGE has proved to be a valuable method of strain typing and will continue to be an important tool for national and international comparison of isolates, nucleic acid sequence-based methods such as MLST have many advantages for strain characterisation including reproducibility, standardisation, auditablity and electronic portability.

In conclusion we regard MLST as the current method of choice to generate high quality data for use in *Campylobacter* risk assessments and attribution approaches, in particular as this technique allows for a meaningful level of discrimination to assess the risk of human *Campylobacter* infection from different exposure sources. However these recommendations do not consider aspects like costs, relative and absolute speed or availability which needs to be taken into account when a practical solution is required. Additionally factors like availability and cost are likely to change over time and may vary considerably between countries.

MLST typing has been applied in the sentinel surveillance study presented in this thesis and some of the source attribution tools presented in this chapter are applied in other chapters of this thesis, including the Dutch model (Chapter 7), the Hald model (Chapters 5 and 7, the PSI (Chapter 7), STRUCTURE (Chapter 4) and the island model (Chapters 3 and 7).

2.6 Appendices

Performance criteria for typing approaches are further described in Appendix A
Bayesian approaches are further described in Appendix D

Chapter 3

Molecular and spatial epidemiology of human campylobacteriosis in New Zealand - source association and genotype-related risk factors

P. Mullner, T. Shadbolt, J. M. Collins-Emerson, A. C. Midwinter, S.E.F. Spencer, J. Marshall, P. Carter, D.M. Campbell, D. J. Wilson, S. Hathaway, N.P. French. Molecular and spatial epidemiology of human campylobacteriosis in New Zealand - source association and genotype-related risk factors. Manuscript prepared for *Epidemiology and Infection*.

3.1 Summary

The epidemiology of campylobacteriosis is complex and, despite being a major public health concern in many countries, is not fully understood. This has delayed the development of successful intervention programs for this disease. New Zealand is a country with a comparatively high, yet unexplained, rate of notifications. In this study the epidemiology of *Campylobacter jejuni* was investigated at the genotype-level over a three year period between 2005 and 2008 using multi locus sequence typing. By combining epidemiological surveillance and population genetics a dominant, internationally-rare strain of *C. jejuni*, namely ST - 474, was identified and the majority of human cases (65.7 %) were found to be caused by only seven different genotypes. Source association of STs was used to identify risk factors at the genotype-level through multivariate logistic regression and a spatio-temporal model. Poultry-associated cases were more likely to be found in urban areas compared to rural areas. In particular, young children in rural areas had a higher risk of infection with non-poultry strains than their urban

counterparts. These finding provide important information for the implementation of pathway specific control strategies.

3.2 Introduction

Cases of campylobacteriosis in New Zealand have been increasing steadily in the last decade peaking at 379 cases per 100,000 people per annum in 2006 (Institute of Environmental Science and Research Limited (ESR)). This situation has been accompanied by a considerable amount of media attention with public calls for urgent action to reduce the unacceptable burden of human disease (Baker, Wilson et al. 2006). Internationally rare clones of *Campylobacter* could be isolated during an epidemic (McTavish, Pope et al. 2008), but their role in the long term occurrence of disease remained unclear. The complex epidemiology of campylobacteriosis has hampered the development of successful control programs for this pathogen in New Zealand and elsewhere. *Campylobacter* spp. have been isolated from a wide range of food sources including poultry, red meat and milk, as well as environmental sources such as wild birds and contaminated water. A variety of risk factors for acquiring campylobacteriosis, such as dog ownership and consumption of offal have been identified (Gillespie, O'Brien et al. 2008; Stafford, Schluter et al. 2008).

The ecological situation in New Zealand is quite unique: there is an extensive agricultural usage of land and the islands are geographically remote, resulting in human, animal and pathogen populations that are relatively isolated. The country has among the highest enteric infectious diseases rates in industrialized countries (Lake, Baker et al. 2000), and the high ratio of domestic production animals to humans, and frequent use of rural water supplies in New Zealand have been proposed as underlying causes (Crump, Murdoch et al. 2001). The poultry supply is different to many developed countries in that suppliers are almost entirely focused on the domestic market and for biosecurity reasons no raw poultry products are imported into the country. Despite little movement of animals and animal products into the country as a result of tight biosecurity measures, New Zealand is exposed to a relatively large number of international travelers. With a population of just over 4 million the country receives 4.5 million international arrivals per annum (1.1 arrivals per resident), an estimated 17 arrivals per square kilometer (http://www.stats.govt.nz/default.htm). The species *Campylobacter jejuni* is responsible for around 90 % of human campylobacteriosis cases and differs in its epidemiology from other *Campylobacter* species (Oporto, Esteban et al. 2007), as illustrated in a case-case comparison for *C. jejuni* and *C. coli* (Gillespie, O'Brien et al. 2002). There is evidence that these differences also apply to different strains (McCarthy, Gillespie et al. 2007), which may affect inferences made at the species level when using epidemiological approaches such as the population attributable risk.

In addition, commonly used typing schemes for *Campylobacter* spp., such as serotyp-

ing, have struggled to provide sufficient resolution for detailed epidemiological studies based on differences between strains. The increasing availability of genotyping methods allows for the study of transmission patterns and risk factors at the strain level, and this is changing the way epidemiologists study infectious diseases (Murray 2002). Multilocus sequence typing (MLST) has emerged as a genotyping technique for *Campylobacter* and has revealed differences in the epidemiology of individual genotypes (Sopwith, Birtles et al. 2008) and allows for the identification of source-associated strains. Based on the sequence of seven, relatively conserved, housekeeping genes, this approach has several properties that make it ideal as a typing tool for epidemiological investigations of highly genetic diverse pathogens. Characteristics such as efficiency, accuracy, reproducibility and the opportunity to share data via an openly accessible database add to this (Dingle, Colles et al. 2001) and the technique is becoming recognised internationally as a valuable tool for the national and international epidemiological characterization of major pathogenic microorganisms such as *C. jejuni* (Urwin and Maiden 2003; Sullivan, Diggle et al. 2005).

In response to the high burden of human disease in New Zealand a long-term sentinel surveillance site was established, to better understand key determinants of human campylobacteriosis by combining longitudinal epidemiological surveillance and genotyping data. In this paper we investigate the molecular epidemiology of *C. jejuni* infection in New Zealand at the molecular level. Patterns of human cases are described and spatio-temporal and multivariable analyses are used to identify risk factors for source associated strains. We compare the molecular epidemiology of *C. jejuni* in New Zealand with findings from overseas, and discuss the implications of our results for the prevention and control of this important zoonosis.

3.3 Material and Methods

3.3.1 Study population

The study was conducted within the Manawatu Health District of New Zealand's North Island. The population consists of 155,072 people who live in a mixture of rural and urban dwellings including both costal and inland areas and is centered around the city of Palmerston North, which has a population of 75,000 people.

3.3.2 Case definition

A case was defined as a laboratory-confirmed sporadic *C. jejuni* case in the Manawatu Health District between March 1st 2005 and February 29th 2008. Outbreak-related cases, except the index case, were excluded. Data from the national public health surveillance system were obtained for cross-validation with notified case numbers in the study region.

3.3.3 Case ascertainment

Cases were prospectively recruited via the regional diagnostic laboratory, which tests all human pathological samples from the region. Samples were tested with a commercial enzyme immunoassay (ProSpecT, Remel, USA), and positive fecal swabs (Amies Carchoal transport swabs, Copan, Italy), were sent to the Hopkirk Research Institute, for isolation of presumptive *Campylobacter*. In brief, swabs were cultured on modified Cefoperazone Charcoal Desoxycholate agar plates (Fort Richard, Auckland, NZ) and in Bolton Broth (Lab M, Bury, England) and incubated in 42 °C in a microaerophilic atmosphere for two days. A single colony resembling *Campylobacter* species was subcultured to Columbia Horse Blood Agar (Fort Richard, Auckland, NZ) and incubated microaerophilically at 42 °C for two days before DNA preparations were made. Isolates were confirmed as *C. jejuni* by polymerase chain reaction. If multiple samples were collected from a single patient, only one isolate was selected.

3.3.4 Genotyping

After speciation, MLST of *C. jejuni* isolates was performed to assign isolates to a sequence type (ST) using seven house-keeping genes: *aspA* (aspartase A), *glnA* (glutamine synthase), *gltA* (citrate synthase), *glyA* (serine hydroxymethyltransferase), *pgm* (phosphoglucomutase), *tkt* (transketolase) and *uncA* (ATP synthase alpha subunit) based on the method outlined by Dingle et al., 2001 (Dingle, Colles et al. 2001).

3.3.5 Epidemiological surveillance

Three years of human data were acquired systematically - initially collected using routine Public Health Service methods, superseded by a sentinel surveillance approach with data collection targets. Case information between February 2005 and June 2006 was acquired using both mailed questionnaires and telephone interviews. To enhance the quality of the surveillance data gathered, from July 1st 2006 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 standard structured EpiSurv case report form used for enteric diseases in New Zealand (Lake, Whyte et al. 2005). The questionnaire gathered demographic variables, risk factor information for

variables within the incubation period, as well as case characteristics, such as whether the case was hospitalised. Exposure history covered aspects such as occupation, recent travel, and indicators of person-to-person transmission. Public health officers were trained to ensure consistency and standardization of interviews.

3.3.6 Data handling and statistical analysis

Epidemiological data collected was merged with the genotyping information in an Access database, which was linked into a Bionumerics database for analysis. Data was quality controlled and validated by techniques such as consistency check, range check and batch totals. R, version 2.7.0, was used for statistical analysis. Distributions were compared using two-sided χ^2 tests and Fisher's exact test for count data.

All STs isolated from human cases were assigned a probability that they were acquired from individual animal reservoirs using the asymmetric island model (Wilson, Gabriel et al. 2008), which is in detail discussed in Chapter 7. Data for this assignment was collected from parallel structured studies at the sentinel surveillance site over the same time-period (French and Molecular Epidemiology and Veterinary Public Health Group Hopkirk Institute; Mullner, Jones et al. 2009). STs with a probability of at least 70 % to have originated from a single source were assigned to that source, thereby creating four different groups of human cases: Poultry-associated ST - 474, other poultry-associated STs, ruminant associated STs and STs not associated with a particular source, of which the latter two were pooled in the analysis due to sample size considerations. The spatial location of each notified human case was given at the meshblock level: the smallest regions defined for the New Zealand census, each containing between zero and about 200 individuals. The spatial resolution is therefore much higher in urban areas. The relative risk of being a case of campylobacteriosis in each meshblock was described using a Bayesian hierarchical model. Relative risk surfaces were prepared for human cases attributable to the common poultry-associated STs, and for non-poultry associated STs.

Let $Y_{i,t}$ represent the number of notified cases of campylobacteriosis in meshblock i and week t. We assume that $Y_{i,t} \sim Poisson(n_i\lambda_{i,t})$, 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

$$U_i \sim N(\sum_{j \in n(i)} \frac{U_j}{\mid n(i) \mid}, \frac{1}{\kappa_u \mid n(i) \mid}), \tag{3.1}$$

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 $R_1, \ldots, R_t$,

$$R_{t+1} - R_t \sim (R_t - R_{t-1}, \frac{1}{\kappa_R}). \tag{3.2}$$

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 mildly informative and conjugate gamma-distributed priors: κ_R $Gamma(5, 1/100)$ and κ_U $Gamma(1, 1/2)$. These priors were chosen to aid convergence of the Markov chain, but to be still sufficiently flexible to allow the model to arrive at the appropriate level of variation for the data.

Multiple covariates were analysed using the defined source associations of strains and odds ratios (OR) were calculated. To reduce the number of covariates and the collinearity among different variable for farm animal exposure items measured on the questionnaire were summarized on biological grounds. The variable 'farm animal contact' therefore comprised 'occupational exposure to farm animals' and 'farm animal contact'. Age was modeled in two groups: five years and under and over five years of age. This was because differences in risk for pre-school children (children first attend school on their fifth birthday in New Zealand), compared to other age-groups, had been observed in other studies and was evident from surveillance data (http://www.nzpho.org.nz/). We constructed multivariable logistic regression models to investigate the relationship between infection with poultry associated strains and a set of exposure variables. Variables for inclusion in the final models were chosen by exploratory univariable analysis.

3.4 Results

3.4.1 Description of cases

A detailed examination of laboratory and surveillance data revealed that only 774 (75 %) of the 1036 swabs submitted between March 1st 2005 and February 29th 2008 could be considered 'primary samples' that originated in the Manawatu area. The remainders were duplicate samples, or samples from outside the study area. Of these 662 (86 %) were culture positive and 581 (88 %) were confirmed *C. jejuni*, of which 502 (86 %) provided MLST data, due to resource constraints (Figure 3.1). We assume, supported by a temporal analysis of the dataset, that the typed samples represent human cases

in the region over the study period. The comparison of demographic data among *C. jejuni* cases for which an isolate was genotyped and *C. jejuni* cases for which no isolate was genotyped revealed no significance difference between typed and not typed cases (Table 3.1). The response rate to the questionnaire, applied to all notified cases in the Manawatu Health district, increased after the introduction of data collection targets in July 2006 from 58 % to 97 % with a 96 - 100 % completion of case report data fields. A significant difference could be observed for the distribution of gender ($p = 0.04$) before ($n = 207$) and after ($n = 252$) the changes to the interview process, with an increase in male respondents as a result of the enhanced surveillance. The age distribution of respondents was not significantly different between the two periods ($p = 0.13$).

Table 3.1: Distribution of selected demographic characteristics among surveillance linked *C. jejuni* cases for which a bacterial isolate was genotyped (typed) and cases for which no isolate was genotyped (nontyped), Manawatu Health District, New Zealand 2005 - 2008.

	Typed (n = 459)†		Nontyped (n = 64)‡		
	No. of cases	%	No. of cases	%	p value*
Sex					
Male	248	54.3	40	63.5	
Female	209	45.7	23	36.5	0.21
Age group (years)					
<20	141	30.7	19	29.7	
20-50	184	40	32	50	
>50	134	29.2	13	20	0.23

* Two-sided p-value from a χ^2 test.
† Sex was unknown for 2 cases.
‡ Sex was unknown for 1 case.

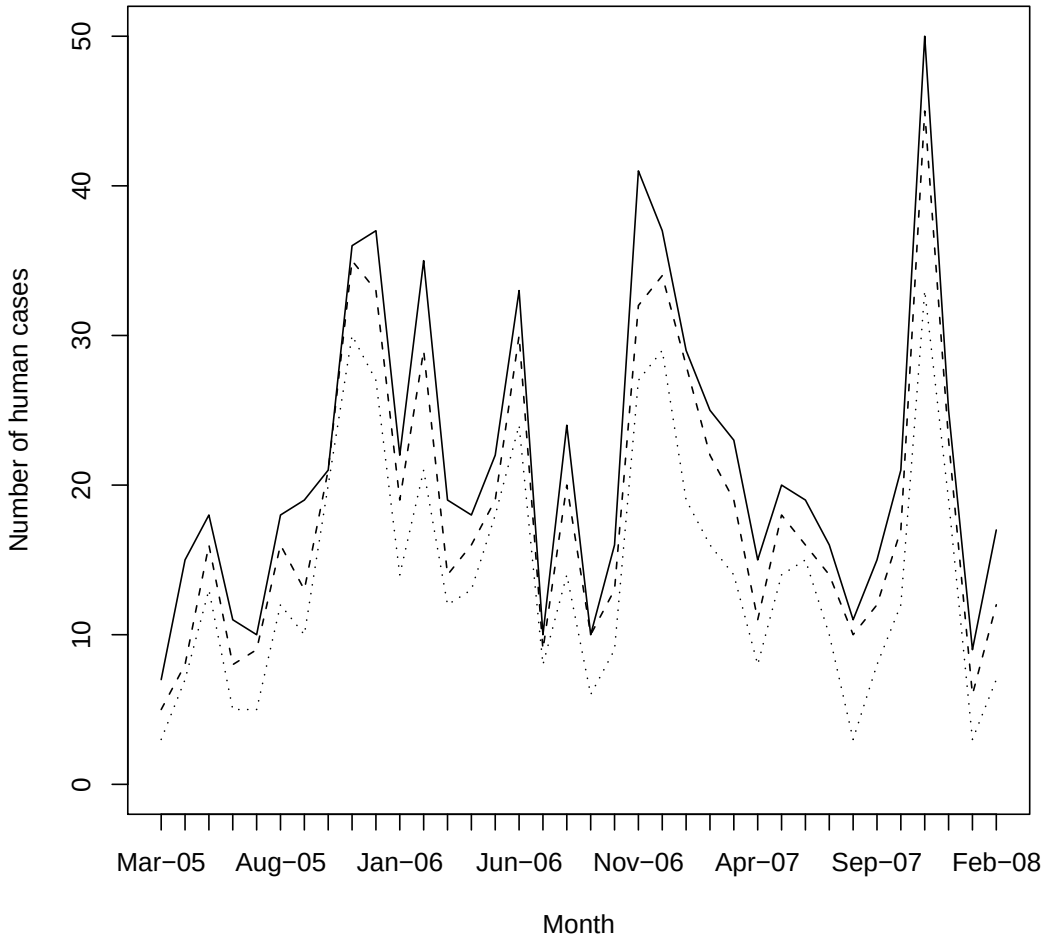

Figure 3.1: The number of ELISA positive (solid line), culture positive (dashed line) and genotyped (dotted line) human samples from the Manawatu surveillance site, New Zealand, in each month between March 2005 and February 2008.

3.4.2 Incidence

A total of 971 cases were notified in the region during the study period compared with 774 cases included in the laboratory recruited study. The number of cases reflected the number of notifications in time with only minor aberrations (Appendix B). Cases numbers increased during each southern hemisphere summer, with an additional peak of cases in winter in 2006. Annual notifications rate (per 100,000 population) in the region for the three years were 206 for the first year, 232 for the second and 188 in the final year of the study.

3.4.3 Distribution and source association of genotypes

A total of 51 different MLST genotypes were detected in our study. Seven STs each accounted for more than 20 cases and amounted to a total of 65.9 % of all cases (Table 3.2). The most dominant genotype was ST - 474 accounting for the largest proportion of human cases in most quarters, contributing a total of 154 cases (30.7 %) over the study period. Other major genotypes were ST - 48 and ST - 45, which could be isolated from 8.4 percent and 8.2 percent, respectively, of human cases. Common STs, occurring in over 20 cases, were ST - 53, ST - 50, ST - 190 and ST - 354. A total of 16 STs were poultry-associated including ST - 474, ST - 45, ST - 48, ST - 190 and ST - 354. Nine STs were associated with ruminant sources including ST - 61, ST - 38 and ST - 2026 (Table 3.2).

Table 3.2: Frequency and source association of human *C. jejuni* MLST genotypes, Manawatu Health District, New Zealand, 2005-2008.

Genotype	Absolute frequency	Relative frequency (%)	Source association*
ST-21	7	1.4	Ruminant
ST-25	1	0.2	Poultry
ST-38	13	2.6	Ruminant
ST-42	19	3.8	Ruminant
ST-45	41	8.2	Poultry
ST-48	42	8.4	Poultry
ST-50	23	4.6	None
ST-52	17	3.4	Poultry
ST-53	27	5.4	None
ST-61	14	2.8	Ruminant
ST-81	1	0.2	Ruminant
ST-137	1	0.2	None
ST-190	21	4.2	Poultry
ST-219	1	0.2	None
ST-257	12	2.4	Poultry
ST-354	23	4.6	Poultry
ST-403	1	0.2	Ruminant
ST-422	3	0.6	Ruminant
ST-436	4	0.8	Ruminant
ST-451	8	1.6	Poultry
ST-459	1	0.2	Ruminant
ST-474	154	30.7	Poultry
ST-520	8	1.6	Poultry
ST-578	1	0.2	Ruminant
ST-583	9	1.8	Poultry
ST-658	2	0.4	None
ST-677	5	1	None

Continued on next page

Table 3.2 – *continued from previous page*

Genotype	Absolute frequency	Relative frequency (%)	Source association*
ST-829	1	0.2	None
ST-1457	1	0.2	None
ST-1517	4	0.8	None
ST-1581	1	0.2	Poultry
ST-1707	1	0.2	None
ST-2026	11	2.2	Ruminant
ST-2219	1	0.2	None
ST-2343	1	0.2	None
ST-2345	3	0.6	Poultry
ST-2350	2	0.4	Ruminant
ST-2391	1	0.2	Poultry
ST-3072	1	0.2	Ruminant
ST-3222	1	0.2	Ruminant
ST-3538	1	0.2	None
ST-3676	2	0.4	None
ST-3711	2	0.4	None
ST-3712	2	0.4	None
ST-3715	1	0.2	None
ST-3717	1	0.2	Poultry
ST-3718	1	0.2	Poultry
ST-3720	1	0.2	None
ST-3727	1	0.2	None
ST-3784	1	0.2	Poultry
ST-3792	1	0.2	Poultry
Total	502		

* source association was determined by applying the island model.

3.4.4 Risk factor analysis

Of the 502 MLST typed cases 459 (90.4 %) could be linked to a surveillance report and thus provided the dataset for the risk factor analysis. This included 142 ST - 474 cases, 174 cases with other poultry associated STs, 48 cases with ruminant associated STs and 95 cases which could not be assigned to a probable source. The genotype distribution between linked and unlinked cases was compared using a two-sided chi-squared test and no significant differences between the two groups could be observed ($p = 0.21$). Except for gender and age, risk factor information was incomplete and 89 cases contained missing data and were omitted from the model. To evaluate possible sources of bias from missing data we investigated the relationship between missing values and potential risk factors. This analysis revealed no significant association for any of the risk factors considered. The relationship between age and ruminant versus poultry associated strains is shown in Figure 3.2. The difference in age profiles is significant using Fischer's exact test for count data ($p = 0.02$). Ruminant genotypes are relatively more common in children under 10 years of age compared to poultry genotypes. Table 3.3 shows results from the multivariable analysis of infection with poultry STs versus non-poultry strains. Farm animal contact was significantly associated with infection with non-poultry strains. The interaction between age-group and location was also found to be significant. Pre-school children (<5 years old) living in rural areas were significanty less likely to be infected with poultry genotypes (OR = 0.47; calculated as 1.5*1.64*0.19 = 0.47 from Table 3.3) and therefore more likely to be infected with non-poultry genotypes, compared to older age groups living in urban areas. In summary, this analysis indicated that poultry-associated cases were relatively more likely to be urban dwelling, >5 year old individuals with no farm animal contact. Conversely non-poultry associated cases were relatively more likely to be pre-school, rural dwelling individuals with contact with farm animals.

Table 3.3: Results of a multivariable case-case risk factor analysis for infection with poultry strains vs non-poultry strains, Manawatu region of New Zealand 2005 to 2008.

Exposure	OR	95 % CI	p value
Male gender	0.81	0.51−1.29	0.38
Age of 5 years and under	1.5	0.69−3.26	0.31
Living in a rural area	1.64	0.75−3.60	0.22
Age of 5 years and under and living in a rural area	0.19	0.05−0.81	0.02
Farm animal contact	0.42	0.26−0.69	<0.001

OR, odds ration; CI, confidence interval

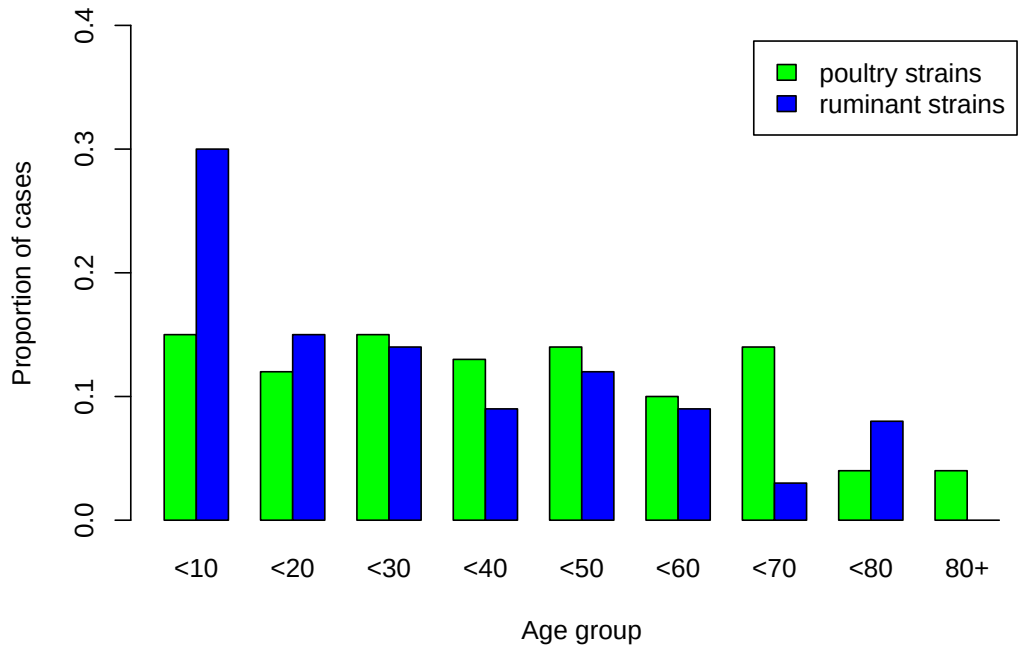

Figure 3.2: The proportion of cases associated with poultry (green bars, N = 316) and ruminant (blue bars, N = 66) genotypes in each age category.

3.4.5 Spatio - temporal analysis

Figure 3.3 shows the estimated smoothed number of cases per week attributable to ST - 474, other poultry STs and non-poultry STs. There is evidence of seasonality in all of the categories and the winter poultry-associated epidemic of 2006 (May - July 2006) is clearly visible. A decline in the weekly incidence of ST - 474 associated cases is evident. Figures 3.4, 3.5 and 3.6 show the relative risk surfaces for the different ST groups. Poultry associated cases are more prevalent in urban areas than rural areas, whereas for non-poultry-associated cases the reverse is true.

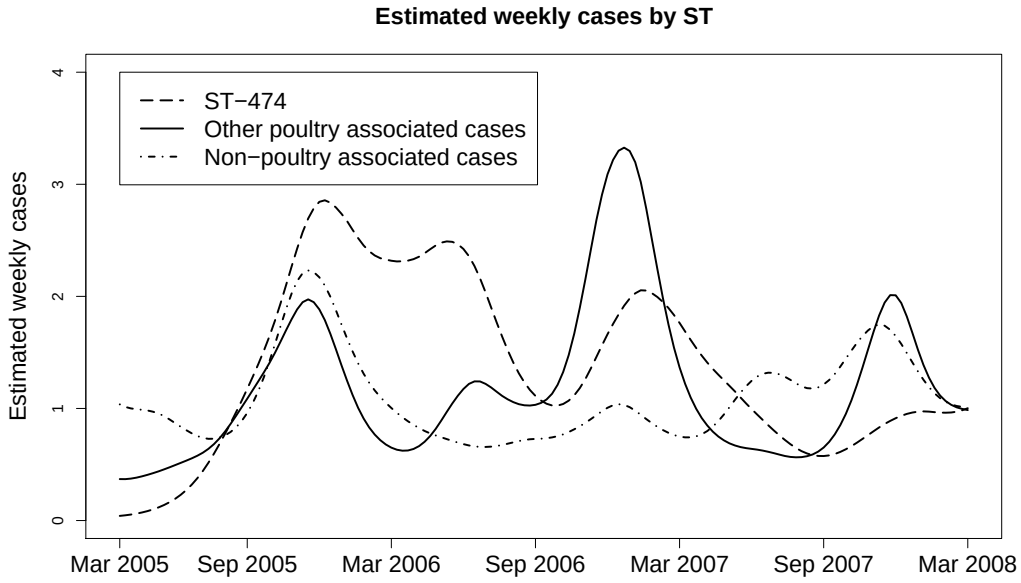

Figure 3.3: Smoothed temporal trends in human campylobacteriosis cases infected with ST- 474, poultry and non-poultry associated strains in the Manawatu between March 2005 and February 2008.

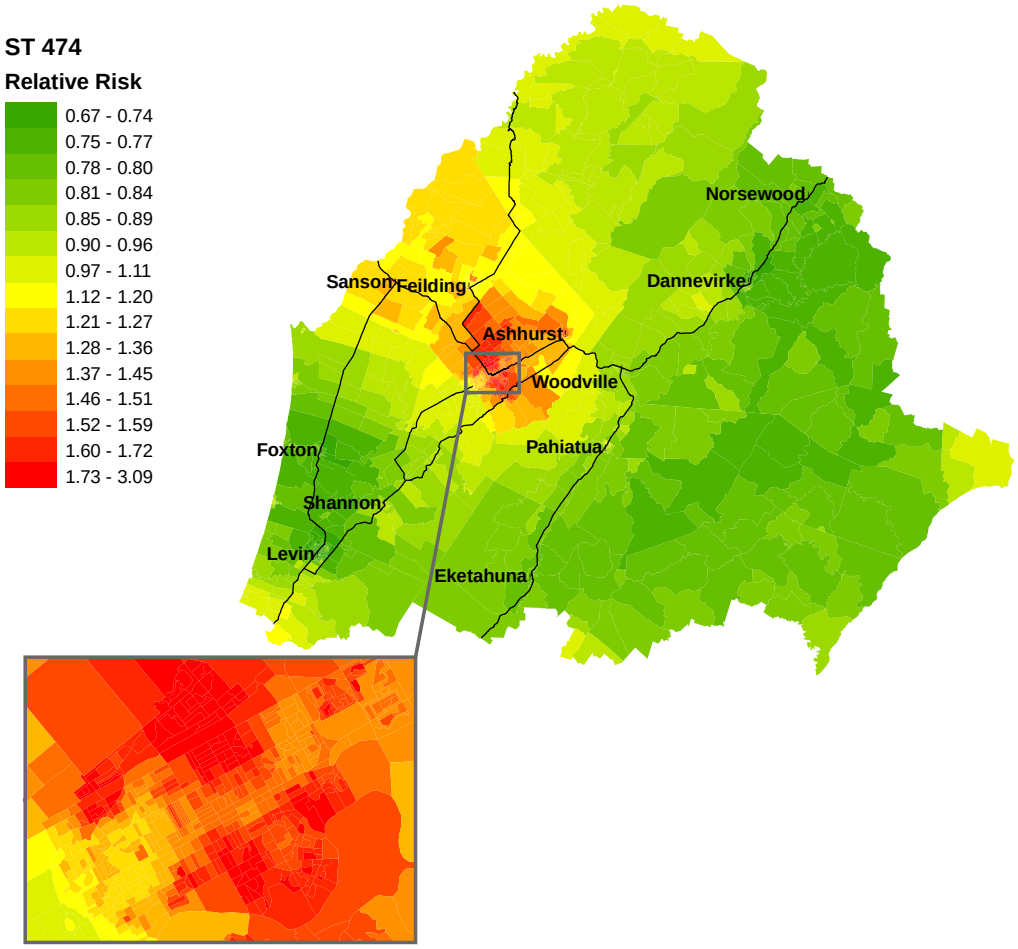

Figure 3.4: Relative risk surface of human campylobacteriosis cases of ST - 474 in the Manawatu between March 2005 and February 2008. The box shows an enlargement of the city of Palmerston North.

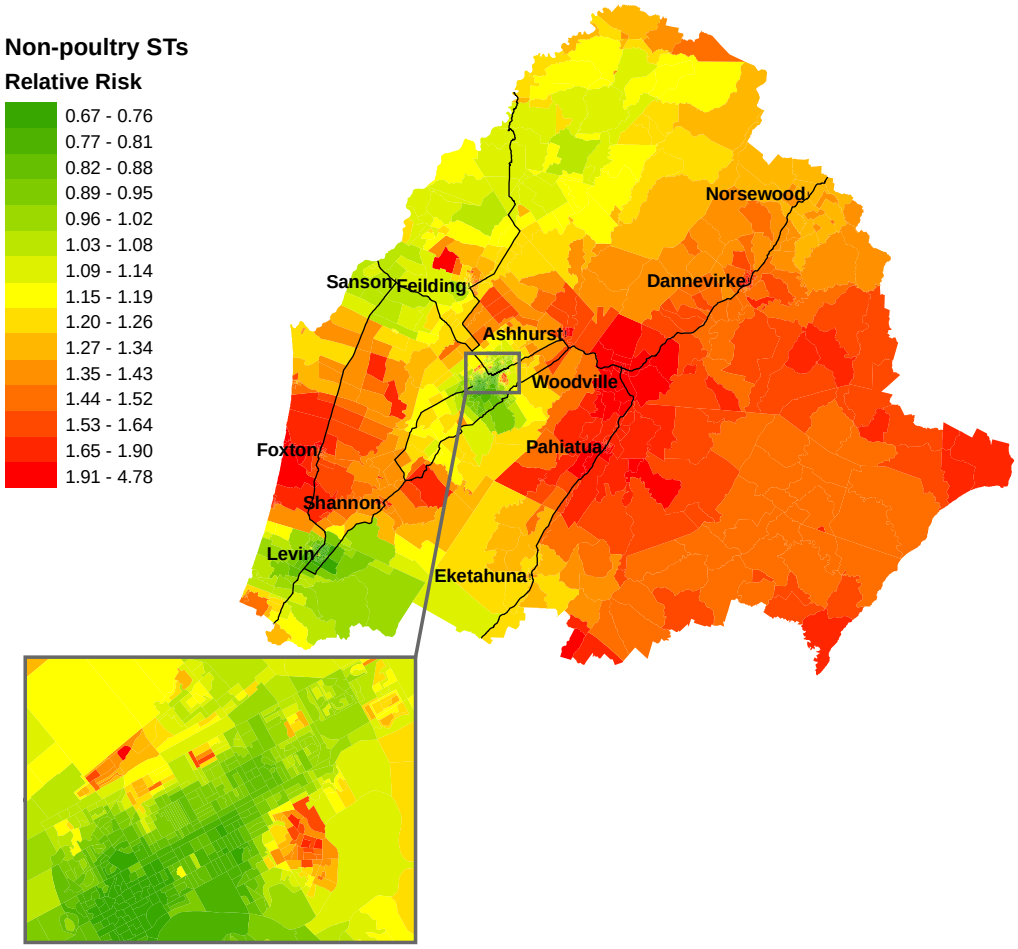

Figure 3.5: Relative risk surface of human campylobacteriosis cases of non-poultry strains in the Manawatu between March 2005 and February 2008. The box shows an enlargement of the city of Palmerston North.

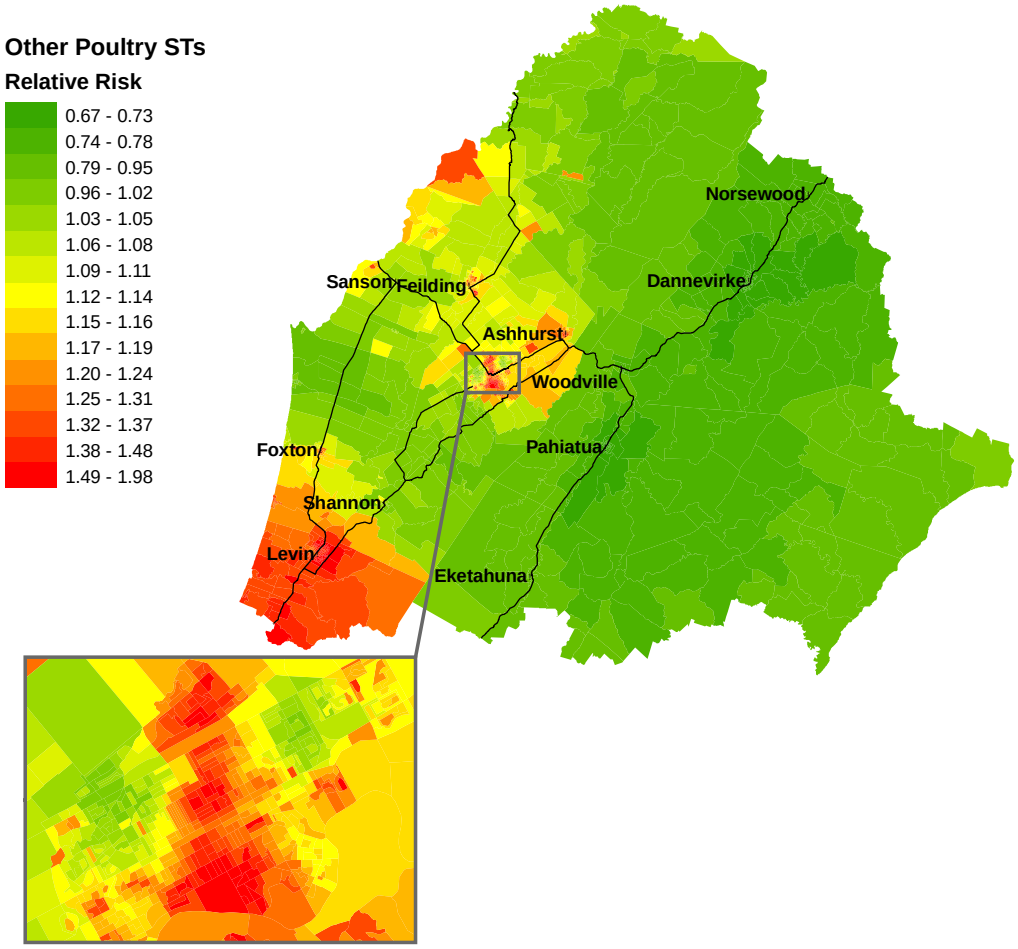

Figure 3.6: Relative risk surface of human campylobacteriosis cases of other poultry strains in the Manawatu between March 2005 and February 2008. The box shows an enlargement of the city of Palmerston North.

3.5 Discussion

Use of a sentinel surveillance site has increased our understanding of the epidemiology of campylobacteriosis in New Zealand. A concentration of resources in the defined site provided a rich source of information, and being more comprehensive than traditional passive surveillance approaches, included advantages such as (i) standardized microbiological and epidemiological methods, (ii) high quality data on individual cases as well as (iii) timeliness and cost effectiveness.

An estimate of the population at risk was available in this study, allowing incidence rates to be calculated. The Manawatu Health district was an area of moderate disease rate in the study period (188 - 232 cases per 100,000) compared with the rest of New Zealand, where in some regions and time periods more than 500 cases per 100,000 population were notified (Institute of Environmental Science and Research Limited (ESR)). From 2005 to 2007 for the whole of New Zealand notification rates ranged from 302 to 379 cases per 100,000 population. Rates in our study population were above those of comparable countries such as the U.S. (12.71 per 100,000), EU-25 (46.1 per 100,000) or Australia (111.8 per 100,000) (Centre for Disease Control and Prevention; European Food Safety Authority; Australian Government. Department of Health and Ageing 2008). Previous studies of the national notification system support that rates of campylobacteriosis in New Zealand are comparatively high (Baker, Sneyd et al. 2007). The observed 2006 nationwide increase in cases during the southern hemisphere winter (McTavish, Pope et al. 2008) is evident in our study as well as the expected seasonal increase in campylobacteriosis during the warmer months (Nylen, Dunstan et al. 2002).

A comparison of cases included in this study with actual cases statutorily notified revealed a good match. Case ascertainment in each month was close to the numbers of notifications. The figures are not identical as the notified cases contain a percentage of non-laboratory confirmed cases associated with outbreaks that were excluded in our case definition. This is highlighted in the data from November 2006 where the number of notifications deviates from the number of cases included in the study, as a result of a major outbreak, which included 26 outbreak-associated cases (McTavish, Pope et al. 2008). In addition differences between notifications and study cases may be caused by a delay between laboratory confirmation and notifications of cases, as well as technical problems such as lost or unidentifiable samples or sample submission to laboratories outside the study region.

Although we found over 50 different MLST sequence types in human samples in our study, the majority of human cases were caused by only seven different STs. The major human genotypes were generally associated with a consistent number of human cases over the time period and appear to be endemic to New Zealand. ST - 474 was the dominant genotype found in our study. Surprisingly this ST has only rarely been reported overseas (McTavish, Pope et al. 2008) and besides human cases has

almost exclusively been detected in samples from poultry in New Zealand (French and Molecular Epidemiology and Veterinary Public Health Group Hopkirk Institute). The link between poultry and human cases of campylobacteriosis has long been identified (Sheppard, McCarthy et al. 2007) and poultry sources have recently been attributed to cause the majority of campylobacteriosis cases in New Zealand (Mullner, Jones et al. 2009), Scotland (Sheppard, Dallas et al. 2009) and the United Kingdom (Wilson, Gabriel et al. 2008). These results are in detail presented in Chapters 5 and 7. Although this study found a dominant clone and several other STs, such as ST - 2345, that are to date only found in samples collected in New Zealand, many genotypes isolated in this study can be commonly found in other countries (Sopwith, Birtles et al. 2006; Gormley, MacRae et al. 2008). This includes major human pathogens such as ST - 45, ST - 48 and ST - 53.

The evaluation of risk factors revealed an increased risk of infection with poultry strains in urban areas and non-poultry strains in rural areas, with the second likely to be driven through direct contact with ruminant farm animals. In addition an interaction between young age and living in rural communities could be observed, resulting in an increased risk of infection with non-poultry strains. The differences depending on the area of location are supported by our spatio-temporal model, which also shows distinct temporal patterns of the different sequence type groups. The latter could indicate the dominance of different transmission pathways in time, such as more frequent animal contact in summer time or during the calving season, or a higher risk of infection with poultry strains during the summer months, due to for example a change in food handling practices over the warmer months. As a consequence different exposure pathways are likely to exist in rural and urban communities, and this will influence the development of control strategies.

This research suggests that the high rates of campylobacteriosis in New Zealand are likely to be a consequence of the dominance of an internationally rare clone, namely ST - 474, as well the presence of common human genotypes such as ST - 45. ST - 474 was the most commonly isolated genotype in the winter epidemic in 2006 in New Zealand (McTavish, Pope et al. 2008). To date this ST has only one submission to the *Campylobacter* PubMLST database (Dingle, Colles et al. 2005) from a chicken sample in the Czech Republic and has only been reported sporadically (Clark, Bryden et al. 2005; Best, Fox et al. 2007), outside New Zealand (French and Molecular Epidemiology and Veterinary Public Health Group Hopkirk Institute 2008; McTavish, Pope et al. 2008; Taboada, MacKinnon et al. 2008). Studies by Taboada and Pope indicated that some *Campylobacter* strains, including ST - 474, are of higher virulence, as measured by an invasion assey, for humans than others (Pope, Wilson et al. 2007; Taboada, MacKinnon et al. 2008). Further research into pathogenicity and virulence genes of this dominant clone could reveal determinants for human infection and com-

mensal colonization and may explain the dominance of this ST in New Zealand. The identification of ST-specific risk factors will facilitate the identification of tailor made control programs and by prioritizing dominating STs, more efficient control programs could be designed.

This study incorporated data from structured longitudinal studies to estimate risks, which are preferable to cross-sectional surveys. Not all cases included in the study were genotyped and the potential misclassification and differences between characteristics of patients with and without genotypes isolates present potential sources of bias. However our analysis revealed no significant differences between the two groups, and we therefore consider the typed cases to be broadly representative of all notified cases from this region. The selection of cases through the diagnostic laboratory is potentially biased as cases providing a stool sample are more likely to be severe. In addition the gathering of information on cases via questionnaires is prone to bias for reasons such as differences in exposure between respondents and non-respondents. Initially the response to the questionnaire was only 57 %, but increased to 97 % in the second phase of the project. Although we observed that this included a higher percentage of male respondents, which was not unexpected (Eaker, Bergstrom et al. 1998), this is unlikely to invalidate the conclusions drawn from this study. Moreover the above average response rate in this study has overall reduced the bias due to non-response, which is inherent in the approach of collecting data via questionnaires. No evidence was found that cases without a linked surveillance report were different from cases with a report, and we therefore conclude that cases included in the analysis are representative of the study population.

Instead of using a network of surveillance sites this study used only one region and the importance of transmission routes is likely to vary regionally (Hearnden, Skelly et al. 2003). However the generalisation of our results from the sentinel surveillance site to the general population is supported by validation studies in two other regions where a similar distribution of MLST types was observed (French and Molecular Epidemiology and Veterinary Public Health Group Hopkirk Institute 2008). New Zealand provides a unique ecology and this is likely to have influenced the national distribution of MLST genotypes and the epidemiology of human cases. However some of the results from this study compare with those made in a similar study in Scotland (Strachan, Gormley et al. 2009), such as the differences in risks of infection in rural and urban communities.

Comparative source attribution studies from the Manawatu sentinel surveillance site provide strong evidence that poultry is the leading cause of human campylobacteriosis in New Zealand, causing an estimated 58-74% of cases (Chapter 7), with varying contributions from different poultry suppliers. In contrast, environmental sources played a relatively minor role. This is supported by the dramatic decline in human notified cases in 2008, which is a likely consequence of industry-specific interventions that were

implemented based on these findings. The analysis presented here complements those results by improving our knowledge of risk factors and transmission pathways to implement more refined control strategies; for example to protect pre-school children from disease through farm animal contact in rural communities and focusing on transmission through poultry products in urban areas. In conclusion, the combination of molecular and epidemiological tools in a sentinel surveillance site has improved our understanding of the epidemiology of campylobacteriosis in New Zealand. The analysis of risk factors at the molecular level has revealed disease characteristics that will influence the design and implementation of control measures. The results from this study underline the importance of studying *Campylobacter* spp. beyond the strain level. Extending this work to study risk factors in space and time in more detail could prove valuable to improve our understanding of the commonly observed regionality and seasonality of *Campylobacter* infections.

3.6 Acknowledgements

I am thankful to all members of the Hopkirk Molecular Epidemiology Team, Environmental Science and Research (ESR), MidCentral Health, Public Health Services, NZFSA (Science Group), and MedLab Central (in particular Lynn Rogers) for their contributions. The spatio-temporal model in this chapter was developed by Simon Spencer (University of Nottingham), who also provided the main contribution to their graphical outputs. The island model results for this manuscript were derived by Jonathan Marshall (Massey University).

3.7 Appendices

Appendix B provides further details about the data and methods applied in this study.

Chapter 4

Molecular epidemiology of *Campylobacter jejuni* in a geographically isolated country with a uniquely structured poultry industry

Petra Mullner, Julie M. Collins-Emerson, Anne J. Midwinter, Philip Carter, Simon E. F. Spencer, Peter van der Logt, Steve Hathaway, Nigel French. Molecular epidemiology of *Campylobacter jejuni* in a geographically isolated country with a uniquely structured poultry industry. Manuscript submitted to *Applied and Environmental Microbiology* on 16th April 2009.

4.1 Abstract

In New Zealand the number of campylobacteriosis notifications increased markedly between the years 2000 to 2007. To examine the link between human cases and poultry a sentinel surveillance site was established to study the molecular epidemiology of *C. jejuni* over a three year period from 2005 to 2008 using multilocus sequence typing. These studies showed that 60.1 - 81.4% of retail poultry carcasses from the major suppliers were contaminated with *C. jejuni*. Differences were detected in the probability and level of contamination and the relative frequency of genotypes in individual poultry suppliers and humans. Some carcasses were contaminated with more than one sequence type (ST) and there was evidence of both ubiquitous and supplier-associated strains; an epidemiological pattern not recognised yet in other countries. The common poultry STs were also common in human clinical cases, thus providing further evidence that poultry is a major contributor to human infection. Both internationally-rare genotypes, such as ST - 3069 and ST - 474, and common genotypes, such as ST - 45 and ST - 48, were

isolated in this study. The dominant human sequence type in New Zealand, ST - 474, was almost exclusively found in isolates from one poultry supplier, providing evidence of a distinctive molecular epidemiology of *C. jejuni* in this country. These results may in part be due to the country's geographical isolation and its uniquely structured poultry industry. Based on this study the national regulatory body implemented industry specific control measures and as a likely consequence of these interventions a substantial drop in human case notifications was observed in 2008.

4.2 Introduction

Campylobacteriosis is the leading enteric zoonosis in the developed world and the majority of cases are caused by *Campylobacter jejuni* (Gormley, MacRae et al. 2008). Poultry sources are suspected to be the major source of human infection with *C. jejuni* (Wagenaar, Mevius et al. 2006; Gormley, MacRae et al. 2008; Sheppard, Dallas et al. 2009) and this is supported by high contamination levels of poultry and the detection of identical *C. jejuni* genotypes in human cases and poultry samples (Kramer 2000). However *C. jejuni* can be isolated from a variety of sources including ruminants (French, Barrigas et al. 2005) and environmental water (Savill, Hudson et al. 2001). Due to the complex epidemiology of this pathogen there is still uncertainty about the relative contribution of individual pathways to the human disease burden (Cowden 1992).

In New Zealand notified campylobacteriosis cases have increased dramatically in the last decade, peaking in 2006 with a total of 15873 notified cases (422.4 per 100,000 population) (Baker, Wilson et al. 2006) and costing the economy an estimated US$ 32 million annually (Scott, Scott et al. 2000). New Zealand has one of the highest enteric infectious diseases rates in industrialised countries (Lake, Baker et al. 2000) and the high ratio of domestic production animals to humans as well as the frequent use of rural water supplies have been postulated as some of the underlying causes (Crump, Murdoch et al. 2001). In addition, the country's poultry industry is uniquely structured: it is almost entirely focused on the domestic market, and no raw poultry products are imported into the country because of biosecurity threats. Due to the geographical isolation and tight border controls New Zealand has remained free of poultry diseases endemic in other countries such as *Salmonella* Enteritidis PT4, *Salmonella* Typhimurium DT 104, Newcastle Disease, and Infectious Bursal Disease. The production of poultry meat in New Zealand is highly integrated; only three companies supply 90% of chicken meat, which represents 95% of poultry meat consumed. These processors own or control most stages of production, processing and distribution. Of these three, one company has one processing plant that distributes nationwide, one company has multiple plants that tend to be more localised in their distribution, except when they make speciality products, which are distributed nationally. The remainder of the companies distribute primarily locally.

To enable regulators to implement food safety programs to reduce human campylobacteriosis, it had become of great interest to better understand the importance and epidemiology of *C. jejuni* in the New Zealand poultry production system. A sentinel surveillance site was therefore established in the Manawatu region to quantify the contribution of different sources, including poultry suppliers, to the human disease burden and to study the molecular epidemiology of *C. jejuni* (French and Molecular Epidemiology and Veterinary Public Health Group Hopkirk Institute 2008). Isolates were typed using multilocus sequence typing (MLST), a method that has major advantages over other methods of genotyping. Further this method offers a large web-based archive of isolates from many sources and countries: the *Campylobacter* PubMLST database (Dingle, Colles et al. 2005). Sequence typing by MLST is now internationally recognised as a valuable approach for national and international epidemiological characterisation of major pathogenic microorganisms such as *C. jejuni* (Urwin and Maiden 2003; Sullivan, Diggle et al. 2005).

In this study data from the sentinel surveillance site was used to study the importance of the poultry sector, and individual producers, as a contributor to the human disease burden of campylobacteriosis in New Zealand. Contamination levels of poultry products with *C. jejuni* were investigated, and human and poultry MLST sequence types (STs) were compared to make inferences about poultry as a source of human infection in New Zealand.

4.3 Material and Methods

4.3.1 *Campylobacter* isolates

Human specimens submitted to MedLab Central, Palmerston North that were positive for *Campylobacter* by ELISA (ProSpecT R, Remel, USA) were sent to the Hopkirk Molecular Epidemiology Laboratory over a three year period from 1st March 2005 to 29th February 2008. Faecal swabs were made using Amies Charcoal transport swabs (Copan, Italy). Over the same period 12-18 fresh whole poultry carcasses, representing the different poultry suppliers in the region, were sampled each month from retail outlets in Palmerston North, with the number of samples collected per supplier reflecting market share. The Manawatu region is predominantly supplied by three companies, the regional plants of Supplier A[1] , and through the nationwide distribution of products from Supplier B, which are dominating the supply with fresh poultry carcasses. In addition a localised company (Supplier C) also distributes fresh whole carcasses product in the region.

[1]Supplier identification letters were assigned arbitrarily

4.3.2 Bacterial culture and identification

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 microaerophilic atmosphere (85% N_2, 10% CO_2, 5% O_2) for two days. A single colony resembling *Campylobacter* was subcultured to Blood Agar (BA) (Fort Richard, Auckland) and incubated microaerophilically at 42 °C for two days before DNA preparations were made. In addition in a small exploratory study multiple colonies (2-5 colonies per sample) were subcultured from individual samples. Cultures were frozen at −80 °C in Glycerol Broth (Difco, USA). Chickens were washed and massaged in 200 mL of Buffered Peptone Water (BPW) (Difco, USA). The chicken wash was centrifuged (16,264 RCF (g), 6 °C, 35 mins, Sorvall RC5B) and the resultant pellet resuspended in 5 mL of BPW. Approximately 3 mL of the resuspended pellet was added to 90 mL of Bolton Broth, which was incubated at 42 °C microaerophilically for two days. After incubation, the broth was subcultured onto mCCDA and incubated microaerophilically at 42 °C for two days. Single colonies resembling *Campylobacter* species were subcultured to BA and incubated microaerophilically at 42 °C for two days before DNA preparations were made. Cultures were frozen at −80 °C. Isolates of *Campylobacter* were speciated by PCR to detect genes associated with either *C. jejuni* or *C. coli*. The *mapA* gene was shown to be found only in *C. jejuni* (Stucki et al, 1995), so primers MapA-F (5'-CTTGGCTTGAAATTTGCTTG-3') and MapAR (5'-GCTTGGTGCGGATTGTAAA-3') were designed to target this gene for its identification. Amplification was performed in a 20 μl reaction containing one unit of Platinum Taq Polymerase (Invitrogen), 100 μ M of each dNTP, 200 nM of each primer (MapA-F, MapA-R, COL3, and MDCOL2), and 1.5 mM $MgCl_2$. 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 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 603 bp product indicated *C. jejuni*.

4.3.3 Sequence typing

After speciation, MLST of *C. jejuni* isolates was performed using seven house-keeping genes: *aspA* (aspartase A), *glnA* (glutamine synthase), *gltA* (citrate synthase), *glyA* (serine hydroxymethyltransferase), *pgm* (phosphoglucomutase), *tkt* (transketolase) and *uncA* (ATP synthase alpha subunit) based on the method as outlined by Dingle et al.(2001). Chromosomal DNA was prepared from freshly grown cultures by boiling for 10 min followed by centrifugation of the disrupted cells. The supernatant was decanted to a fresh tube and used for amplification. The amplifications were per-

formed in a 25 μl volume reaction using Applied Biosystems AmpliTaq Gold mastermix (Applied Biosystems, Auckland New Zealand) and 5 pmoles of each primer. Products were sequenced on an ABI 3130XL automated DNA sequencer using ABI BigDye v3.1 (Applied Biosystems) following the manufacturer's instructions. Sequence data was collated 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. Alleles that did not give clear results were re-amplified and sequenced using primers sets published by Miller, Peterson et al., (2005) using the same protocol as above. Enumeration of *Campylobacter* on poultry carcasses. From October 2006, *Campylobacter* on all chicken carcasses were enumerated using a Wasp spiral plater (Don Whitley, England) and a manual spread plate. Duplicate
mCCDA plates were inoculated with 50 μl (spiral plater) or 1ml (spread plate) aliquots of chicken wash and 100 μl (spiral plater) aliquots of resuspended chicken wash pellet. The plates were incubated microaerophilically at 42 °C for 2 days. Colonies were counted manually or by using a plate reader (aCOLyte, Synbiosis, England).

4.3.4 Statistical analysis of *Campylobacter* count data

The aim was to estimate both the proportion of carcasses positive and the levels of *Campylobacter* present on positive carcasses. 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. Bacterial count data have two components: firstly whether bacteria are present or absent, and secondly how many are present. In order to capture these two components in our model we used a Bayesian zero-inflated Poisson model (Robinson, Brown et al. 2009). The model features two linear predictors: one informing the probability that the bacteria will be present and the other informing the number of bacteria that will be obtained given that they are present. For the zero-inflation part of the model we assumed that sample s was positive for *Campylobacter* with probability p_s. Since p_s is a probability, the logit link function was used when forming the linear predictor. We wished to model the change in the probability of contamination quarter-by-quarter for each company, and so we included a random effect for each quarter/company pair. Thus, $logit(p_s) = \alpha_0 + \alpha_{c(s),q(s)}$, 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:

- $\alpha_0 \sim N(0, 10)$
- $\alpha_{c,q} \sim N(0, \kappa_\alpha^- 1)$
- $\kappa_\alpha \sim Gamma(2, 2)$

The second part of the model quantifies 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. $Y_i \sim Pois(V_i, \mu_i)$, 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 used the log link function and included random effects for each company / quarter pair as before. To attempt to separate the laboratory variation from the sample variation, we also included 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 assumed weakly informative priors:

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

The output from these models is presented as a series of graphs describing the probability of a carcasses containing *Campylobacter*, by supplier and by quarter, and the estimated number of viable *Campylobacter* on positive carcasses - again by supplier and quarter. This method ensures that all the individual replicate counts for each sample are analysed appropriately.

4.3.5 Statistical Analysis of ST distributions

The proportional similarity index (PSI) or Czekanowski index is an objective and simple measure of the area of intersection between two frequency distributions (Rosef, Kapperud et al. 1985) that can estimate the similarity between the frequency distributions of bacterial subtypes from different sources. The PSI is calculated by: $PS = 1 - 0.5 \sum_i \mid p_1 - q_i \mid = \sum_i min(p_i, q_i)$, where p_i and q_i represent the proportion of strains belonging to type i out of all strains typed from sources P and Q (Feinsinger, Spears et al. 1981; Rosef, Kapperud et al. 1985). The value for PSI ranges between one for identical frequency distributions, to zero for distributions with no common types. Bootstrap confidence intervals for this measure were estimated (Garrett, Devane et al. 2007). STRUCTURE 2.2 (Pritchard, Stephens et al. 2000; Falush, Stephens et al. 2007) was used as a frequency-based clustering model to investigate the population structure of *C. jejuni* in the New Zealand poultry supply. This Bayesian MCMC

method attempts to assign individuals to clusters on the basis of their genotypes, while simultaneously estimating allele frequencies for each cluster. This approach assumes a model with K populations, each of which is characterised by a set of allele frequencies at each locus and assumes that loci are unlinked and at equilibrium with one another within populations (Pritchard, Stephens et al. 2000). Following the approach taken by Sheppard, Dallas, et al (2009) we further assumed the NO ADMIXTURE model and that the allele frequencies were uncorrelated. Since we are interested in exploring the structure of the distribution of STs between the three poultry companies, we only considered the case in which there are three clusters (K=3) and we used the source of the isolates as prior information for the model. We set the parameter
GENSBACK=1 - this parameter represents the number of generations back that Structure will consider that a sample may have had an ancestor from a different cluster. The model was run with a burn-in period of 100,000 iterations and thereafter sampling of the chain occurred over a further 100,000 iterations. The posterior probabilities of the final model were visualised as confluent, stacked bar plots using the *Distruct* program. Subpopulations are represented as different shades, and individuals are depicted as bars partitioned into shaded segments that correspond to the posterior probabilities of assignment (Rosenberg 2004).

4.4 Results

Retail poultry samples had a high prevalence of contamination with *Campylobacter* spp. throughout the three year period. Overall some 78.9% of carcasses were positive for *Campylobacter*, and 72.7% were positive for *C. jejuni*, with estimates of prevalence of *C. jejuni* for the individual suppliers ranging from 60.1 to 81.4% (Table 4.1). The estimated proportion of positive carcasses (Figure 4.1) and the counts conditional on being positive (Figure 4.2) are shown for supplier A, B, and C and each quarter from October 2006 to February 2008. There was a moderately high probability of contamination for all suppliers throughout the study period, with over 80% of carcasses from Supplier A being positive throughout 2007, and over 60% of carcasses positive for most suppliers in most quarters (4.1). The average number of *Campylobacter* colony forming units (cfu) on positive carcasses was initially estimated to be high (over 104 cfu/carcass) in Supplier A, but declined steadily over the study period, whereas the levels in Suppliers B have remained at a relatively constant but lower level throughout, and those for Supplier C declining since mid 2007 from a medium to a low level of contamination (Figure 4.2).

A total of 62 different STs were detected in human and poultry samples. The four major human STs were ST - 474, ST - 48, ST - 45 and ST - 53, accounting for 30.7%, 8.4%, 8.2% and 5.4% of human cases respectively (Table 4.2). The dominant human ST, namely ST - 474, was very prevalent in samples from Supplier A, but was

rarely detected in samples from other poultry suppliers. Other STs such as ST - 45 and ST - 48 occurred frequently in human samples and poultry samples from all three suppliers. Of the 62 STs detected 51 were found in human samples, 18 in samples from Supplier A, 17 in samples from Supplier B and 11 in samples from Supplier C. Twenty-nine human STs, representing 14.5% of all human cases, were not found in any of the poultry samples. Supplier A had seven unique STs, Supplier B had ten, and Supplier C five (Table 4.2). ST - 3609, which was unique to Supplier B, was isolated from 12.6% of this supplier's positive samples but has not been isolated from humans. This highly prevalent ST is closely related to ST - 48, a common human pathogen, differing only by one point mutation in the gltA gene. To date ST - 3609 has only been found in New Zealand.

One of the ten humans cases, from which two isolates were typed, was infected with two different STs. Between two and five isolates were typed from a total of 58 poultry samples, and (up to of three) different STs were found in 22 individual samples. Different ST's were isolated from 6 out of 16 samples from Supplier A, 24 out of 28 from Supplier B and 2 out of 4 samples from Supplier C (Table 4.3).

The PSI was calculated to explore formally the similarity of genotypes from human cases and the different poultry suppliers. This analysis revealed a significantly higher degree of similarity between STs identified in human cases and genotypes from Supplier A (PSI = 0.58, 95% CI 0.48 - 0.64) in comparison with the other suppliers (Table 4.4). Isolates from the other suppliers were less similar to human isolates with an estimated PSI of 0.32 (95% CI 0.26 - 0.36) and 0.28 (95% CI 0.21 - 0.31) respectively. When comparing similarity between poultry suppliers, the greatest similarity was observed between Suppliers A and C (PSI=0.46, 95% CI 0.27-0.50) and the lowest similarity was observed between Suppliers A and B (PSI=0.26 95% CI 0.18-0.35).

The population structure of *C. jejuni* in New Zealand poultry suppliers was explored using STRUCTURE 2.2. The posterior probabilities for each isolate are presented as a stacked bar plot in Figure 4.3 where the stacked bars represent the probability of each individual isolate belonging to each of the three clusters. This figure allows a visual assessment of the three clusters against the three different poultry suppliers. Isolates from Supplier A showed the most diversity, with isolates having reasonable probability of originating from any of the three clusters (0.611, 0.154, 0.234 for clusters 1 to 3 respectively). On the other hand isolates from Supplier B showed the highest probability of belonging to cluster 2 (0.845), with minor contributions from the clusters 1 and 3 (0.065, 0.09 respectively). Supplier C was similar in its diversity to Supplier A, with probabilities of originating from clusters 1,2 and 3 of 0.195, 0.149 and 0.657, but had the highest probability in cluster 3 instead of cluster 1.

Table 4.1: Proportion of poultry samples positive for *Campylobacter* spp. in the Manawatu March 2005 - February 2008.

	Number of samples	Number culture positive	Prevalence (%) presumptive *Campylobacter*[a]	Number confirmed *C. jejuni*	Estimated prevalence (%) in *C. jejuni*
Supplier A	239	203	84.9	189	80.7
Supplier B	196	136	69.4	117	60.1
Supplier C	65	55	84.6	50	81.4
All Suppliers	500	394	78.9	356	72.7

[a] estimate based on proportion of speciated samples that were *C. jejuni*.

Table 4.2: Relative frequency of sequence types (STs) in samples in% and total number of isolates from human cases and poultry suppliers in the Manawatu.

ST	Human cases	Supplier A	Supplier B	Suppliers C
21	1.4	0	0.8	0
25	0.2	1.5	0	5.7
38	2.6	0	0	0
42	3.8	3.1	0	0
45	8.2	27.5	12.6	25.7
48	8.4	0.8	22.7	20
50	4.6	3.8	17.6	0
52	3.4	4.6	0	0
53	5.4	7.6	3.4	8.6
61	2.8	0	0	0
81	0.2	0	0	0
137	0.2	0	0	0
190	4.2	6.1	0	0
219	0.2	0	0	0
227	0	0.8	0	0
257	2.4	7.6	3.4	0
354	4.6	2.3	0	0
403	0.2	0	0	0
422	0.6	0	0	0
436	0.8	0	0	0
451	1.6	0	0	5.7
459	0.2	0	0	0
474	30.7	20.6	1.7	0
520	1.6	3.8	0	5.7
578	0.2	0	0	0
583	1.8	6.1	0	11.4
658	0.4	0	0	0

Continued on next page

Table 4.2 – *continued from previous page*

ST	Human cases	Supplier A	Supplier B	Supplier C
677	1	0	0.8	0
829	0.2	0	0	0
1457	0.2	0	0	0
1517	0.8	0	5.0	0
1581	0.2	0	4.2	0
1707	0.2	0	0	0
1818	0	0	0.8	0
1911	0	0	0.8	0
2026	2.2	0	0	0
2219	0.2	0	0	0
2343	0.2	0	0	0
2345	0.6	0.8	9.2	0
2350	0.4	0	0	0
2391	0.2	0.8	0	0
2397	0	0	2.5	0
2535	0	0	0	2.9
3072	0.2	0	0	0
3222	0.2	0	0	0
3230	0	0	0	2.9
3538	0.2	0	0	0
3609	0	0	12.6	0
3676	0.4	0	0	0
3711	0.4	0	0	2.9
3712	0.4	0	0	0
3715	0.2	0	0	0
3717	0.2	0	0	8.6
3718	0.2	0	0	0
3719	0	0	0.8	0
3720	0.2	0	0	0
3721	0	0.8	0	0

Continued on next page

Table 4.2 – *continued from previous page*

ST	Human cases	Supplier A	Supplier B	Supplier C
3725	0	0	0.8	0
3726	0	1.5	0	0
3727	0.2	0	0	0
3784	0.2	0	0	0
3792	0.2	0	0	0
Isolates	502	131	119	35

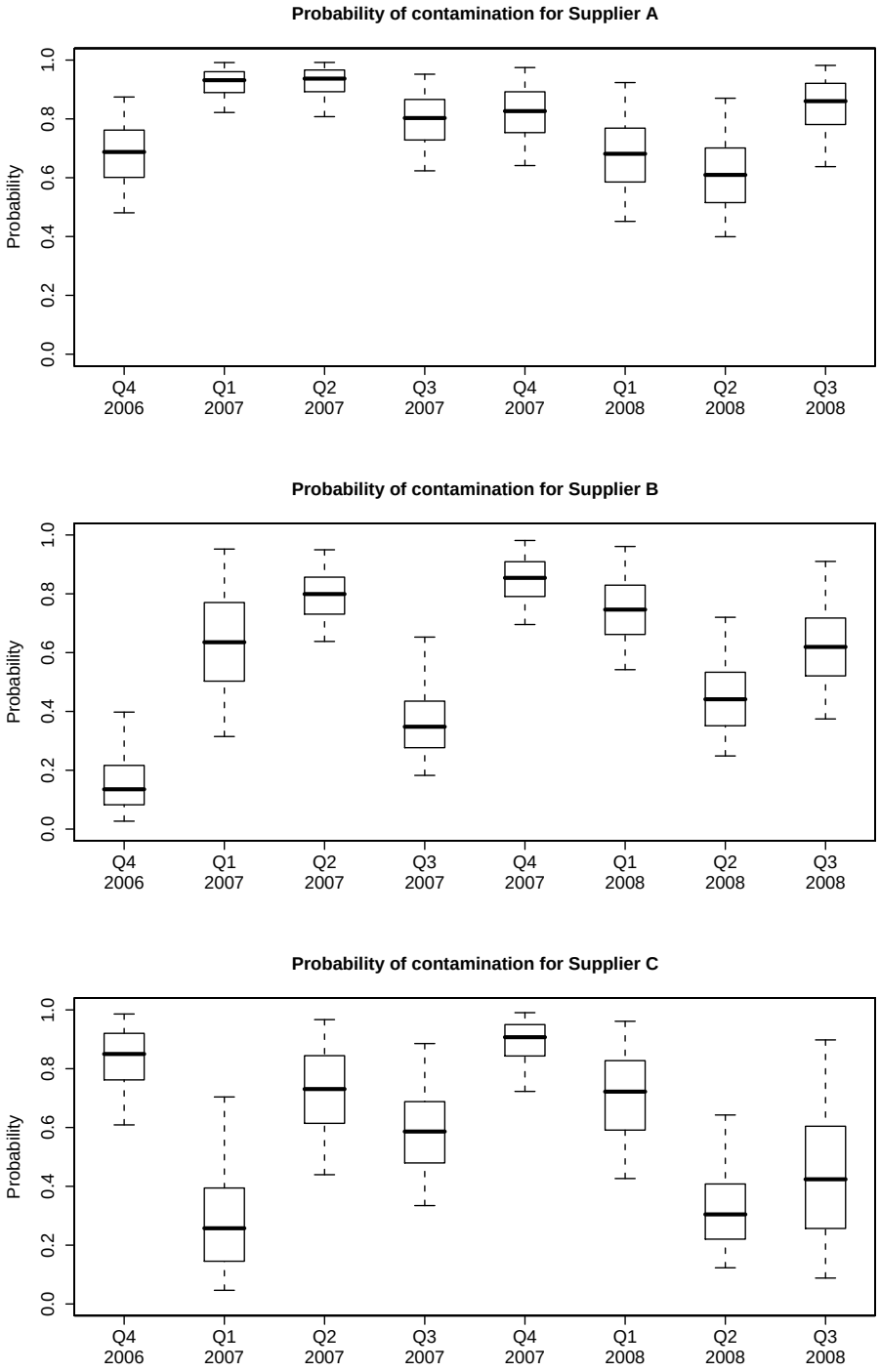

Figure 4.1: *Campylobacter* on poultry carcasses by quarter: probability of contamination for each supplier

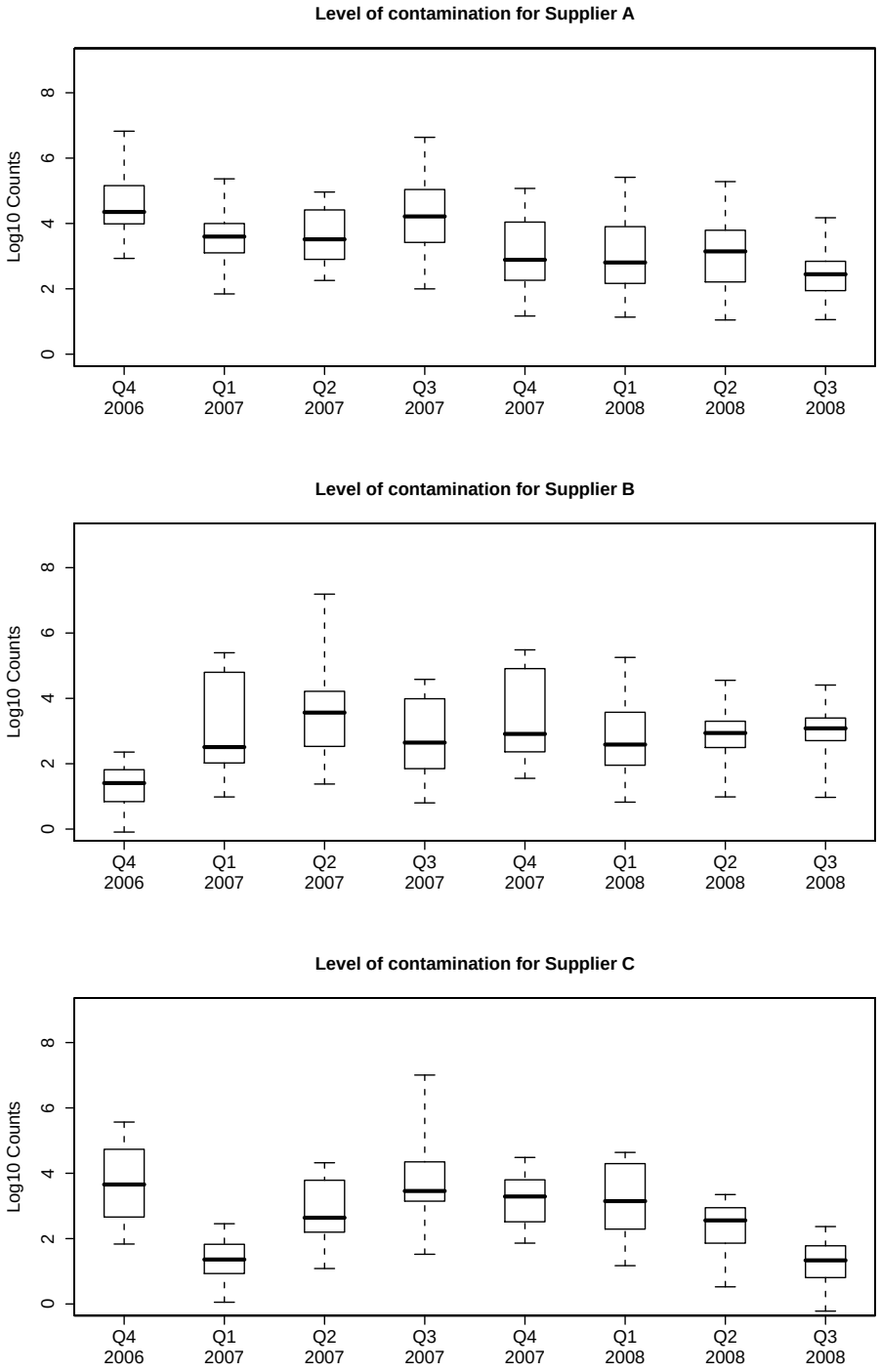

Figure 4.2: *Campylobacter* on poultry carcasses by quarter: level of contamination on positive carcasses for each supplier

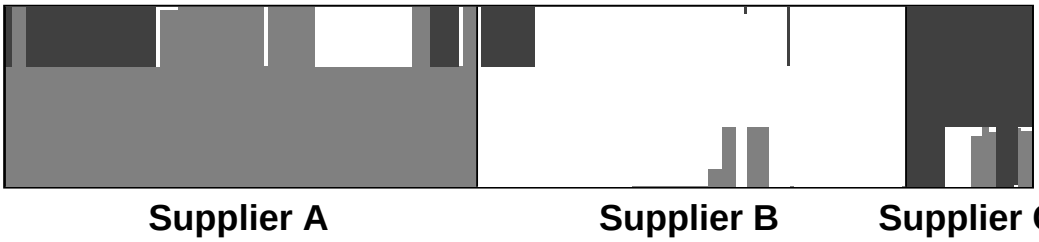

Figure 4.3: Population structure of *C. jejuni* in different poultry suppliers estimated by cluster analysis using STRUCTURE 2.2. Posterior probabilities of each isolate belonging to cluster 1 (light grey) cluster 2 (white) and cluster 3 (dark grey). Isolates are grouped by suppliers, and ordered by ST.

Table 4.3: Multiple typed samples.

	Human	Supplier A	Supplier B	Supplier C
Samples with multiple STs	1/10	6/16	24/38	2/4
Proportion of multiple STs	0.1	0.38	0.63	0.5
Maximum number of different STs found in one sample	2	2	3	2

Table 4.4: The Proportional Similarity Index[a] for each supplier compared with each other and to the distribution of human genotypes, with 95% bootstrapped confidence interval. .

	Supplier A	Supplier B	Supplier C
Supplier A	1	-	-
Supplier B	0.26 (0.18, 0.35)	1	-
Supplier C	0.46 (0.27, 0.50)	0.36 (0.22, 0.44)	1
Human cases	0.58 (0.48, 0.64)	0.32 (0.26, 0.36)	0.28 (0.21, 0.31)

[a] Higher values indicate a strong similarity between the STs identified in the different sources.

4.5 Discussion

In New Zealand the prevalence of *Campylobacter* spp. in poultry product is generally higher than for other meats (Wong, Hollis et al. 2007), and in other countries a similar high prevalence in chicken meat has been observed, e.g. 60.9% in a UK study (Little, Richardson et al. 2008) and 60% in a study from Belgium (Habib, Sampers et al. 2008). In this study 78.9% of fresh carcasses tested positive and differences in contamination probability and levels were observed for the different suppliers. High contamination of retail poultry in New Zealand was also reported by a study conducted in 2003 and 2004 by Wong et al. (Wong, Hollis et al. 2007). In addition the prevalence and mean counts of *Campylobacter* spp. positive carcasses were previously found to vary between different processing plants in New Zealand (Chrystal, Hargraves et al. 2008). However these observations were not significant, which may have been a result of the study's design and power. In Belgium, recently similar supplier-specific observations were made (Habib, Sampers et al. 2008). Our findings support the view that humans are highly exposed to *Campylobacter* spp. through retail poultry in New Zealand (Baker, Wilson et al. 2006), and revealed that contamination probability and levels differ between the individual suppliers.

We examined the similarity and relatedness of genotypes between human and poultry isolates from the different suppliers to explore the diversity of *Campylobacter* STs in the New Zealand poultry supply and to investigate further the importance of poultry products as a contributor to the burden of human disease. A total of 51 different STs were detected in human samples in our study. Of these 21 STs were also isolated from poultry sources and these STs accounted for 85.5% of human cases. This overlap between human and poultry-associated genotypes is consistent with poultry being important in disease transmission in New Zealand (Eberhart-Phillips, Walker et al. 1997; Baker, Wilson et al. 2006).

This study revealed the existence of both supplier-specific and ubiquitous strains, a total of 15 STs were found to be unique to one of the suppliers. Although with increased sample size and therefore higher detection rates this number could have been smaller, this is still an unexpected finding. Whereas the dominating human pathogen ST - 474 is commonly found in samples from Supplier A (20.6% of samples), this ST was only rarely isolated from other sources in the sentinel surveillance site, including samples from Supplier B (1.6% of samples) and bovine, ovine and environmental samples (French and Molecular Epidemiology and Veterinary Public Health Group Hopkirk Institute 2008). As another example, ST - 3609, was commonly isolated in Supplier B (12.6% of samples) but has not yet been detected in any other supplier, or in any human clinical cases. On the other hand ST - 45, ST - 48, ST - 50 and ST - 53 were ubiquitous in the poultry supply and these STs together accounted for about a quarter of human cases (26.6%). This suggests that both ubiquitous and poultry-supplier associated strains contribute to

human campylobacteriosis in New Zealand. As a consequence the relative contribution of individual suppliers to the public health burden may differ substantially. Further studies from this sentinel surveillance site have used risk attribution models to quantify the contribution of different sources to the human disease burden. These approaches estimated that 80% of human cases originate from poultry, with the contributions of the individual suppliers ranging from 7 - 58% (Mullner, Jones et al. 2009).

In an explorative analysis, using only a small subset of samples, we detected a high proportion of poultry samples that were contaminated with more than one ST. Similar findings have been made in other studies (Kramer 2000) and this further underlines the importance of a considering the consequences of selecting and testing only one presumptive isolate per sample. In this study we are assuming that on average by only typing one isolate per sample we would be detecting isolates at the relative frequently that they occur in all samples. However this does affect the probability of detecting all strains present in a source, and the existence of multiple strains in individual samples may decrease the likelihood of discovering all strains. The significance of statistical consideration in sample size calculation to detect all present strains was recently discussed by Dopfer et al. (Dopfer, Buist et al. 2008).

In this study, the PSI was calculated to assess formally the similarity between the distribution of bacterial subtypes from the different poultry suppliers and human samples. The highest PSI was observed between genotypes from Supplier A and human samples providing further evidence for the link between this supplier and human cases. This is supported by the finding that ST - 474 has a very low prevalence in samples from non-poultry sources in New Zealand (French and Molecular Epidemiology and Veterinary Public Health Group Hopkirk Institute 2008; McTavish, Pope et al. 2008). It is notable that the similarity of Supplier A genotypes to human genotypes is higher than the similarity between poultry suppliers. This observation may be explained by both the high market share of Supplier A, and hence the relatively high exposure of the human population to strains from this supplier, and the vertically-integrated structure of the industry which is likely to reduce between-supplier transmission. In the sentinel surveillance site poultry genotypes shared a higher similarity with human genotypes, as estimated by the PSI, compared to other sources, such as ruminants or wildlife (French and Molecular Epidemiology and Veterinary Public Health Group Hopkirk Institute 2008). The program STRUCTURE was used as an exploratory method to identify genotypic clusters between isolates from the different poultry suppliers. This analysis suggests that genotype distribution is structured by the individual suppliers, for example the majority of isolates from Supplier A originate from Cluster 1: the same cluster which contains the major human pathogen ST - 474. Each company has a different cluster associated with it and this provides further support for the existence of supplier-specific transmission pathways - most likely as a consequence of the vertically

integrated nature of the New Zealand poultry industry, whereby suppliers control all stages from primary production to processing with no imports and very little or no overlap between companies. The differences in the diversity of genotypes in the three clusters could be a result of different levels of biosecurity between suppliers, where presumably low biosecurity would lead to a higher diversity of genotypes present in the cluster associated with that supplier. In addition breeding stock may play a role in the importance of the different clusters. It would be interesting to see if similar findings can be made in other countries with a different industry structure and to investigate further the reasons behind this diversity.

It is striking that the majority of human cases were caused by an internationally rare ST, namely ST - 474, which is highly prevalent in the New Zealand poultry industry. This ST - 474 was the most commonly isolated genotype in the winter epidemic in 2006 in New Zealand (McTavish, Pope et al. 2008). To date this ST has only one submission to the *Campylobacter* PubMLST database (Dingle, Colles et al. 2005) from a chicken sample in the Czech Republic and was, besides in samples from New Zealand (French and Molecular Epidemiology and Veterinary Public Health Group Hopkirk Institute 2008; McTavish, Pope et al. 2008; Taboada, MacKinnon et al. 2008), only reported sporadically (Clark, Bryden et al. 2005; Best, Fox et al. 2007). Studies by Taboada and Pope showed that some *Campylobacter* strains, including ST - 474, are of higher virulence for humans than others (Pope, Wilson et al. 2007; Taboada, MacKinnon et al. 2008). Further research into pathogenicity and virulence genes of this dominant clone could reveal determinants for human infection and commensal colonisation and may explain the dominance of this ST in New Zealand.

Other STs detected in our study are also rare overseas or have never been found elsewhere (French and Molecular Epidemiology and Veterinary Public Health Group Hopkirk Institute 2008). This may in part be due to reporting bias, since it is likely that not all detected strains are described in published studies or submitted to the MLST database, in particular as MLST has just recently emerged as new approach for typing of *Campylobacter* spp. However many countries have recently submitted isolates to the shared database and conducted large scale studies using MLST typing (Colles, Johnes et al. 2003; Strachan, Gormley et al. 2009; Wilson, Gabriel et al. 2008). In addition to this many of the internationally common *Campylobacter* STs, such as ST - 45 and ST - 48, can be found in human and poultry samples in New Zealand. These STs are widespread in other countries (Mickan, Doyle et al. 2007; Colles, Jones et al. 2008; Gormley, MacRae et al. 2008) , which is reflected by a wide range of isolates included in the *Campylobacter* PubMLST database. The introduction of a variety of animals and animal products to a geographically isolated country in days of early settlement in combination with a now tight border biosecurity would provide a framework for a distinct ecosystem of *Campylobacter* spp. in New Zealand, which may have originated

from imported pathogens during early settlement. In support of this hypothesis French et al. (French, Midwinter et al. 2008) isolated *C. jejuni* genotypes from wild bird faeces in New Zealand that were associated with wild bird populations in the northern hemisphere.

A structured longitudinal sampling approach was applied in this study, which includes the simultaneous collection of data from human clinical cases and corresponding disease source to represent exposure. This sampling strategy has several advantages, such as a higher representativeness over commonly used approaches such as cross-sectional surveys, which may not be concurrent in space and time. This study has revealed the unique molecular epidemiology of *C. jejuni* in the Manawatu region of New Zealand and provided valuable information on the strong link between poultry and human campylobacteriosis. Comparative studies conducted in Christchurch and Auckland show a similar distribution of genotypes in both humans and poultry (French and Molecular Epidemiology and Veterinary Public Health Group Hopkirk Institute 2008), suggesting that the results from the Manawatu are likely to be similar to that in other regions and allow for extrapolation to the whole of New Zealand. Using novel approaches to source attribution the majority of human cases (80%) in the Manawatu were estimated to be attributable to poultry sources (Mullner, Jones et al. 2009). This study also has provided significant input in the New Zealand Food Safety Authority's (NZFSA) *Campylobacter* Management Strategy in Poultry (http://www.nzfsa.govt.nz/consumers/food-safety-topics/foodborne-illnesses/ campylobacter/strategy/index.htm) and with the strategic introduction of a range of industry and regulatory control measures, there has been a more than 50% drop in human cases in 2008 to the lowest notification rate in 16 years.

4.6 Acknowledgements

I would like to thank all members of the Hopkirk Molecular Epidemiology Team, Environmental Science and Research (ESR), MidCentral Health (in particular Tui Shadbolt), Public Health Services, NZFSA (in particular Science Group, Roger Cook, Donald Campbell and Sharon Wagener), and MedLab Central for their contributions. I want to thank in particular Sarah Moore for maintaining the database and Lynn Rogers for enabling the cooperation with MedLab Central. The statistical count analysis and STRUCTURE model results for this manuscript were derived by Simon Spencer (University of Nottingham), who also provided the main contribution to their graphical outputs.

4.7 Appendices

Appendix C provides further details into the data and methods applied in this study.

Chapter 5

Source Attribution of Food Borne Zoonosis in New Zealand: A Modified Hald Model

P. Mullner, G. Jones, A. Noble, S. Spencer, S. Hathaway, and N. French. Source Attribution of Food Borne Zoonosis in New Zealand: A Modified Hald Model. *Risk Analysis*. Accepted for publication 30 January 2009.

5.1 Abstract

A Bayesian approach was developed by Hald et al. (2004) to estimate the contribution of different food sources to the burden of human salmonellosis in Denmark. This paper describes the development of several modifications which can be used to adapt the model to different countries and pathogens. Our modified Hald model has several advantages over the original approach, which include the introduction of uncertainty in the estimates of source prevalence and an improved strategy for identifiability. We have applied our modified model to the two major food borne zoonosis in New Zealand, namely campylobacteriosis and salmonellosis. Major challenges were the data quality for salmonellosis and the inclusion of environmental sources of campylobacteriosis. We conclude that by modifying the Hald model we have improved its identifiability, made it more applicable to countries with less intensive surveillance and feasible for other pathogens in particular with respect to the inclusion of non-food sources. The wider application and better understanding of this approach is of particular importance due to the value of the model for decision making and risk management.

5.2 Introduction

Source attribution is the process of determining what proportion of a particular disease is acquired from a given source (e.g. poultry) and through a given pathway (e.g. food, water, person-to-person transmission). This 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 (Batz, Doyle et al. 2005). Most quantitative risk assessments commonly deal with one pathogen occurring in a single food commodity and are targeted to identify options for prevention, intervention and control. In contrast, source attribution models provide information about the public health impact of all important sources and pathways (Havelaar, Braeunig et al. 2007). These methods are intended to provide decision makers with a set of tools for priority setting to achieve a more targeted control of diseases (Havelaar, Braeunig et al. 2007). Overall these tools are now vital components for the prioritization of hazards and interventions in food systems (Doyle and Erickson 2006).

Classical approaches to source attribution include full risk assessments, analysis and extrapolation of surveillance or outbreak data, and analytical epidemiological studies (Adak, Cowden et al. 1995; Eberhart-Phillips, Walker et al. 1997; Gillespie, O'Brien et al. 2002; Gillespie 2003). However recently source attribution has received a considerable amount of attention, in particular since the development of a source attribution model for salmonellosis by Hald et al. (2004). The Hald model makes explicit use of *Salmonella* typing data by using differences in the relative frequency of occurrence of *Salmonella* subytpes in individual sources to quantify their contribution to the human disease burden. The Hald model has now successfully been applied to salmonellosis in several EU member states (Pires, Nichols et al. 2008; Wahlstroem, Anderson et al. 2008). We have modified the Hald model to make it more generic and then applied it to New Zealand's major food borne zoonosis, namely campylobacteriosis and salmonellosis. Depending on the pathogen a different subset of modifications was chosen. Those modifications included the inclusion of non-food sources and a new approach to estimating the prevalence of different pathogen subtypes in a source in the absence of large scale surveillance.

Campylobacteriosis is the most frequently found food borne zoonosis in almost all developed countries (Oberhelman and Taylor 2000). However the development of successful attribution models for this pathogen has been held back by the lack of an appropriate typing scheme as well as its complex epidemiology, which includes environmental reservoirs (French, Barrigas et al. 2005). With the recent application of Multilocus Sequence Typing (MLST) to *Campylobacter jejuni* a typing technique has become available, which can be used to investigate the importance of individual components of the food chain as sources of human *C. jejuni* infection (Dingle, Colles et al.

2002). This technique was developed to understand bacterial population structures and is able to show host association for this pathogen (Colles, Jones et al. 2003; French, Barrigas et al. 2005; Kwan, Barrigas et al. 2008). An additional challenge to source attribution for *Campylobacter* is modelling the role of environmental contamination. Although it is relatively straightforward to estimate the prevalence of a pathogen in a particular type of retail product, such as broiler carcasses, and combine this with the estimated consumption within a population to get an estimate of the exposure through this food source, this approach is hard to apply to the environment as a source of disease. While the environment as a disease source can be thought of as a proxy for unmeasured wildlife sources, it may also be a transmission pathway for pathogens present in livestock sources such as bovines. In order to adapt the Hald model, which originally uses the amount of food source consumed to apportion human cases, changes had to be made to key parameters of the model.

Sparse data was a major problem with our salmonellosis model. The original application of the model makes use of Denmark's extensive surveillance system and as a consequence point estimates for the prevalence of the different *Salmonella* types in the animal food sources could be used. In New Zealand data were much sparser and an approach was developed to represent uncertainty in the prevalence matrix. In addition information about prevalences from different studies had to be combined to derive the best possible estimates. In the next section we review briefly the Hald model. We then describe our modifications, which we illustrate by analysing data on salmonellosis and campylobacteriosis in New Zealand. In alignment with the Hald model and other approaches it is assumed that cases with a history of travel in the incubation period, have acquired the infection overseas. As a consequence known travelers were excluded form the attribution process. In our campylobacteriosis model travel status was known for all cases whereas in the salmonellosis model where unknown the status was estimated as proposed by Hald et al.

5.3 The Bayesian risk attribution model by Hald et al.

The Hald salmonellosis model was published in 2004 and compares the number of human cases caused by different subtypes with their prevalence in different food sources, weighted by the amount of food source consumed. To do this effectively it requires a heterogeneous distribution of bacterial subtypes among the different animal and food sources. This model is a further development of the so called Dutch model (Hald 2002), a frequentist model, which compares the number of reported human cases caused by a particular bacterial subtype with the relative occurrence of that subtype in each source. By using a Bayesian approach implemented in the WinBUGS software, the Hald model can explicitly include and quantify the uncertainty surrounding each of the parameters. A detailed analysis of the Dutch and the Hald model on Dutch data has been made by

Hald et al.(2002), and gives comparable results for attribution.

In the model o_i represents the number of human cases of type i and the expected number of cases of *Salmonella* type i from source j is denoted by λ_{ij}. Using the parameters defined in Table 5.1, assume that

$$o_i \sim Poisson(\sum j\lambda_{ij}), \tag{5.1}$$

and that

$$\lambda_i j = M_j p_{ij} q_i a_j. \tag{5.2}$$

Table 5.1: Description of parameters of Hald model.

Parameter	Description
λ_{ij}	expected number of cases / year of type i from source j
M_j	amount of food source j consumed
p_{ij}	prevalence of type i in source j
q_i	bacteria dependent factor for type i
a_j	source dependent factor for source j
o_i	number of human cases of type i

Here the bacteria dependent factor q_i combines survivability, virulence and pathogenicity of the pathogen to estimate the ability of that type to cause disease (measured in cases per dose of bacteria in the population). On the other hand the source dependent factor a_j summarizes the ability to act as vehicle for foodborne infections including factors such as the environment provided for the bacteria through storage and preparation (measured in doses of bacteria per kg of infected material consumed). Usually q_i is defined to equal 1 for some reference type, thereby defining implicitly the size of a 'dose'.

One difficulty of this modelling approach is the disparity between the number of data points and the number of parameters to be estimated. Given I types and J sources, there are $I+J$ parameters (q_i and a_js) but only I independent data points (the observed case totals o_i), so the model is not identifiable: estimates cannot be obtained from the data alone and will be sensitive to the priors used. This non-identifiability raises the question of whether the output values are more a product of the assumptions made than the data from which they were derived. The pooling of bacterial subtypes into groups with similar characteristics is one way of addressing this problem, as fewer parameters

then have to be estimated. This was done in the Hald salmonellosis model by assuming that q_i is of equal value for *S.* Enteritidis and *S.* Typhimurium subtypes. In addition it was assumed that a_j is equal for some foods, e.g. Danish and imported pork. The priors chosen for q_i and a_j:

- $q_i \sim Uniform(0, 10)$
- $a_j \sim Uniform(0, 0.01)$

were assumed to be non-informative.

Data quality and representativeness of the data have been identified as key determinants of the successful adaptation of this model (Pires, Nichols et al. 2008). However this has not yet been formally explored and proof of this could be provided by simulating degrees of data quality and representativeness and describe how the model performs in response. In the original model approximately 25% of human cases were assigned to an unknown source. This is a result of the grouping scheme where some Salmonellae are pooled in a category of 'others including not typed.' Hald et al. concluded that this fraction could be reduced by introducing more subtypes individually into the model (Hald, Vose et al. 2004).

5.4 A modified Hald model

A set of modifications was developed to make the Hald model more generic. This includes a methodology for incorporating uncertainty in the prevalence parameters. We have also taken a different approach to achieving identifiability and to the setting of priors. In addition approaches were developed to include potentially pathogenic subtypes and to avoid food consumption weights when considering environmental sources. A subset or all of these modifications can be used to apply the model to a particular pathogen and dataset. The modifications chosen will depend on the data quality available and the pathogen's epidemiology.

5.4.1 Modelling prevalence uncertainty

In our modified model uncertainty is introduced in the estimates for the prevalence p_{ij} of type i in source j. The original model uses data from Denmark's intensive surveillance system as a justification for keeping source prevalences fixed. In order to make the model more generic and therefore applicable to other countries and pathogens it was necessary to introduce uncertainty into these estimates. The priors for the p_{ij} were chosen to be independent $Beta(\alpha_{ij}, \beta_{ij})$ distributions. 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 Bayesian analysis of the prevalence data

for each source. Depending on the data we have developed a standard and a novel approach to estimating prevalence. In the standard approach to the analysis of the prevalence, we assumed that $p_{ij} = \pi_j r_{ij}$, where π_j is the prevalence over all types in source j and r_{ij} is the relative occurrence of type i in the successfully typed isolates from source j. The priors used were $r_{ij} \sim Dirichlet(1, 1, ..., 1)$ and $\pi_j \sim Beta(1, 1)$.

Good data on the prevalence of different subtypes in a source for the standard approach may not be available and we have developed an alternative approach to use data from different studies, including data provided by routine surveillance as well as small scale surveys. In this novel approach we consider three different types of study, investigating relative prevalence r_{ij}, prevalence over all types π_j, or absolute prevalence $p_{ij} = \pi_j r_{ij}$. For ease of exposition we assume only three pathogen types. The three data types are:

1. Typed positives only: observe X_1, X_2 and X_3 samples positive for types 1 to 3 out of $X_1 + X_2 + X_3 = N_x$. The contribution to the likelihood function will be

$$L_1(r_1, r_2) \propto r_1^{X_1} r_2^{X_2} (1 - r_1 - r_2)^{X_3}. \tag{5.3}$$

2. Typed positives and negatives: observe Y_1, Y_2 and Y_3 samples positive for types 1 to 3 out of a total of N samples.

$$L_2(\pi, r_1, r_2) \propto (\pi r_1)^{Y_1} (\pi r_2)^{Y_2} (\pi - \pi r_1 - \pi r_2)^{Y_3} (1 - \pi)^{N_Y - Y_1 - Y_2 - Y_3} \tag{5.4}$$

3. Untyped positives and negatives (overall prevalence): observe Z positives out of N_z samples.

$$L_3(\pi) \propto \pi^Z (1 - \pi)^{N_Z - Z} \tag{5.5}$$

Putting these all together gives the likelihood function from all available data as

$$L_1(\pi, r_1, r_2) \propto r_1^{X_1+Y_1} r_2^{X_2+Y_2} (1 - r_1 - r_2)^{X_3+Y_3} \pi^{Z+Y_1+Y_2+Y_3} (1 - \pi)^{N_Z+N_Y-Z-Y_1-Y_2-Y_3} \tag{5.6}$$

If we start with independent priors, $dirichlet(1, 1, 1)$ for $(r_1, r_2, 1 - r_1 - r_2)$ and $beta(1, 1)$ for π we get independent posteriors (by factorization theorem):

$$(r_1, r_2, 1 - r_1 - r_2) \sim dirchlet(X_1 + Y_1 + 1, X_2 + Y_2 + 1, X_3 + Y_3 + 1) \tag{5.7}$$

$$\pi \sim beta(Z + Y_1 + Y_2 + Y_3 + 1, N_Z + N_Y - Z - Y_1 - Y_2 - Y_3 + 1) \tag{5.8}$$

Where data are available in different time periods, for example in consecutive years, it may be preferable to smooth the observed numbers of cases over time using a weighted

moving average. These posteriors incorporating the available prevalence information could now be used as priors in the source attribution program. We have found the full scheme to be too complex for our WinBUGS updater, so we have implemented it as a two-stage process as follows.

First equations 5.7 and 5.8 were implemented in WinBUGS and posterior means and standard deviations (SD) for the prevalences p_{ij} were obtained. The parameters of a beta distribution (α_{ij} and β_{ij}) were chosen to match the mean and SD of each the prevalences. In order to avoid convergence problems for very small values of α_{ij}, we enforced a minimum $a = 1$ and set the corresponding β_{ij} to match the mean only. The values of α_{ij} and β_{ij} for the different prevalences p_{ij} were then used to specify beta priors in the risk attribution model for each source.

5.4.2 Splitting data into different time periods

An alternative, or complementary approach for achieving identifiability, is to divide the observation period into a number of intervals and to estimate prevalences separately in each time interval t. Equations 5.1 and 5.2 now become

$$o_{it} \sim Poisson(\sum j\lambda_{ijt}) \tag{5.9}$$

,

$$\lambda_{ijt} = M_{jt}p_{ijt}q_i a_j \tag{5.10}$$

where all data values and parameters are assumed to change between time periods, except for q_i and a_j, which are assumed to be constant over time. This increases the number of independent data values while keeping the number of parameters to be estimated constant, and will lead to identifiability provided the prevalences p_{ijt} do vary with time.

5.4.3 Hierarchical model for bacterial parameters

To achieve identifiability in the model, the number of q_i parameters must be reduced to be less than the number of observations o_i. Hald et al. do this by assuming that some of the q_i are equal. We prefer to model the q_i hierarchically as random observations from a hypothetical distribution of bacterial characteristics. We used a lognormal distribution $log(q_i)$ $N(0, \tau)$, where τ is a parameter controlling the variation in characteristics between types. A prior distribution is needed for τ; we used a fairly diffuse gamma prior $\tau \sim gamma(0.01, 0.01)$. We conducted a sensitivity analysis and found that variations in the choice of prior did not significantly influence the results. The mean of 0 on the log scale parallels Hald et al.'s approach of fixing one q_i to equal 1. This random effects approach allows for the possibility that some parameters may

be similar, as in the Hald model, but without forcing them to be equal. If genuine prior information is available on relative characteristics, this can be incorporated. The model is now identifiable because the I virulence parameters are replaced by a single τ parameter describing the virulence distribution. This is particularly useful if there are a large number of subtypes.

5.4.4 Exponential prior for source specific parameters

We also changed the priors for the a_j parameter. Independent prior distributions for the source dependent factors were chosen as a_j $Exponential(\lambda)$. This prior is fairly uninformative, but prevents the a_j from becoming too large, whilst not specifying a strict upper boundary as in the Hald model. An alternative would be to induce priors on the a_j by speculating on the likely observed number of cases for an 'average' pathogen type given a fixed prevalence and amount consumed of source j. The priors for a_j will depend on the inclusion of food consumption weights and their unit of measurement. In our analysis we investigated the sensitivity of the results to the value of λ chosen for the prior.

5.4.5 Avoid food consumption weights

For some pathogens such as *C. jejuni* the environment is a well known source of infection (French, Barrigas et al. 2005) and adjustment has to be made to include it into the model, as the original was focused on animal-food sources. The Hald model uses a food consumption weighting factor M_j in its specification. To include the environment in the model we removed M_j because weighting exposure to the environment in a comparable way to consumption of a food source (as measured in e.g. kg of the product produced for consumption) could not be achieved in a sensible way and the M_j are not essential to the model. With the M_j in the model the related food source dependent factor a_j can be interpreted as the relative ability of a certain food source to cause disease. By removing M_j the information about the amount of food consumed in a population is not explicitly included in the model but will be absorbed in the values for a_j. In general the estimated values for a_j (and q_i) are simply multiplication factors to arrive at the most probable solution given the observed data and their size gives an idea about the different ability of the food sources to transmit disease (or the bacteria subtypes to cause disease). In the modified model a high a_j for example may reflect a high exposure (e.g. a large market share), but not necessarily a high ability of the individual food source to cause disease, as a consequence the a_j are less meaningful and less comparable. This modification is not unique to our models and has also been applied to the adaptation of the Hald model to salmonellosis in Sweden (Wahlstroem, Anderson et al. 2008). We extend this work by outlining the theory behind this adaptation and describe the consequences for the interpretation of the results.

5.4.6 Including potentially pathogenic subtypes

The Danish data used in the Hald model only consists of one particular type of data: *Salmonella* types, which have been found in both human cases and at least one food source. The Danish data has a pooled group 'Other *Salmonella* including not typable isolates' as well as the major individual *Salmonella* types. As a consequence the origin of approximately 25% of the human cases was classified as 'unknown'(Hald, Vose et al. 2004). However depending on the resolution of the typing method a dataset can consist of three data types: ST's occurring in both human cases and at least one of the sources (like in the Danish data set), those only occurring in humans and those only occurring in at least one of the sources but not in humans. We chose to also include the third type, the potentially pathogenic ST's into the model. These can be found in the sources but have not (yet) been detected in humans. We assume that these subtypes are potentially pathogenic but rarely occur in humans. We did not include human types that haven't been found in any of the sources. In the absence of information relating them to any of the major sources we are assuming those types may have come from unknown, possibly exotic, sources. In consequence the model only attributes cases of subtypes that have been detected in a source. To attribute types undetected in a source inferences from genotypic relatedness may be used as proposed by Wilson et al.(2008). However this only applicable where genotyping data is being used, and therefore has not been applied as yet to *Salmonella*, which is routinely investigated by phage typing.

5.5 Application of the modified Hald model

5.5.1 Campylobacteriosis

Details of campylobacteriosis model

Data for campylobacteriosis source attribution was generated by a sentinel surveillance study for *C. jejuni* in the Manawatu region of New Zealand conducted between 2005 and 2008 (French and Molecular Epidemiology and Veterinary Public Health Group Hopkirk Institute 2008) and the allocation is based on 481 MLST typed human cases. The Manawatu study consists of structured parallel studies of isolates from domestic human cases of *C. jejuni* and environmental and food sources in a defined geographical area over a three year period. Retail chicken samples were used to represent poultry sources since in New Zealand chicken represent 95% of poultry meat consumed. Details of the study are discussed in French et al. (2008) and further work is underway to differentiate between individual poultry sources. All isolates in the dataset were completely typed and each *C. jejuni* subtype was modeled individually.

The following modifications were applied to this pathogen:

- Modelling prevalence uncertainty using the standard approach
- Using a hierarchical model for bacterial parameters
- Using an exponential prior for source specific parameters
- Avoiding food consumption weights
- Including potentially pathogenic subtypes

Samples from the posterior distribution for this model were 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 with a thinning of 10. The prior distribution for τ was assumed to be $\tau \sim gamma(0.01, 0.01)$. The sensitivity of the model for two different priors for the food source factor a_j, the standard prior $a_j \sim Exponential(0.0020$ and a less restricting prior $a_j \sim Exponential(0.01)$, was assessed. The model's sensitivity against different groupings of the sources was tested. The source 'poultry' was split into the major three suppliers and the source 'ovine' samples was split into 'retail' and 'on-farm samples.' In a second step we assessed the models sensitivity to the inclusion of potentially pathogenic *C. jejuni* subtypes.

Results of campylobacteriosis model

The risk model for *C. jejuni* attributes 474 human cases, a close match to the observed number of 481 cases. Overall observed (oi) and expected cases (λ_{ij}) for the individual *Campylobacter* subtypes were in good agreement (Figure 5.1). Out of the 474 cases 379 were attributed to poultry (80%), 48 cases to bovine (10%), 44 cases to ovine (9%) and 4 cases (1%) to the environment (Figure 5.2, left graph). Including uncertainty in the prevalence estimates widens the credible intervals of the attribution estimates and mainly affects the estimates for bovine and poultry. The bovine estimate is affected the most; its point estimate changes from 48 to 68 cases and the upper credible interval limit increases from 160 to 180 (Figure 5.2, left graph).The effect of a change of prior from $a_j \sim Exponential(0.002)$ to $a_j \sim Exponential(0.01)$ was minor, predictions were within 15% of the standard approach (Figure 5.2, right graph).

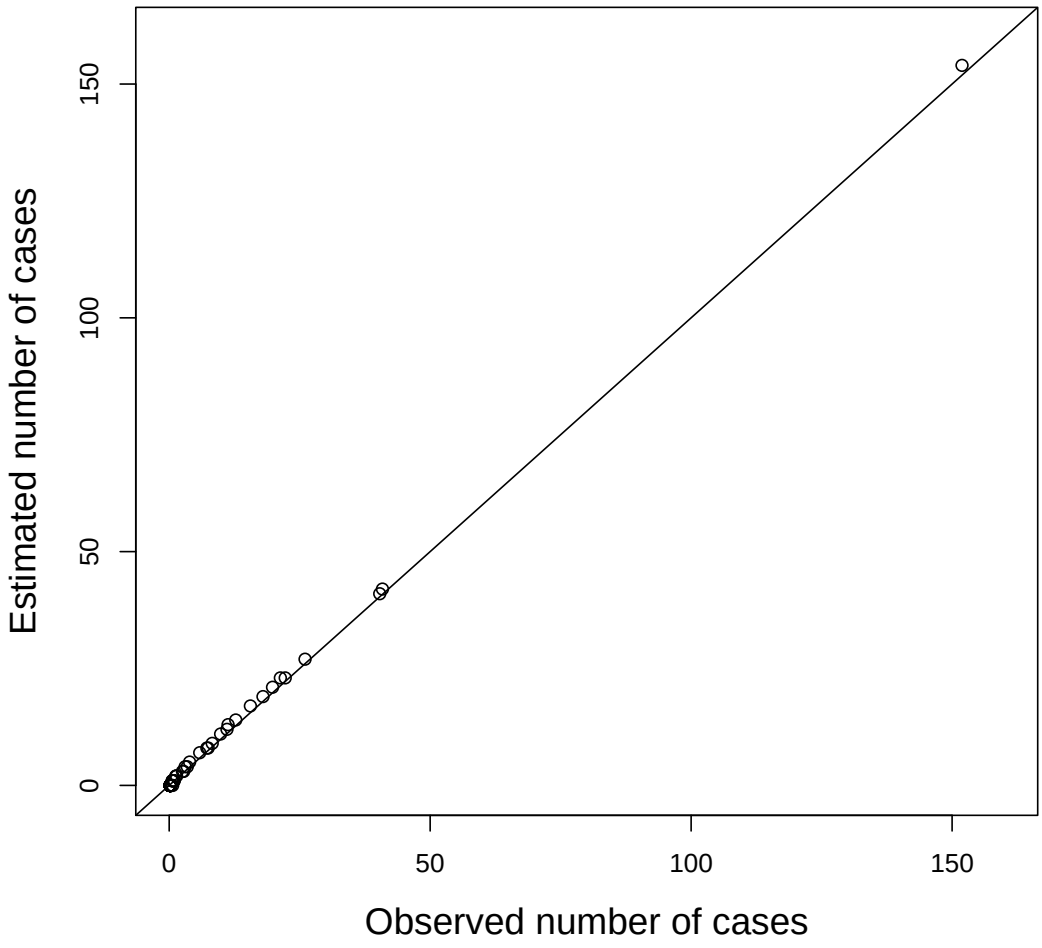

Figure 5.1: Plot of the observed (o_i) and expected (λ_{ij}) cases for the individual *Campylobacter* subtypes

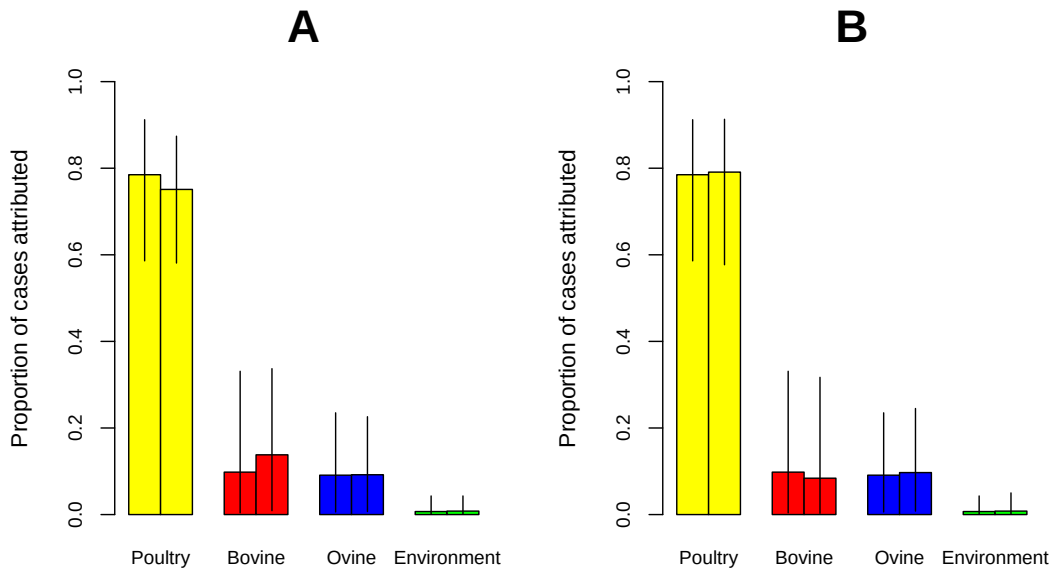

Figure 5.2: Attribution results based on 481 campylobacteriosis cases on the Manawatu region of New Zealand to several sources. Median proportion of cases attributed to each source with 95% Bayesian credible intervals.
The left graph (A) shows results from the modified Hald model with (on the left) and without (on the right)uncertainty in the prevalence matrix
The right graph (B) shows the sensitivity of the modified Hald for different priors for a_j. The standard model using $a - j \sim Exponential(0.002)$ is represented in the left column and $a_J \sim Exponential(0.01)$ in the right column.

To test the sensitivity of the model to changes in the underlying assumptions several alternative models were developed (Table 5.2). The impact of the choice of model on the source attribution estimates is documented in Figure 5.3. The model showed a low sensitivity to changes in the source grouping. It was able to distinguish between ovine on-farm and retail exposure without affecting the overall estimate and similarly could distinguish between the three major New Zealand poultry suppliers. It was however more sensitive to the inclusion of potentially pathogenic *C. jejuni* subtypes, which affected all estimates, except the one for ovine sources.

Estimates for the food source dependent factor a_j, the bacteria dependent factor q_i and the shape parameter tau are presented in Table 5.3. Poultry was estimated to have the highest value for a: 127.7 (95% CI 56.7–247.6) followed by bovine, ovine and the environments. The bacterial types with the highest values for q were: ST-474 with a median estimate of 19.4 (95% CI 9.6–42.4), ST-38 with a median estimate of 16.5 (95% CI 4.8–66.3) and ST-354 with a median estimate of 14.0 (95% CI 5.1–45.6).

Table 5.2: Description of standard and alternative models to test robustness of campylobacteriosis model.

Model number	Model name	Source grouping	Inclusion of potentially pathogenic subtypes
I	Standard	poultry	yes
		bovine	
		ovine	
		environment	
II	No extra types	poultry	no
		bovine	
		ovine	
		environment	
III	Split ovine	poultry	yes
		bovine	
		retail ovine	
		on-farm ovine	
		environment	
IV	Split chicken	poultry supplier 1	yes
		poultry supplier 2	
		poultry supplier 3	
		bovine	
		ovine	
		environment	

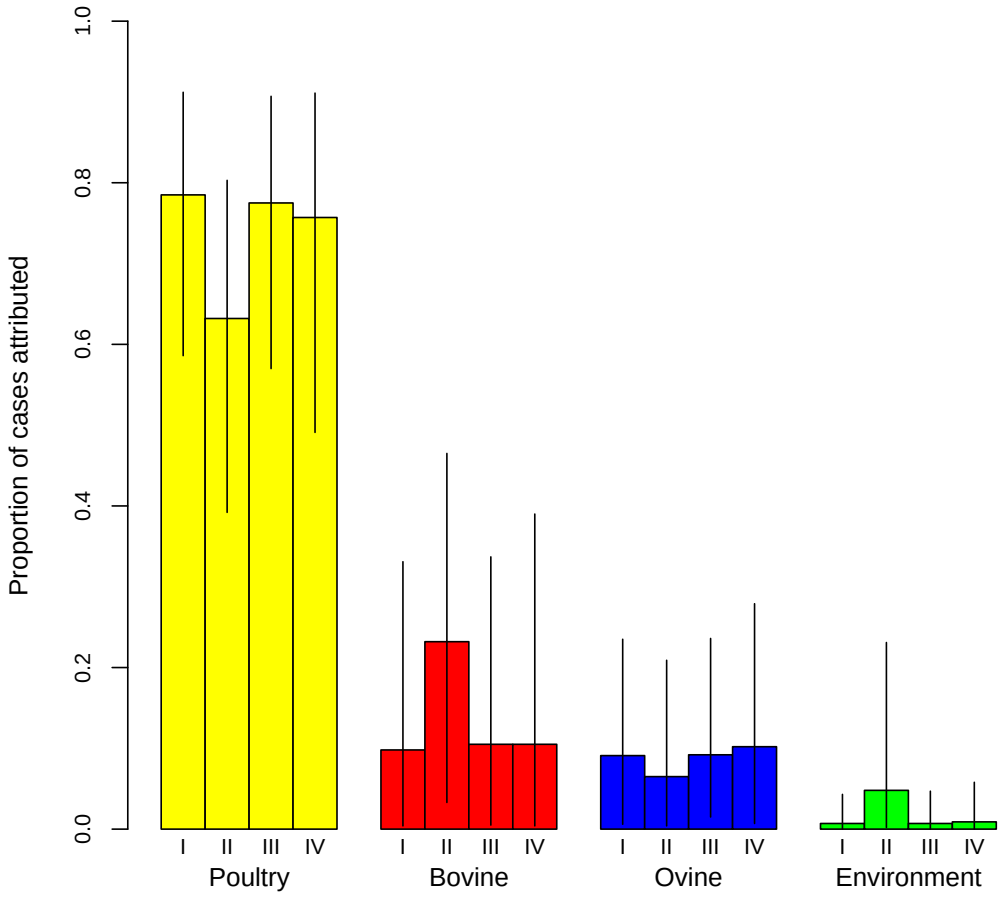

Figure 5.3: Sensitivity analysis of source attribution results to different grouping of sources and inclusion of minor *Campylobacter* types. The different models are described in detail in Table 5.2. Models I to IV are shown from the left to the right and summary estimates are shown where sources were split up. Median number of cases attributed to each source with 95% Bayesian credible intervals.

Table 5.3: Estimates for a_j, q_i and tau in standard campylobacteriosis model.

Node	Description	Median	Lower 95% CI	Upper 95% CI
a[1]	Poultry	127.70	56.67	247.60
a[2]	Bovine	60.44	2.26	252.60
a[3]	Ovine	41.20	2.41	137.30
a[4]	Environment	6.51	0.22	39.11
q[1]	ST–21	5.43	1.66	18.76
q[2]	ST–25	0.75	0.11	3.67
q[3]	ST–38	16.47	4.82	66.28
q[4]	ST–42	5.26	2.26	12.98
q[5]	ST–45	2.27	1.08	5.07
q[6]	ST–48	4.87	2.21	11.28
q[7]	ST–50	1.86	0.91	3.94
q[8]	ST–52	8.15	2.91	24.86
q[9]	ST–53	4.11	1.89	9.74
q[10]	ST–61	6.44	2.36	20.83
q[11]	ST–137	1.53	0.18	9.66
q[12]	ST–190	5.74	2.56	13.05
q[13]	ST–257	3.08	1.14	8.25
q[14]	ST–354	13.97	5.13	45.60
q[15]	ST–422	1.46	0.34	6.16
q[16]	ST–436	3.81	0.98	14.59
q[17]	ST–451	1.53	0.56	4.24
q[18]	ST–474	19.42	9.64	42.38
q[19]	ST–520	2.67	0.94	7.52
q[20]	ST–583	2.52	0.93	7.17
q[21]	ST–677	4.89	1.28	19.57
q[22]	ST–1517	1.56	0.44	5.28
q[23]	ST–1581	0.68	0.10	3.29
q[24]	ST–2026	5.01	1.75	17.21

Continued on next page

Table 5.3 – *continued from previous page*

Node	Description	Median	Lower 95% CI	Upper 95% CI
q[25]	ST–2345	0.91	0.22	3.13
q[26]	ST–2350	2.55	0.43	14.00
q[27]	ST–2391	1.18	0.15	7.15
q[28]	ST–3072	1.32	0.16	8.73
q[29]	ST–3538	1.52	0.18	9.84
q[30]	ST–3676	2.60	0.44	14.27
q[31]	ST–3711	1.93	0.33	8.89
q[32]	ST–3717	0.84	0.12	4.52
q[33]	ST–177	0.53	0.04	4.61
q[34]	ST–227	0.49	0.04	3.83
q[35]	ST–393	0.54	0.04	4.63
q[36]	ST–526	0.55	0.04	4.76
q[37]	ST–618	0.53	0.04	4.42
q[38]	ST–694	0.55	0.04	4.92
q[39]	ST–1030	0.54	0.04	5.02
q[40]	ST–1115	0.53	0.04	4.35
q[41]	ST–1191	0.53	0.04	4.41
q[42]	ST–1223	0.58	0.04	4.76
q[43]	ST–1225	0.52	0.04	4.42
q[44]	ST–1243	0.56	0.04	5.04
q[45]	ST–1818	0.49	0.04	3.61
q[46]	ST–1911	0.49	0.04	3.67
q[47]	ST–2347	0.58	0.04	5.17
q[48]	ST–2354	0.55	0.04	4.78
q[49]	ST–2381	0.47	0.03	4.08
q[50]	ST–2392	0.55	0.04	4.11
q[51]	ST–2397	0.38	0.03	2.67
q[52]	ST–2535	0.48	0.03	3.82
q[53]	ST–2584	0.55	0.04	5.11
q[54]	ST–2619	0.55	0.03	4.77

Continued on next page

Table 5.3 – *continued from previous page*

Node	Description	Median	Lower 95% CI	Upper 95% CI
q[55]	ST–3230	0.49	0.03	3.97
q[56]	ST–3232	0.48	0.04	4.41
q[57]	ST–3301	0.57	0.04	4.90
q[58]	ST–3609	0.18	0.02	0.94
q[59]	ST–3610	0.50	0.04	3.90
q[60]	ST–3640	0.58	0.04	5.05
q[61]	ST–3655	0.57	0.04	4.71
q[62]	ST–3656	0.60	0.04	5.26
q[63]	ST–3657	0.56	0.04	4.86
q[64]	ST–3658	0.57	0.04	5.20
q[65]	ST–3659	0.54	0.04	4.45
q[66]	ST–3660	0.58	0.04	4.84
q[67]	ST–3661	0.55	0.04	4.72
q[68]	ST–3662	0.56	0.04	4.72
q[69]	ST–3663	0.56	0.04	4.65
q[70]	ST–3664	0.55	0.04	5.15
q[71]	ST–3672	0.56	0.04	4.66
q[72]	ST–3673	0.59	0.04	4.74
q[73]	ST–3674	0.59	0.04	4.63
q[74]	ST–3675	0.57	0.04	4.48
q[75]	ST–3714	0.55	0.04	4.97
q[76]	ST–3716	0.51	0.03	4.31
q[77]	ST–3719	0.47	0.04	3.50
q[78]	ST–3721	0.47	0.04	3.96
q[79]	ST–3722	0.56	0.04	4.19
q[80]	ST–3723	0.55	0.04	4.43
q[81]	ST–3724	0.52	0.04	4.46
q[82]	ST–3725	0.50	0.04	3.71
q[83]	ST–3726	0.42	0.04	3.06
q[84]	ST–UA[a]	0.53	0.04	4.52

Continued on next page

Table 5.3 – *continued from previous page*

Node	Description	Median	Lower 95% CI	Upper 95% CI
q[85]	ST–UA	0.53	0.04	4.42
q[86]	ST–UA	0.52	0.04	4.35
q[87]	ST–UA	0.57	0.04	5.35
q[88]	ST–UA	0.51	0.04	4.11
q[89]	ST–UA	0.49	0.04	4.06
q[90]	ST–UA	0.51	0.04	3.78
q[91]	ST–UA	0.55	0.04	5.12
tau		0.49	0.26	0.95

[a] These STs were novel and not yet unassigned at this point of the analysis

5.5.2 Salmonellosis

Details of salmonellosis model

d In the New Zealand dataset the *Salmonella* subtype was known for all typed human cases, however the resolution of subtypes in the food source was lower. As a consequence human subtypes were reported in the same categories as the types from the food sources. This was based on the scheme commonly used in New Zealand for surveillance purposes, which includes a total of 16 types, one of which is a pooled category of less frequently occurring subtypes. This group, called 'other *Salmonella*', was very diverse, and initially included a total of 95 different subtypes, many of which were associated with foreign travel and have not been detected in any source in New Zealand. Estimates of food source prevalence were calculated for 13 *Salmonella* types and together with the occurrence of these 13 types in human cases used to attribute cases. Three of the subtypes have not been detected in any food sources and were therefore excluded from the model, based on the conclusions drawn in section 5.6. These cases account for 2-4 % of all domestic sporadic cases annually.

Human case data were derived from New Zealand's National Surveillance system EpiSurv (Institute of Environmental Science and Research Limited (ESR)). For the food sources in the absence of large scale high quality surveillance data information from different studies were used to estimate the prevalences of the different *Salmonella* types in the food sources. These included data provided by routine surveillance as well as small scale surveys. The amount of data amount was similar for all sources except pork for which data were very sparse and in parts unrepresentative. Pork is the only meat product in New Zealand that is not routinely tested and very few isolates are typed from pork annually. The consequences of this are considered in the discussion of our results.

As an example for the novel prevalence estimation approach we are presenting here results for *Salmonella* in beef and veal in 2003. For this source three different data sets were available.

- Data summarizing isolates submitted to the National Reference Laboratory for *Salmonella* (X),
- Data from a major retail survey conducted in 2003 and 2004 (Y1),
- Data from the National Microbiological Database (NMD), which standardizes New Zealand official export assurances to overseas markets (Y2).

These three data sets were combined into a data matrix (Table 5.4) and consecutively analyzed as described in 5.4.1.

The following modifications to the Hald model were applied for this pathogen:

- Modelling prevalence uncertainty using the novel approach
- Splitting data into different time periods
- Using an exponential prior for source specific parameters
- Using a hierarchical model for bacterial parameters

We modeled a combined estimate for the years 2002 to 2004 for salmonellosis, using individual year data on human cases and source prevalence. A pooled (three-year) value for each q_i and a_j was estimated, while source attribution estimates (λ_j) were calculated individually for each year. As in the Hald model, equality in the type-specific bacteria dependent parameters for *S.* Typhimurium subtypes was assumed.

The model was fitted using MCMC techniques with the software WinBUGS 1.4.3. The code was developed from the original code by Hald et al. Five independent Markov chains, with widely dispersed starting values were run for 80 000 iterations after a burn-in period of 40 000 iterations, with a thinning of 50. Convergence was monitored using the method developed by Gelman and Rubins (Toft, Innocent et al. 2007). 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(Toft, Innocent et al. 2007). The prior distribution for τ was assumed to be $\tau \sim gamma(0.01, 0.01)$. As for the campylobacteriosis model the sensitivity of the model for two different priors for a_j namely $a_j \sim Exponential(0.002)$ and $a_j \sim Exponential(0.01)$ was assessed.

Results of salmonellosis model

Table 5.5 shows median prevalence estimates and corresponding Bayesian credible intervals for *Salmonella* (p_j) and individual *Salmonella* subtypes (p_{ij}) in beef and veal in 2003 as estimated by our novel approach. Based on a total of 963 observed cases the risk model apportions an estimated 981 human cases in 2003. As in our campylobacteriosis model, observed (o_i) and expected cases (λ_{ij}) for the individual subtypes were very close in value (data not shown). The majority of cases was attributed to pork (60%, 591 cases) followed by poultry (21.2%, 209 cases) and beef and veal (11.5%, 113 cases). Eggs and lamb and mutton are estimated to be minor sources of infection with 3.2% and 1.4% of cases apportioned (Fig. 4). The sensitivity of the model against changes in the prior for a_j was tested. Changing the prior caused a minor change in the estimates for the two minor sources eggs, poultry, and lamb and mutton (Figure 5.4).

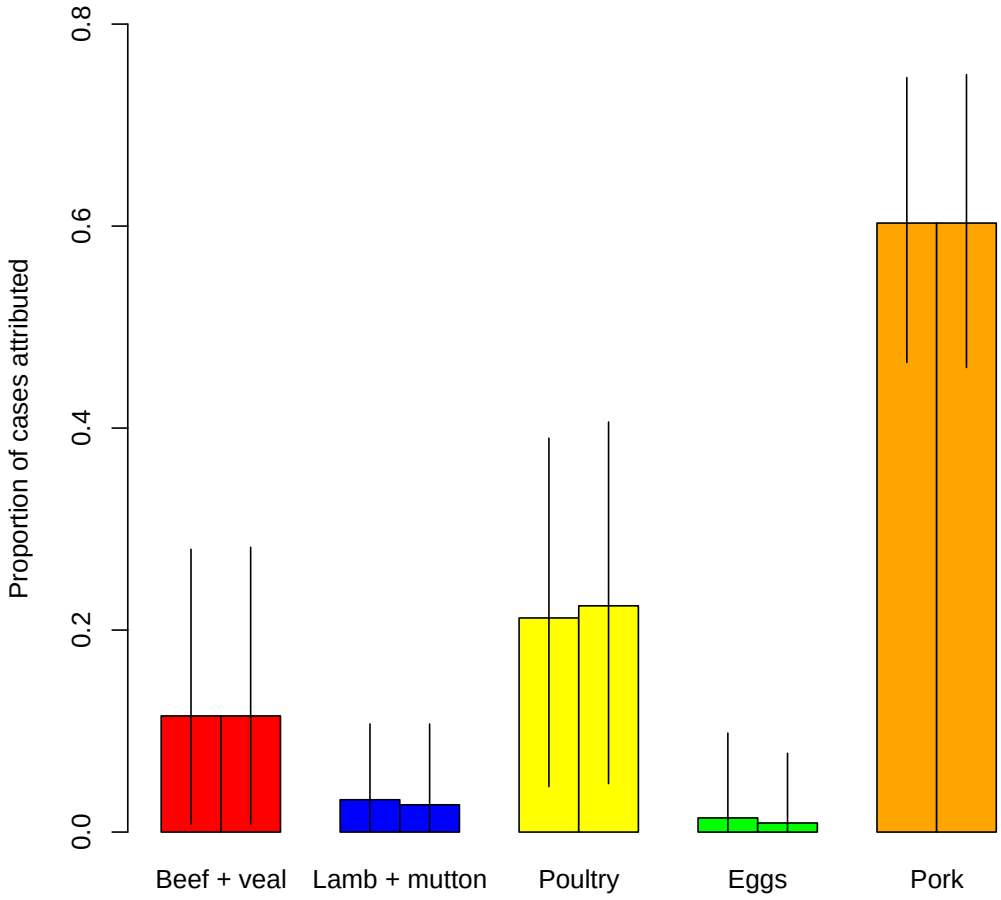

Figure 5.4: Attribution of 891 human salmonellosis cases in New Zealand in 2003. Median proportion of cases attributed to each source with 95% Bayesian credible intervals. The graph shows the sensitivity of the modified Hald model for different priors for a_j. The standard model using $a_j \sim Exponential(0.002)$ is represented in the left column and $a_j \sim Exponential(0.01)$ in the right column. These estimates are based on sparse data and need to be interpreted with caution

Table 5.4: Data matrix for *Salmonella* subtype prevalence estimation in beef and veal in 2003.

	Salmonella subtype i														
	1	2	3	4	5	6	7	8	9	10	11	12	13	Total isolates	Total samples
X (2002)	53	3	0	5	24	0	0	33	13	27	15	11	61	245	-
X (2003)	30	2	3	1	10	5	0	26	17	0	25	4	78	201	-
X (2004)	0	0	0	1	0	0	0	1	0	0	0	0	0	2	-
X^a	**33.1**	**2.1**	**2.3**	**1.8**	**12.3**	**3.8**	**0**	**26.2**	**15.4**	**5.4**	**21.8**	**5.2**	**70.7**	**200.1** (N_x)	**-**
Y_1	0	0	0	1	0	0	0	1	0	0	0	0	0	2	294
Y_2 (2002)	6	0	0	0	1	0	0	0	0	1	1	1	2	12	2138
Y_2 (2003)	4	0	0	0	4	0	0	1	3	0	0	2	2	16	1768
Y_2 (2004)	6	0	0	0	2	0	0	0	0	0	1	0	2	11	1549
Y_2^b	5	0	0	0	2.8	0	0	0.5	1.5	0.3	0.5	1.3	2	13.8	1805.8
Y^c	**5**	**0**	**0**	**1**	**2.8**	**0**	**0**	**1.5**	**1.5**	**0.3**	**0.5**	**1.3**	**2**	15.8	**2099.8** (N_y)

Included data X, Y_1 and Y_2 and methods are discussed in text
[a]X is calculated by combining data from 3 years using moving weighting averages. These are typed isolates only
[b]Y_2 is calculated by combining data from 3 years using moving weighting averages. These are typed isolates with prevalence data
[c]Y is calculated by combining Y_1 and Y_2

Table 5.5: Prevalence estimates in percent for *Salmonella* in beef and veal in 2003.

	Median	2.5% CI[a]	97.5% CI[a]
p_i for *Salmonella* subtype i			
S. Brandenburg	0.133	0.073	0.227
S. Enteritidis PT 9a	0.009	0.002	0.028
S. Heidelberg	0.01	0.002	0.029
S. Infantis	0.012	0.003	0.033
S. Other/unknown	0.054	0.026	0.103
S. Saintpaul	0.015	0.005	0.038
S. Thompson	0.002	0	0.014
S. Typhimurium DT 1	0.097	0.052	0.172
S. Typhimurium DT 101	0.06	0.03	0.112
S. Typhimurium DT 135	0.022	0.008	0.049
S. Typhimurium DT 156	0.079	0.041	0.142
S. Typhimurium DT 160	0.024	0.009	0.054
S. Typhimurium other/unknown	0.253	0.145	0.41
Overall prevalence p	0.7813	0.4591	1.221

[a] Bayesian credible interval

5.6 Discussion

5.6.1 Improving identifiability

The full model as specified by Hald et al. is over-parameterized. This was solved in the original model by assuming equality in some of the type-specific bacteria dependent and food source specific parameters. Our solution involves using a hierarchical model for the bacterial parameters, and combining data from different time periods to achieve identifiability. It is biologically plausible to assume that the q_i for all pathogen subtypes come from a common distribution. A major advantage of this approach is that no assumptions about equal value for any of the q_is have to be made. In our salmonellosis approach it was possible to model the q_is hierarchically while also assuming equality for *S.* Typhimurium and Enteritidis subtypes, as in the original approach. This could not be done for campylobacteriosis, as to date there is no evidence that would justify a grouping of MLST types. It was assumed in the Hald salmonellosis model that the source dependent factor a_j is equal for some foods, e.g. for Danish and imported pork. We have not been able to identify sources for which this assumption could be made in any of our models. To further address identifiability the modified model, as applied to salmonellosis in New Zealand, used human case and animal food source data from individual years while estimating a pooled q_i and a_j over a three year period. This change improves the ratio of data points and parameters and considerably improved the performance of our model. The assumption behind this is that those factors are constant in time. This would, as an example, include the ability of beef to cause disease, which results from a variety of factors such as the survivability of this pathogen in this food source. The assumption that $a - j$ and q_i are constant in time may not always be accurate, in particular for the q_i, which has to be carefully considered depending on the context. Prevalence of endemic *Salmonella* types is often observed to fluctuate in time and this pathogen is also well known for emerging types causing outbreaks within populations (Ben Aissa, Al-Gallas et al. 2007). Changes in consumer behaviour such as an increased consumption of raw eggs would have an impact on the true value of a_j. By careful consideration the above modelling of pooled estimates can be a way of improving model performance.

5.6.2 Choice of prior for q_i and a_j

The original Hald model uses universal uniform priors for q_i and a_j, which contain very little information about the parameters. In our modified approach q_i is modelled as a random effect and its variation is controlled by a hyperparameter τ. By taking this approach we are assuming a common distribution for all q_i, which adds to our knowledge and improves the model. A valuable extension to this, particular suitable for salmonellosis, would be to take a nested hierarchical approach, allowing e.g. for

random effects between *S.* Typhimurium and *S.* Enteritidis subtypes. The choice of prior for τ may be influential so a sensitivity analysis should be carried out to investigate the effect of different priors. We used an exponential distribution to model a_j. Our sensitivity analysis for both our models showed a low sensitivity of the estimates to the exact exponential distribution chosen. A major advantage of taking a Bayesian approach is that by running the model for different periods, the priors for q_i and a_j, could be updated and improved by including estimates of a previous model as prior for the updated model. At the moment, the model does not address the possibility of interaction between the food a_j and the bacteria related factor q_i. This would allow for the biologically plausible possibility that certain subtypes are more or less likely to survive and cause disease, dependent on the food source they appear in. There is evidence in the literature for such an interaction (Vora, Senecal et al. 2003; Birk, Rosenquist et al. 2006).

5.6.3 Introduction of uncertainty in prevalence matrix

In general, for a full Bayesian approach, uncertainty in the prevalence p_{ij} estimates should be included. In particular, in the absence of intensive surveillance data as used in the original Hald model, uncertainty around the estimates cannot be ignored as our results would otherwise overestimate the level of precision. Technically the inclusion of uncertainty around the prevalence estimates in the model removes zeros from the prevalence matrix, possibly increasing the number of parameters and making the model more difficult to fit. However if assuming zero prevalence is incorrect then the model will be mis-specified. We have evaluated the effect of introducing uncertainty in the prevalence p_{ij} estimates for our campylobacteriosis model. As a consequence in this model attribution confidence intervals were moderately wider and point estimates slightly changed for two of the sources. By incorporating this additional layer of uncertainty the model now more closely reflects the true uncertainty in the risk estimates.

5.6.4 Novel approach to prevalence estimation

In general the Hald model requires intensive monitoring of all relevant food animals, foods and humans followed by the application of discriminatory epidemiological typing methods (Hald 2002). The general concept however is applicable where the consequences of using biased or sparse data are being considered. Firstly a distinction has to be made between data quality (or representativeness) and data quantity (or sparsity). Sparse data will result in a low precision of the prevalence estimate and as a consequence a higher uncertainty in the whole model. On the other hand un-representative data will lead to bias in the estimates and this effect will need to be carefully discussed. The New Zealand *Salmonella* animal food source dataset has been created from several fragmented data sources from different origins. This data can clearly never be as

representative as large scale standardized surveillance. In practice however, even if a known major source of a disease is closely monitored, this is almost never the case for all known sources, in particular for minor sources such as game birds. The approach we developed has enabled us to formally estimate prevalence in a source, including the uncertainty surrounding our estimate. By objectively combining several sets of existing information we could maximise our knowledge and come up with the best possible estimate. Applying this approach has also helped us to identify data gaps and to evaluate existing surveillance systems. A way forward would be to address data quality and quantity in the analysis, by incorporating them in the prior for a_j and p_{ij}.

5.6.5 Splitting data into different time periods

To achieve identifiability, the observation period can be divided in a number of intervals and prevalence estimated separately while q_i and a_j are assumed to be constant over time, as we have illustrated in our salmonellosis model. In our campylobacteriosis model we estimate the contribution of different sources to the human disease burden over a three year period, while the original Hald model produces an annual estimate. Overall the model offers the opportunity for dynamic attribution modelling in time and this is planned to be further explored by our research group. In particular for disease with a seasonal component, such as campylobacteriosis, this would be of great interest and could greatly contribute to our understanding of disease dynamics.

5.6.6 Avoid food consumption weights and modelling environmental sources

We have reasoned that it is without major consequences to avoid food consumption weights by taking the M_j out of the model, but it will make the a_j less comparable. Removing the M_j may be necessary when no reasonable and comparable consumption estimates are available, for example when environmental sources are considered. It is worth mentioning that in the original Hald model M_j is either measured in tonnes (for meat sources) or counts (for eggs). Since disease transmission from a farm animal source may be via both food and non-food pathways, farm animal sources may be better represented taking this approach. Estimating the contribution from farm animal reservoirs (or amplifying hosts) by using the distribution of genotypes present in both food and on-farm faecal material we attempt to capture the contribution from both food and non-food pathways. The importance of such non-food pathways is underlined by assessments of exposure to *Campylobacter* (Evers, Van Der Fels-Klerx et al. 2008). This adaptation offers many opportunities to model non-food sources and has been applied previously to include travellers as a disease source (Wahlstroem, Anderson et al. 2008). Including specific environmental pathways as additional 'sources' in the model requires careful consideration. In the case of our wild bird environmental source, this

represents a non-food pathway in which we assume humans are directly exposed to faecal material from a reservoir of non-food producing wild animals. Our water environmental source is, however, not strictly speaking a reservoir or amplifying host itself, but represents an exposure pathway that is non-food, but could contain faecal material from both food producing and non-food producing animals. Therefore some of the human cases attributed to the separate environmental category could be considered as additional cases arising from farm animals via a non-food pathway. However although some ruminant-associated and poultry-associated genotypes were isolated from environmental water, the majority belonged to genotypes associated with wildlife. We therefore conclude that, although some of the cases attributed to environmental water are likely to originate from non wildlife sources, this is likely to be a very small fraction. The most common *C. jejuni* subtype in water was ST-2381 - this genotype has not been identified anywhere other than in New Zealand water (French and Molecular Epidemiology and Veterinary Public Health Group Hopkirk Institute 2008). In addition, wild birds and water isolates showed the lowest similarity with human isolates (French and Molecular Epidemiology and Veterinary Public Health Group Hopkirk Institute 2008). Given that only 1% of cases are attributed to environmental sources, the effect of how environmental pathways and reservoirs are classified and treated in the model is likely to be negligible. Our approach has included such sources, however this needs to be better understood and more work is underway to improve our knowledge of the epidemiology of *Campylobacter* in wildlife.

5.6.7 Including potentially pathogenic subtypes

We chose to include potentially pathogenic ST's into the model which occur in the sources but have not (yet) been detected in a human sample. The model assigns each of these potentially pathogenic subtypes a low probability of causing a case. This approach assumes that these types are not apathogenic but rare human pathogens. By introducing these potentially apathogenic types into the model we account for the proportion of subtypes from a source, which are proven and common human pathogens. This further advances our knowledge of the overall importance of a source and this is reflected in the estimate. However if these types are truly apathogenic, the few estimated cases for these subtypes will be an overestimate. Due to the low number of attributed cases this accounts for, this will not affect the model much and is outweighed by the insight we gain on the relative frequency of common pathogens in the sources. As an example only 26% of the *C. jejuni* isolates typed from environmental water were types that were also found in our human samples, compared with 92% of isolates from poultry (French and Molecular Epidemiology and Veterinary Public Health Group Hopkirk Institute 2008). In our campylobacteriosis model we evaluated the sensitivity of the model using this approach. A moderate change in the estimates for some sources

could be observed. A moderate increase of the cases attributed to sources with a high proportion of types detected in humans (poultry in our example) and a reduced number of cases attributed to a source with a low proportion of types detected in humans (water in our example) could be observed.

5.6.8 Campylobacteriosis model

We have, for the first time, extended the Hald model to a pathogen other than *Salmonella*. Our campylobacteriosis model produces outputs similar to those of other approaches (French and Mullner 2008; Mullner, Spencer et al. 2008) and further work is on the way to use the results from these different models as cross-validation of our results. Our modified model shows low sensitivity against changes in the prior for a_j, as well as different grouping of the sources, which suggests a stable model. The application of the model has been supported by the quality of the available dataset as well as application of a new *Campylobacter* typing technique, which provides information on host association. Poultry has been identified as the major source of human campylobacteriosis due to *C. jejuni* in New Zealand causing an estimated 80% of human cases. The importance of bovine and ovine sources in disease transmission has been highlighted with an estimated 10% and 9%, respectively, of human cases attributed to these sources. Given the relatively low prevalence and pathogen counts in ruminant meat samples (Wong, Hollis et al. 2007; French and Molecular Epidemiology and Veterinary Public Health Group Hopkirk Institute 2008), and the age and spatial distribution of ruminant-associated human cases, they are more likely to be the result of environmental and occupational, rather than food-borne exposures which will require tailored control strategies. However our model alone cannot readily distinguish between food and non-food pathways arising from farm animal reservoirs. Further inference can only be gained from additional epidemiological and microbiological data. This underlines the importance of a holistic approach to disease prevention when multiple transmission pathways exist for an individual source. A combination of environmental, occupational and food safety interventions will be necessary to reduce the number of cases from bovine and ovine sources. Our risk attribution results do have wide credible intervals, which is only in part caused by the uncertainty in our prevalence matrix. Firstly these wide confidence intervals may reflect *C. jejuni*'s complex epidemiology and the resulting uncertainty about disease origin. Secondly our allocation was based on 481 human cases compared with 3268 cases in the Hald salmonellosis model. An increase of human cases may be a way forward to increase precision of the estimates. The model is focused on the main human pathogen *C. jejuni*, which is responsible for an estimated 90% of human campylobacteriosis cases and differs in its epidemiology and infection pathways from other *Campylobacter* species (Oporto, Esteban et al. 2007). Studies including the rarer species of *Campylobacter* will be needed to model the total burden of human

campylobacteriosis in New Zealand.

5.6.9 Salmonellosis model

We have adapted the Hald salmonellosis model to New Zealand. The major challenge we encountered was the data available to estimate prevalence of *Salmonella* subtypes in the food sources. In any of the three years the risk model has identified pork as the major source of human salmonellosis followed by poultry and beef and veal. This contradicts the fact that it is commonly believed that pork consumption is not an important animal food pathway for this disease in New Zealand due to the low prevalence and rare isolation from domestic pork and live pigs (Titus 2007)). However in the European Union pork is often implicated as a source and reservoir of human salmonellosis, and commonly involved in outbreaks (European Food Safety Authority 2008). Given that the data for pork were much sparser and more biased than for any of the other sources, the results have to be interpreted with care as this may have had a major influence on the amount of cases attributed. If only few isolates are typed from a source and these include the major human types, the sparse data will lead to an overestimate of the importance of that source. Not including a major source in the model may however cause severe bias due to cases form pork being wrongly assigned to other sources. It is therefore preferable to include the source into the model, fill existing data gaps and take the poor data quality into account when interpreting the results from the model. Layer hens (eggs) have been identified as major source of the disease elsewhere (Hald, Vose et al. 2004), but due to the absence of particular pathogenic Salmonellae in New Zealand eggs this is likely not to be the case in New Zealand and very few cases appear to be attributable to eggs. We believe that the ranking of poultry, beef and veal, lamb and mutton, and eggs by our model has provided us with valuable information about the contribution of these sources to the disease burden in New Zealand. Attribution estimates for salmonellosis in NZ include wide credible intervals. We believe this is a result of the large remaining uncertainty in the prevalence matrix and estimates will be more precise once more data become available.

Due to the differences in the surveillance scheme between New Zealand and Denmark we have been able to allocate all cases to a food source. A group exists in our dataset, which pools minor, but completely typed, Salmonellae subtypes together. In the Danish data the pooled group also contains non typed isolates and as a consequence the source of these types is unknown. We make the assumption that the minor types, which in combination account for 8-11% of domestic-sporadic human cases annually, can be proportionally attributed between the sources. If data becomes available, it is preferable to assign as many individual subtypes as possible. The original Hald model did not converge with our salmonellosis dataset. This has also been observed in other countries (Pires, Nichols et al. 2008). After modifying the model convergence was

achieved, using a high number of iterations. The validity of these estimates is highly dependent on the quality of the data used, and sources of bias should be taken into consideration when interpreting the results. This is illustrated by the apparent high attribution of pork in our study - which is likely to be biased by poor quality data on prevalence for this source. Therefore a logical next step for source attribution of salmonellosis in New Zealand will be improving on the quality of the data in these food sources. Although we have concerns about the validity of the source attribution estimates, the application of the modified Hald model to salmonellosis was a useful exercise to better understand the model, to identify and address obstacles to its adaptation, and has lead the way to its extension to another pathogen.

5.7 Conclusions

We have extended the Hald model for a more generic application, by creating a set of modifications, which can be individually applied to the model depending on the context. These modifications improve the model's identifiability by various means such as by using a hierarchical model for the bacterial parameter q_i. We have modified the model to consider the contribution from environmental, non-food pathways, including putative wildlife sources, and applied this modification to campylobacteriosis in New Zealand. In addition, novel methods have been developed to introduce uncertainty around the prevalence estimates and to thereby enable the application of this approach in the absence of a very intensive national surveillance system for all major sources of disease.

5.8 Acknowledgements

I would like to thank all members of the Hopkirk Molecular Epidemiology Team, Environmental Science and Research (ESR, in particular P. Carter and C. Kliem), Mid-Central Health, Public Health Services, MedLab Central and the New Zealand Food Safety Authority (NZFSA, in particular the Science Group and S. Wagener) for their help with this manuscript. Thanks to T. Hald and S.M Pires for providing the code for the Hald model and their support in adapting the model. Also thanks to M. Hazelton for his contribution to our discussions and Jonathan Marshall for his help with the graphs.

5.9 Appendices

Chapter 6 and Appendix D provide further details into the data and methods applied in this study.

Chapter 6

Salmonella in human cases and animal food sources in New Zealand 2000 to 2004

6.1 Summary

A Bayesian risk assessment model was developed in Denmark (Hald, Vose et al. 2004) to quantify the contribution of different animal food sources to the number of human cases of salmonellosis. This report describes the development of a dataset for an equivalent New Zealand risk attribution model. Available data from human and animal food source surveillance were collected, descriptively analysed and used to model the burden of human salmonellosis from five different animal food sources. In the process a new prevalence estimation approach was developed, which combines different types of data. This report presents and analyses the data used, and thereby completes the description of the methodology and attribution results used by Mullner et al. (2009) and described in Chapter 5.

6.2 Introduction

Salmonellosis is among the commonest causes of food borne diarrhoeal disease worldwide (Hald, Wegener et al. 2005). Illnesses caused by the majority of non thyphoidal *Salmonella* serotypes range from mild to severe gastroenteritis, however in some patients bacteraemia, septicaemia and a variety of associated longer-term conditions develop. Severe illness from salmonellosis is further exacerbated by the emergence of strains that are resistant to multiple antibiotics (FAO/WHO 2002). In industrialised countries the main source of infection are animal derived products, notably fresh meat and poultry product, whereas in developing countries contaminated water, vegetables and human contacts are believed to play a major role in disease transmission (Hald, Wegener et al. 2005). The wide distribution of *Salmonella* reflects the flexibility of the bacterium to adapt to many different environments and contributes to its success

as a multi source zoonotic agent for humans (Nielson 2004). Despite many efforts to prevent and control the disease, the incidence of human salmonellosis in most industrialized countries has remained high (Hald, Wegener et al. 2005). With increasing travel and trade *Salmonella* is easily transferred between different continents as well as within regions on the globe, and has consequently become an international trade issue (FAO/WHO 2002; Nielson 2004).

The Acute Gastrointestinal Illness (AGI) Study estimated that more than five million working days are lost each year in New Zealand due to gastrointestinal illness (Lake, Adlam et al. 2007). Salmonellosis is, after campylobacteriosis, the second most frequently reported zoonosis in New Zealand. The rate of salmonellosis notifications has ranged between 26.5 to 62.3 cases per 100,000 population between 2000 and 2008 (Institute of Environmental Science and Research Limited (ESR)). Figure 6.1 shows the number of notifications from 2000 to 2008.

Salmonella belongs to the Gram-negative Enterobacteriaceae family together with *Escherichia coli*, *Yersinia* and *Shigella* and shares with those pathogens characteristics related to structure, function and habitat. The Kauffmann-White scheme is typically used to further divide *Salmonellae* into serotypes. More than 2300 serotypes have been described worldwide and the majority of these belong to *S.* enterica subsp. enterica (Nielson 2004). Changes in the relative frequency of serotypes in different reservoirs can be observed over short periods of time, sometimes within one or two years (Pan American Health Organisation 2001). The serotype information on human isolates of *Salmonella* in New Zealand reveals a dynamic situation with two serotypes emerging since 1997: *S.* Typhimurium DT 160 and *S.* Brandenburg. The former is found amongst poultry isolates and the latter associated with abortions in sheep and occupational disease amongst farmers and farm workers. New Zealand is fortunate in having a poultry industry in which the types of *Salmonella* that have caused major problems overseas, such as *S.* Enteritidis PT 4 and *S.* Typhimurium DT 104 are not endemic in this country and import controls are in place to maintain this status. It is generally assumed that New Zealand cases of salmonellosis caused by these types are acquired overseas (Lake, Hudson et al. 2004).

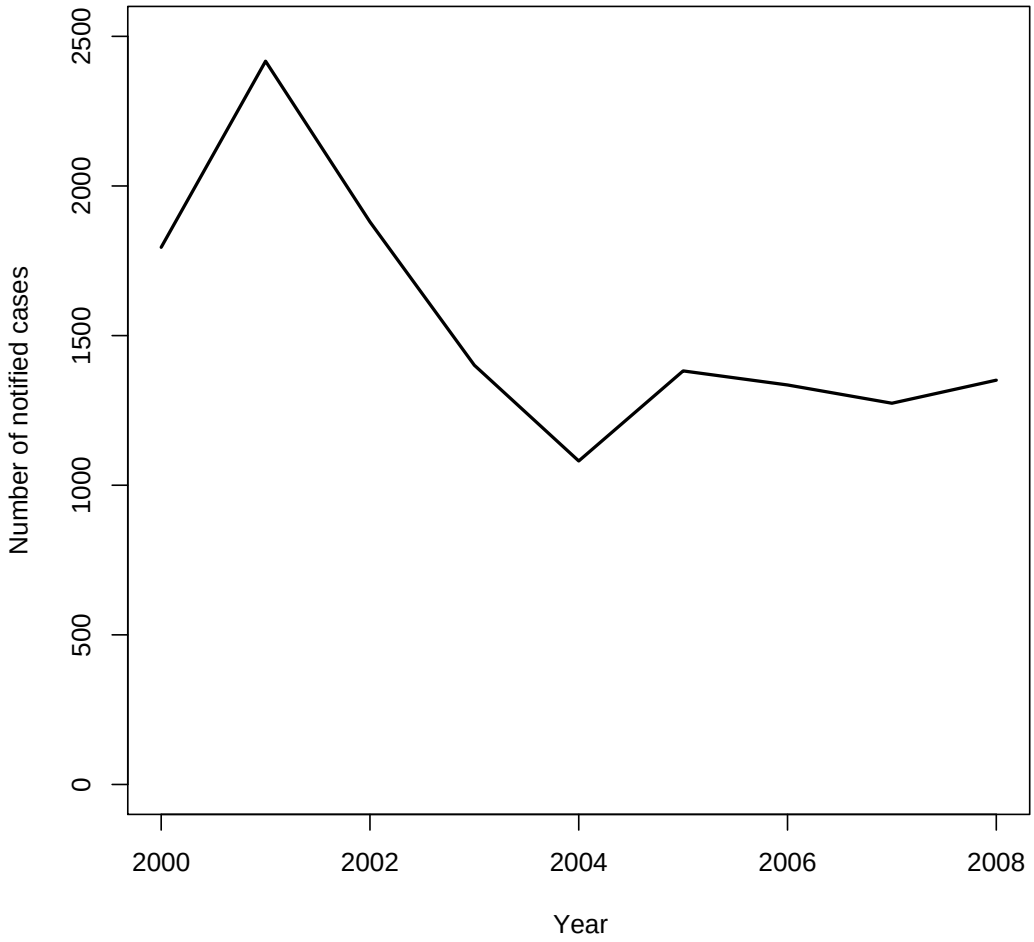

Figure 6.1: Number of notified cases of salmonellosis in New Zealand 2000 - 2008.

6.2.1 Sources of data in New Zealand to estimate the prevalence of *Salmonella* in animals, food and humans

New Zealand sources of data on *Salmonella* include the Enteric Reference Laboratory (ERL), the National Microbiological Database (NMD) as well as the notifiable disease surveillance database (EpiSurv). The Enteric Reference Laboratory (http://www.esr.cri.nz/competencies/communicabledisease/Pages/ERL.aspx) is responsible for the provision of national reference services on notifiable diseases of humans, including enteric pathogens. The data support the epidemiological surveillance of enteric organisms of public health significance. The ERL creates monthly tables of submitted human *Salmonella* isolates as well as quarterly tables of non-human isolates. EpiSurv, New Zealand's notifiable diseases surveillance database is operated by the Institute of Environmental Science and Research Ltd (ESR) on behalf of the Ministry of Health. In general, notifiable disease surveillance activities in New Zealand are carried out by both local and national authorities. Until December 2007 this surveillance was undertaken primarily by general practitioner notification to regionally based medical officers of Health. General practitioners and public health units are required to notify ESR of any cases of notifiable disease under the Health Act 1956. Although, the notifications were based upon laboratory confirmation of pathogens, they were not laboratory notified. In December 2007 significant changes in the notification process were implemented (Ministry of Health 2007). For most of the cases, subtyping information from the ERL can be linked into the EpiSurv database. Key EpiSurv data fields provide information on case demographics, clinical features and risk factors (Institute of Environmental Science and Research Limited 2007; Ministry of Health 2007).

The National Microbiological Database (NMD) commenced in 1997 as microbiological sampling strategy. It standardizes New Zealand official export assurances to overseas markets and permits evaluation of Hazard Analysis and Critical Control Points (HACCP) programmes (Duncan 2008). The National Reference Laboratory for *Salmonella* is situated within the ERL and provides summary tables of isolates from different projects. These isolates come from a broad range of sources, including diagnostic submissions of sick animals, structured national and local surveys of live animals, meats at slaughter or retail and animal feed. The results are consecutively grouped into broad categories such as 'bovine origin' before publication on the ERL website (Cook 2008).

In addition to the above, the New Zealand Food Safety Authority (NZFSA) commissions microbiological surveys, in particular to support ongoing risk profiles or risk assessments. Data from these surveys can be used to create estimates of *Salmonella* prevalence. For example a nationwide survey of 807 diced or minced samples of retail meats was conducted between 1st August 2003 and 9th November 2004 to obtain an estimate of pathogen prevalence in different meats (Institute of Environmental Science and Research Limited (ESR) 2003/2004; Cook 2008).

6.3 Material and Methods

6.3.1 Human cases

The analysis of human data was based on a matched dataset of EpiSurv reported cases and laboratory data of human (non-typhoid) cases from the ERL covering the period from 2000 - 2004. *Salmonella* types were grouped according to the scheme commonly used in NZ for surveillance purposes. Travel and outbreak status of the cases was determined using the EpiSurv database (for details see Appendix A).

6.3.2 Animal food source prevalence

Available national resources were used to estimate the prevalences of *Salmonella* serotypes in different food sources. These resources include data provided by the ERL, the NMD and several structured surveys. Data from adjacent time periods were included, assessed for their relevance and representativeness and weighted accordingly using moving weighted averages. Estimates of prevalence were obtained as described in Chapter 5.

6.3.3 Amount of food source consumed

To estimate the amount of food sources consumed data from the National Nutritional Survey (NNS) (New Zealand Food Safety Authority 1997) were updated using information obtained from the apparent consumption data collected by Meat and Wool New Zealand from 1997 to 2004 (Meat & Wool New Zealand 2006). It was assumed that the trend observed in the overall consumption would also be reflected in the consumption data. A consumption trend was calculated for each food source (e.g. a decrease in apparent consumption for eggs of 8.7% from 1997 to 2004) which was then applied to update the NNS data.

6.3.4 Analysis

Collected data was stored and manipulated in Microsoft Excel and analysed using R, version 2.7.0, and WinBUGS 14. The novel prevalence estimation approach presented in

Chapter 5 was used to estimate prevalence of *Salmonella* in food sources. Distributions were compared using two-sided χ^2 tests.

6.4 Descriptive analysis of data set: Human cases 2000 - 2004

6.4.1 Overview

Between 2000 and 2004 salmonellosis notifications declined from 2393 notified cases in 2001 to 1076 cases in 2004 (Table 6.1). The proportion of notified cases that were typed ranged from 84 to 90% and did not show significant variations in the time period under investigation. The proportion of cases categorised as travellers based varied from 9.1 - 11% in 2000 - 2003 and showed an overall high of 17.4% in 2004. However travel status was not reported for 20.1 - 27.5% of cases in any given year. Cases that were reported to be outbreak related decreased from 9.7% in 2000 to 0.7% in 2004.

Table 6.1: Typing, outbreak and travel status of human salmonellosis cases from 2000-2004.

Year	2000	2001	2002	2003	2004
Laboratory confirmed	1982	2605	2067	1601	1229
Notified (EpiSurv)	1788	2393	1906	1400	1076
% typed	88	83.5	89.1	89.1	89.1
% travelled	10.5	9.1	10.3	11	17.4
% travel status unknown	20.1	21.6	27.5	26.8	22
% outbreak related	9.7	5.1	7.7	4	0.7

6.4.2 *Salmonella* serotypes occurring in human cases

The occurrence of *Salmonella* serotypes in time showed a strong variation between individual serotypes (Table 6.2). While some serotypes such as *S.* Saintpaul and *S.* Thompson could be isolated from a constant number of cases, numbers for others such as *S.* Typhimurium DT 160 and *S.* Brandenburg varied over time. For example, *S.* Brandenburg cases declined from 168 to 82 cases during the study period, whereas cases for *S.* Typhimurium increased from 167 in 2000 to 629 in 2001 and declined again to 190 in 2004. Figure 6.2 illustrates temporal trends of selected serotypes. The pooled group of "other" serotypes was found to be large consisting of 11.2 - 20.0% of all cases in any year, including 95 different *Salmonella* types between 2002 and 2004 (Appendix D).

Table 6.2: Number of cases of *Salmonella* subtypes in 2000 - 2004.

Serotype	2000	2001	2002	2003	2004
Brandenburg	168	105	87	51	82
Enteritidis other/unknown	17	59	40	38	46
Enteritidis PT 4	37	19	29	17	16
Enteritidis PT 9a	58	61	60	56	36
Heidelberg	0	105	12	5	2
Infantis	24	55	73	75	54
Montevideo	9	3	20	33	8
Other	203	223	229	180	193
Saintpaul	18	14	31	27	29
Thompson	7	14	19	6	18
Typhimurium DT 1	110	135	197	89	53
Typhimurium DT 101	104	58	38	44	27
Typhimurium DT 135	328	215	130	58	26
Typhimurium DT 156	81	83	72	68	46
Typhimurium DT 160	167	629	485	256	190
Typhimurium other/unknown	243	221	177	244	144
All typed cases	1788	2393	1906	1400	1076

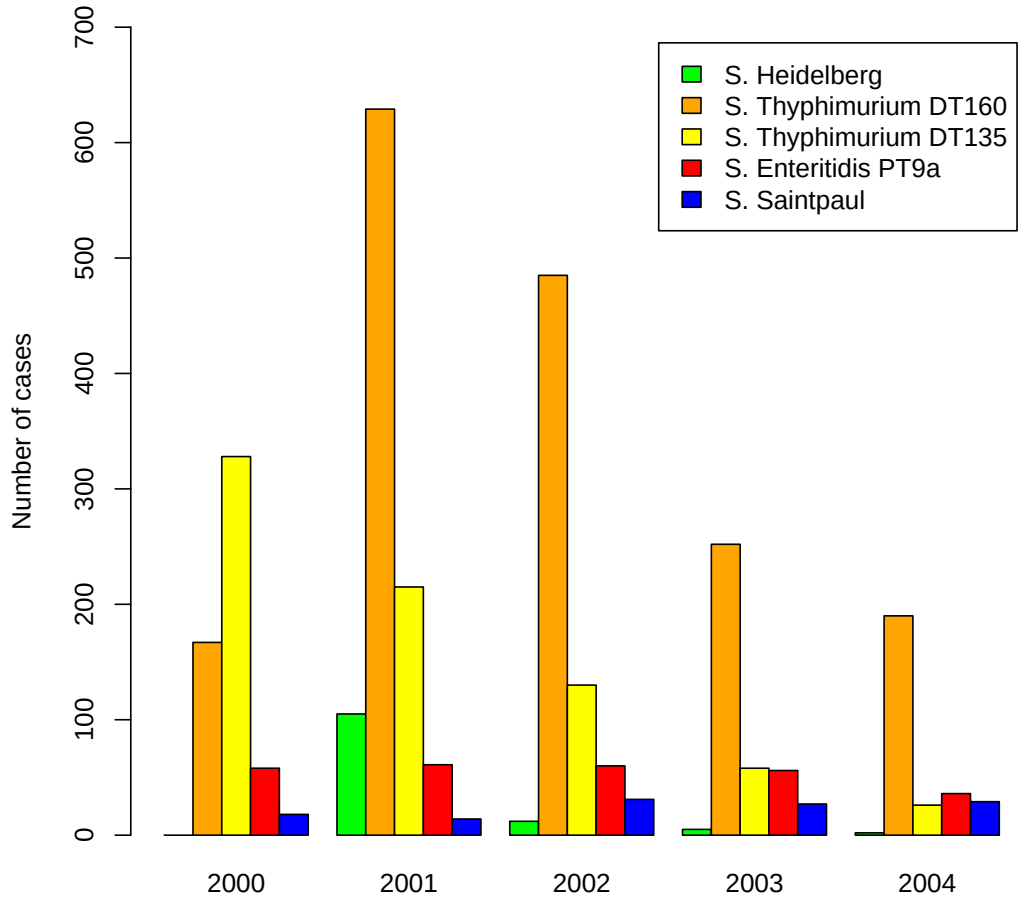

Figure 6.2: Temporal trends of selected *Salmonella* serotypes detected in humans 2000 - 2004.

6.4.3 Serotype and travel status

Table 6.3 shows a ranking of the proportion of cases of individual serotypes with a known history of travel. Travel status varied markedly between different serotypes with the proportion of travellers ranging from 0% (e.g. *S.* Brandenburg in 2001) to 76% (*S.* Enteritidis PT 4 in 2000). The differences in the proportion of travellers between serotypes were statistically significant ($p<0.0005$). A variation in time within individual serotypes could also be observed.

Table 6.3: Percentage of cases of *Salmonella* serotypes with a history of travel 2000 - 2004.

Year	2000-2004	2000	2001	2002	2003	2004
Serotype	%	%	%	%	%	%
Enteritidis other/unknown	57.5	41.2	62.7	50	68.4	65.2
Enteritidis PT 4	55.5	75.7	57.9	75.9	11.8	56.3
Other	43.2	46.3	38.1	42.8	41.1	47.7
Montevideo	39.8	22.2	66.7	45	15.2	50
Heidelberg	26.2	0	2.9	8.3	20	100
Thompson	14.5	28.6	0	5.3	33.3	5.6
Typhimurium other/unknown	7.4	3.3	8.6	9	4.1	11.8
Infantis	5.8	16.7	3.6	4.1	2.7	1.9
Saintpaul	5.1	11.1	0	0	7.4	6.9
Typhimurium DT 101	3.4	1	1.7	0	6.8	7.4
Typhimurium DT 135	2.5	3.1	1.9	2.3	5.2	0
Typhimurium DT 1	1.6	1.8	1.5	0	1.1	3.8
Brandenburg	1.1	1.2	0	0	2	2.4
Typhimurium DT 160	1	1.2	0.8	1.2	0.8	1.1
Enteritidis PT 9a	1	1.7	3.3	0	0	0
Typhimurium DT 156	0.7	0	0	0	1.5	2.2
All cases	11.7	10.5	9.1	10.3	11	17.4

6.4.4 Serotype and outbreak status

Table 6.4 shows a ranking of the proportion of cases of individual serotypes that were associated with outbreaks. This attribute varied considerably in time for different serotypes. As an example figures for *S.* Heidelberg ranged from 0 to 20% and for *S.* Typhimurium DT 135 from 2.3 to 38.6% in individual years. For some serotypes these proportions appeared to be more stable in time e.g. *S.* Infantis (0 - 3.9%) and *S.* Enteritidis PT 4 (0% in all years). The percentage of cases recorded as outbreak associated for all subtypes decreased from 9.7% in 2000 to 0.7% in 2004.

Table 6.4: Outbreak association of *Salmonella* serotypes 2000 - 2004.

Year	2000-2004	2000	2001	2002	2003	2004
Serotype	%	%	%	%	%	%
Typhimurium DT 135	14.7	5.5	6.7	38.6	2.3	5.7
Brandenburg	9.5	30.2	7.9	0	0	0
Typhimurium other/unknown	7.7	14	3.8	11.6	1.3	0
Enteritidis PT 9a	7.6	9.6	10.3	5.4	4.7	0.5
Typhimurium DT 156	5.2	4.5	5	3.4	7.8	0
Heidelberg	5	0	0	0	20	0
Typhimurium DT 101	5	0	14.3	5.3	0	0
Montevideo	2.2	5.2	0	0	3.6	0
Typhimurium DT 1	2.2	0	3.5	5.3	0	0
Typhimurium DT 160	1.9	0	1.2	0	0	6.5
Thompson	1.7	0	0	0	6.7	0
Infantis	1.5	1	0	3.9	1.1	0
Enteritidis other/unknown	1.2	2.4	0	2.3	0	0
Saintpaul	0.9	0	0	0	3.7	0
Enteritidis PT 4	0	0	0	0	0	0
Other	0	0	0	0	0	0
Overall	6.8	9.7	5.1	7.7	4	0.7

6.5 Descriptive analysis of data set - animal food sources 2002-2004

6.5.1 Overview

Thirteen of the 16 *Salmonella* subtypes, as used in the scheme for human surveillance, could be detected in the five animal food sources included in the study in any of the three years analysed. The human pathogens *S.* Enteritidis PT 4, *S.* Enteritidis other/unknown and *S.* Montevideo could not be detected in an animal food source between 2002 and 2004. *S.* Heidelberg was not found in an animal food source in 2002, but was found in beef and veal in 2003 and 2004. Table 6.5 gives an overview of the annual detection of individual subtypes. In total five serotypes were only detected in a single animal food source in any one of the three years investigated, these include *S.* Saintpaul in chicken in 2002 and *S.* Heidelberg in beef and veal in 2003. Data were generally sparser for 2004 than for the other two years. In addition results were only available for a total of 15 pork isolates, compared with 796 for chicken, the source with the highest number of isolates typed.

The prevalence of *Salmonella* serotypes in animal food sources from 2002 to 2004 was estimated using data from the NMD, ERL, MOAS study (Institute of Environmental Science and Research Limited (ESR) 2003/2004), one survey of *Salmonella* in eggs (Wilsom 2007) as well as two reports on *Salmonella* in pork (Wong 2005; Gilbert, Lake et al. 2008). Estimates and their uncertainty were obtained by using the WinBUGS implemented approach developed by Mullner et al. (Mullner, Jones et al. 2009). Prevalence estimates are expressed in percent.

Figure 6.3 shows the median of the prevalence estimates over all subtypes from 2002 - 2004 for the different animal food sources. In general prevalence of *Salmonella* was found to be low ranging from a median prevalence of 0.06 to 1.53% throughout all sources and years. Poultry was estimated to have the highest prevalence (1.53% in 2003 and 2004) followed by beef and veal (0.81% in 2003 and 2004). *Salmonella* prevalence in all red meats was well below 1% with a slightly increasing trend for lamb and mutton from 0.53% in 2002 to 0.86% in 2004.

Table 6.5: Number of *Salmonella* serotypes isolated from different animal food sources 2002 - 2004.

Year	Serotype	Source type				
		Beef/veal	Pork	Lamb/mutton	Chicken	Eggs
2002	Brandenburg	59	3	177	5	0
	Enteritidis PT 9a	3	0	0	0	0
	Heidelberg	0	0	0	0	0
	Infantis	5	0	0	53	16
	Other	25	2	116	165	10
	Saintpaul	0	0	0	1	0
	Thompson	0	0	0	3	0.2*
	Typhimurium DT 1	33	0	0	9	1
	Typhimurium DT 101	13	0	4	5	1
	Typhimurium DT 135	28	0	1	30	9
	Typhimurium DT 156	16	0	3	4	0.2*
	Typhimurium DT 160	12	0	1	92	16
	Typhimurium other/unknown	63	1	4	27	5
2003	Brandenburg	34	0	100	16	1
	Enteritidis PT 9a	2	0	1	2.5*	1
	Heidelberg	3	0	0	0	0
	Infantis	1.5*	0	2	31	19
	Other	14	0	68.5*	160	14
	Saintpaul	5	0	0	0	0
	Thompson	0	0	0	0	0.2
	Typhimurium DT 1	27.5*	0	0	67.5*	31
	Typhimurium DT 101	20	0	2	8	2
	Typhimurium DT 135	0	0	2	17	3
	Typhimurium DT 156	25	1	0	13	1.2*
	Typhimurium DT 160	6	2	0	22	6
	Typhimurium other/unknown	80	2	6	43	12

Continued on next page

Table 6.5 – *continued from previous page*

Year	Serotype	Source type				
		Beef/veal	Pork	Lamb/mutton	Chicken	Eggs
2004	Brandenburg	6	3	5	0	0
	Enteritidis PT 9a	0	0	0	1.5*	0
	Heidelberg	0	0	0	0	0
	Infantis	1.5*	0	2	1	6
	Other	2	0	1.5*	12	0
	Saintpaul	0	1	2	0	0
	Thompson	0	0	0	0	0.2*
	Typhimurium DT 1	1.5*	0	0	4.5*	0
	Typhimurium DT 101	0	0	0	0	0
	Typhimurium DT 135	0	0	0	0	0
	Typhimurium DT 156	1	0	0	0	0.2*
	Typhimurium DT 160	0	0	1	0	1
	Typhimurium other/unknown	2	0	0	3	0
Total		489	15	499	796	156.2*

* non-integer numbers are the result of using weighting moving averages

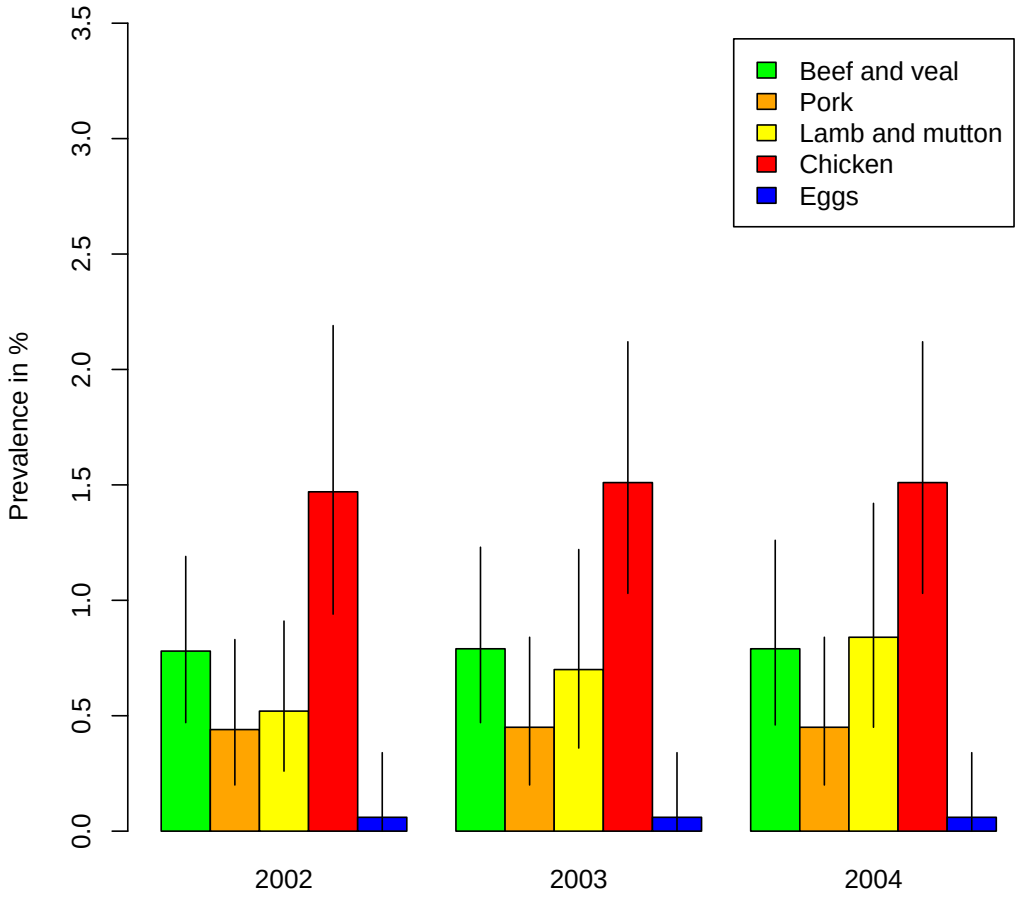

Figure 6.3: Estimated median *Salmonella* prevalence in different food sources 2002 - 2004.

6.5.2 Beef and veal

Twelve different *Salmonella* subtypes were isolated from beef and veal samples between 2002 and 2004. *S.* Thompson was the only serotype not found in beef and veal and total of 489 isolates were typed from this source (Enteric Reference Laboratory 1995 - 2004). Prevalence of *Salmonella* in beef and veal was estimated by using data from the MOAS study, data obtained from the ERL as well as by using information from the NMD. *S.* Typhimurium other/unknown was estimated to be the most prevalent subtype group with a median prevalence between 0.18 and 0.27% followed by *S.* Brandenburg (0.13 - 0.18%) (Table 6.6).

Table 6.6: Median prevalence estimates and 95% CI for *Salmonella* serotypes in beef and veal in 2002, 2003 and 2004.

Year	2002	2003	2004
Salmonella serotype			
Brandenburg	0.18 (0.11-0.29)	0.13 (0.07-0.22)	0.13 (0.07-0.22)
Enteritidis PT 9a	0.01 (0-0.02)	0.01 (0-0.03)	0.01 (0-0.03)
Heidelberg	0.00 (0-0.01)	0.01 (0-0.03)	0.01 (0-0.04)
Infantis	0.01 (0-0.03)	0.01 (0-0.03)	0.01 (0-0.04)
Other	0.08 (0.05-0.14)	0.05 (0.03-0.10)	0.05 (0.02-0.10)
Saintpaul	0.01 (0-0.02)	0.01 (0-0.02)	0.02 (0.01-0.05)
Thompson	0.00 (0-0.01)	0.00 (0-0.01)	0.00 (0-0.02)
Typhimurium DT 1	0.08 (0.05-0.14)	0.01 (0.05-0.17)	0.09 (0.05-0.18)
Typhimurium DT 101	0.05 (0.02-0.08)	0.06 (0.03-0.11)	0.06 (0.03-0.12)
Typhimurium DT 135	0.05 (0.03-0.10)	0.02 (0.01-0.05)	0.00 (0-0.02)
Typhimurium DT 156	0.06 (0.03-0.10)	0.08 (0.04-0.14)	0.09 (0.04-0.16)
Typhimurium DT 160	0.04 (0.02-0.07)	0.02 (0.01-0.05)	0.02 (0.01-0.05)
Typhimurium other/unknown	0.18 (0.11-0.29)	0.25 (0.14-0.41)	0.27 (0.15-0.45)

6.5.3 Pork

Five *Salmonella* subtypes were isolated from the small number of pork samples typed, namely: *S.* Brandenburg, *S.* Saintpaul, *S.* Other/unknown, *S.* Typhimurium DT 156, *S.* Typhimurium DT 160 and *S.* Typhimurium. Prevalence of *Salmonella* in pork was estimated by using data from the MOAS study, data obtained from the ERL as well as a *Salmonella* in pork risk profile and a microbiological survey on imported and domestic pork. The most prevalent types in pork were *S.* other/unknown, *S.* Brandenburg and *S.* Typhimurium other/unknown with a median estimated prevalence ranging from 0.03 to 0.07% (Table 6.7).

Table 6.7: Median prevalence estimates and 95% CI for *Salmonella* serotypes in pork in 2002, 2003 and 2004.

Year	2002	2003	2004
Salmonella serotype			
Brandenburg	0.06 (0.01-0.19)	0.03 (0.00-0.14)	0.07 (0.01-0.20)
Enteritidis PT 9a	0.02 (0-0.10)	0.02 (0-0.10)	0.02 (0-0.11)
Heidelberg	0.02 (0.00-0.10)	0.02 (0-0.10)	0.02 (0-0.10)
Infantis	0.02 (0-0.10)	0.02 (0-0.10)	0.02 (0-0.10)
Other	0.07 (0-0.20)	0.05 (0.01-0.17)	0.05 (0.01-0.17)
Saintpaul	0.02 (0-0.10)	0.02 (0-0.10)	0.03 (0-0.13)
Thompson	0.02 (0-0.10)	0.02 (0-0.10)	0.02 (0-0.10)
Typhimurium DT 1	0.02 (0-0.10)	0.02 (0-0.10)	0.02 (0-0.10)
Typhimurium DT 101	0.02 (0-0.10)	0.02 (0-0.10)	0.02 (0-0.1)
Typhimurium DT 135	0.02 (0-0.10)	0.02 (0-0.10)	0.02 (0-0.10)
Typhimurium DT 156	0.02 (0-0.11)	0.03 (0-0.13)	0.02 (0-0.11)
Typhimurium DT 160	0.03 (0-0.12)	0.05 (0.01-0.17)	0.03 (0-0.13)
Typhimurium other/unknown	0.04 (0.01-0.15)	0.05 (0.01-0.17)	0.03 (0-0.131)

6.5.4 Lamb and mutton

Ten different *Salmonella* subtypes could be isolated from lamb and mutton samples from 2002 to 2004. In total 499 isolates were typed from this source. Prevalence of *Salmonella* in lamb and mutton was estimated by using data from the MOAS study, data obtained from the ERL as well as information from the NMD. The two most commonly subtypes detected in this food source were *S.* Brandenburg (median prevalence 0.32 - 0.43%) and *S.* other/unknown (median prevalence 0.16 - 0.29%) (Table 6.8).

Table 6.8: Median prevalence estimates and 95% CI for *Salmonella* serotypes in lamb and mutton in 2002, 2003 and 2004.

Year	2002	2003	2004
Salmonella serotype			
Brandenburg	0.32 (0.16-0.57)	0.37 (0.19-0.65)	0.43 (0.22-0.73)
Enteritidis PT 9a	0.00 (0-0.01)	0.00 (0-0.02)	0.01 (0-0.03)
Heidelberg	0.00 (0-0.01)	0.00 (0-0.01)	0.00 (0-0.02)
Infantis	0.00 (0-0.01)	0.00 (0-0.03)	0.02 (0-0.04)
Other	0.16 (0.08-0.28)	0.25 (0.13-0.44)	0.29 (0.15-0.51)
Saintpaul	0.00 (0-0.01)	0.00 (0-0.02)	0.01 (0-0.03)
Thompson	0.00 (0-0.01)	0.00 (0-0.01)	0.00 (0-0.02)
Typhimurium DT 1	0.00 (0-0.01)	0.00 (0-0.01)	0.00 (0-0.02)
Typhimurium DT 101	0.01 (0-0.02)	0.01 (0-0.03)	0.01 (0-0.04)
Typhimurium DT 135	0.00 (0-0.01)	0.01 (0-0.02)	0.01 (0-0.03)
Typhimurium DT 156	0.00 (0-0.01)	0.00 (0-0.02)	0.00 (0-0.02)
Typhimurium DT 160	0.00 (0-0.01)	0.00 (0-0.02)	0.01 (0-0.03)
Typhimurium other/unknown	0.01 (0-0.02)	0.02 (0.01-0.05)	0.03 (0.01-0.07)

6.5.5 Chicken

The source with the highest number of isolates was chicken with typing data available for 796 isolates, and twelve different *Salmonella* types were detected. Prevalence of *Salmonella* in poultry was estimated by using data from the MOAS study, data obtained from the ERL as well as information from the NMD database. *S.* Other/unknown, *S.* Infantis, *S.* Typhimurium DT 1, DT 135 and DT 160 were the most frequently found subtypes in this food source out of which *S.* Other was the most prevalent with estimates ranging from 0.44 to 0.62% (Table 6.9).

Table 6.9: Median prevalence estimates and 95% CI for *Salmonella* serotypes in chicken in 2002, 2003 and 2004.

Year	2002	2003	2004
Salmonella serotype			
Brandenburg	0.04 (0.02-0.08)	0.05 (0.03-0.10)	0.07 (0.04-0.13)
Enteritidis PT 9a	0.01 (0-0.02)	0.01 (0-0.03)	0.02 (0-0.04)
Heidelberg	0.00 (0-0.01)	0.01 (0-0.01)	0.00 (0-0.02)
Infantis	0.16 (0.10-0.26)	0.14 (0.08-0.22)	0.12 (0.07-0.20)
Other	0.44 (0.27-0.67)	0.61 (0.41-0.87)	0.62 (0.41-0.90)
Saintpaul	0.01 (0-0.02)	0.00 (0-0.02)	0.00 (0-0.02)
Thompson	0.02 (0.01-0.05)	0.01 (0-0.02)	0.00 (0-0.02)
Typhimurium DT 1	0.09 (0.05-0.16)	0.21 (0.13-0.32)	0.25 (0.16-0.38)
Typhimurium DT 101	0.02 (0.01-0.05)	0.03 (0.01-0.06)	0.03 (0.01-0.07)
Typhimurium DT 135	0.24 (0.14-0.37)	0.07 (0.04-0.13)	0.05 (0.03-0.10)
Typhimurium DT 156	0.02 (0.01-0.05)	0.04 (0.02-0.08)	0.05 (0.02-0.10)
Typhimurium DT 160	0.27 (0.17-0.42)	0.14 (0.09-0.22)	0.08 (0.05-0.15)
Typhimurium other/unknown	0.12 (0.07-0.19)	0.15 (0.09-0.23)	0.15 (0.09-0.24)

6.5.6 Eggs

Eleven *Salmonella* types were found in egg samples between 2002 and 2004 (Table 10). Prevalence of *Salmonella* in eggs was estimated by using data obtained from the Enteric Reference Laboratory (ERL) as well as information from a Risk profile of *Salmonella* both in and on eggs which included two separate surveys (Institute of Environmental Science and Research Limited (ESR) 2004). *S.* Infantis was the most prevalent subtype in eggs (0.02%) (Table 6.10).

Table 6.10: Median prevalence estimates and 95% CI for *Salmonella* serotypes in eggs in 2002, 2003 and 2004.

Year	2002	2003	2004
Salmonella serotype			
Brandenburg	0 (0-0.01)	0 (0-0.01)	0(0-0.01)
Enteritidis PT 9a	0 (0-0.01)	0 (0-0.01)	0 (0-0.01)
Heidelberg	0 (0-0.01)	0 (0-0.01)	0 (0-0.01)
Infantis	0.02 (0-0.11)	0.02 (0-0.12)	0.02 (0-0.11)
Other	0.01 (0-0.05)	0.01 (0-0.04)	0.01 (0-0.04)
Saintpaul	0 (0-0.01)	0 (0-0.01)	0 (0-0.01)
Thompson	0 (0-0.01)	0 (0-0.01)	0 (0-0.01)
Typhimurium DT 1	0 (0-0.03)	0.01 (0-0.06)	0.01 (0-0.07)
Typhimurium DT 101	0 (0-0.01)	0 (0-0.01)	0 (0-0.01)
Typhimurium DT 135	0.01 (0-0.04)	0 (0-0.02)	0 (0-0.02)
Typhimurium DT 156	0 (0-0.01)	0 (0-0.01)	0 (0-0.01)
Typhimurium DT 160	0.01 (0-0.06)	0.01 (0-0.04)	0.01 (0-0.03)
Typhimurium other/unknown	0 (0-0.02)	0.01 (0-0.03)	0.01 (0-0.03)

6.6 Descriptive analysis of data set - food consumption data

Food consumption was estimated by using data from the National Nutritional Survey (NNS) 1997 (New Zealand Food Safety Authority 1997), which was updated with the trend observed from the apparent consumption figures from 1997 to 2004 as generated by Meat and Wool New Zealand from 1997 to 2004 (Meat & Wool New Zealand 2006). Beef and veal was estimated to be the food consumed in the largest quantity (74.1 g / person / year) followed by poultry (49.9 g / person / day). The time trend in apparent consumption showed a high variation between the different sources. For example poultry consumption increased by 45.1% between 1997 and 2004 while over the same time period lamb and mutton consumption decreased by 26.5% (Table 6.11).

Table 6.11: Estimated food consumption in New Zealand 2002 - 2004.

	Beef/veal	Pork	Lamb/mutton	Chicken	Eggs
NNS g/person/day	87.9	32.3	13.7	34.4	23.2
Proportion of total	45.9	16.87	7.15	17.96	12.11
Time trend 1997-2004(%)	-15.7	24.09	-26.5	45.06	-8.7
Updated proportion	37.94	20.52	5.15	25.55	10.84
updated g/person/day	74.1	40.08	10.07	49.9	21.18

6.7 Discussion

The descriptive analysis of the human dataset showed that both endemic strains (i.e. *S.* Saintpaul), that produce a relatively constant number of cases every year, and epidemic strains (e.g. *S.* Heidelberg), were present in New Zealand from 2000 to 2004. This is a pattern commonly observed for *Salmonella* (Pan American Health Organisation 2001) and it is likely that some reservoirs produce a constant number of human cases in time, while others change their contribution, or new reservoirs for individual strains are emerging. Possible triggers for these changes could be the introduction of new serotypes to New Zealand in new food products, or changes in food handling practices and consumption patterns.

The travel status of a *Salmonella* case was determined by the EpiSurv questionnaire. This attribute was used in the analysis to determine the number of domestic cases of *Salmonella* for use in the risk attribution model. It is commonly assumed that cases with a history of travel during the incubation period have acquired the infection overseas and, as a consequence the incidence of such cases is less likely to be reduced by domestic interventions. Travel association varied between serotypes and indicates different transmission pathways of individual serotypes. However it must be borne in mind that the variable 'previous travel' is merely an indicator that the infection may have been acquired overseas. Travel status was not available for all cases but could be estimated where the history was unknown (Mullner, Jones et al. 2009).

Outbreak association was used to determine the number of sporadic cases for the risk attribution model and it could be observed that the proportion of outbreak related cases varied by serotype. Interestingly the peak of cases of *S.* Heidelberg in 2001 could not be attributed to an increased number of outbreak-related cases, which leads to the hypothesis that the excess cases were sporadic in nature. However they may have been associated with a intermittent common source and thereby although apparently sporadic, linked in a generalised epidemic.

On the other hand data for *S.* Typhimurium DT 1 showed an increased percentage (39%) of outbreak related cases in a year (2002) with a higher than average number of cases. This supports the hypothesis that the increase in cases was due to an increased number of outbreaks. Whether a rise in cases is due to an increase in sporadic cases or due to an increase in outbreak cases provides important information for the development of control strategies. Although the outbreak status was known for all cases, this attribute may be subject to underreporting. The observed variations in outbreak association could therefore indicate a true variation in the epidemiology of individual serotypes or could be due to differences in reporting in individual years. In general it is reasonable to assume that a proportion of outbreak related cases are not reported as such, which is of particular importance when underreporting is linked to the serotype or any epidemiological criteria. The proportion of outbreak related cases decreased from

10% in 2000 to 1% in 2004. It would be of interest to investigate if the decline over time was caused by changes in the surveillance system, or if other underlying causes can be identified.

The large number of human *Salmonella* types classified as other serotypes (11 - 20% of human cases) included 95 different *Salmonella* types, which may suggest a range of infection pathways. In addition, this group showed a low proportion of outbreak related cases (1 - 4%) and a high proportion of reported travellers (38 - 48%). This leads to the hypothesis that these *Salmonella* types are more likely to be acquired overseas, which is supported by the popularity of New Zealand as a holiday destination, and the fact that New Zealanders themselves frequently travel overseas. As the number of cases of these *Salmonella* types did not decrease in time like the total numbers of cases, these presumably 'exotic' cases may show a distinct epidemiology that is not influenced by changes in the overall case numbers. However some of the types in this category are likely to originate from (minor) New Zealand reservoirs and sources such as imported foods. Pooling a large number of serotypes into one category made it impossible to make any hypothesis beyond the group level, and a more detailed data and analysis will be required to make further inferences about the epidemiology of these cases.

The routine human surveillance data and the high proportion of subtyped cases allowed for a detailed descriptive analysis at the serotype level. However as with all observational studies sources of bias should be carefully considered (Rothman and Greenland 1998). Notified cases only represent the tip of the pyramid, cases that have been picked up by the national surveillance systems. This may results in higher notification rates in children or communities with a higher socio-economic status. In addition the collection of epidemiological data based on questionnaires may be biased by the interview situation and by low response rates (Eaker, Bergstrom et al. 1998).

To model the risk of human salmonellosis from different sources the Hald source attribution model (Hald, Vose et al. 2004) used data from Denmark's intensive animal surveillance system. The large amount and the quality of data collected allowed in their case to use point estimates of the prevalence of different *Salmonella* subtypes in major animal food sources. Such good data does not exist in all countries, and the estimation of prevalence, including the uncertainty which surrounds them, can be a challenging task. Available data may originate from several surveys for example covering different time periods, sample numbers and data types and can therefore not easily be combined. In our approach animal food source prevalence of the different *Salmonella* types was estimated by using different national resources and the data were modelled using a newly developed technique which combines different types of data and also allows for uncertainty in the prevalence estimates (Chapter 5). A major problem with using different datasets is that they can differ with respect to the time period they were assessing, the number of samples taken as well as the sampling strategy (different de-

nominator). The assumption has to be made that subtypes from the different datasets, which might include diagnostic submissions from live animals, as well as retail meat samples reflect human exposure from a source (or reservoir). In consequence an increase in the frequency of individual subtypes could be the result of an increase of diagnostic submissions from samples from aborted calves rather than a higher contamination of meat products for human consumption. This would results in a higher contamination of the animal environment, and a higher risk of infection through for example occupational exposure, rather than through exposure from the consumption and preparation of meat.

In common with a previous publication based on food microbiological surveys in New Zealand (Wong, Nicol et al. 2007), the prevalence of *Salmonella* in animal food sources was estimated to be low ranging from 0.5 to 1.5% in the five animal food sources investigated. In international comparison these prevalence rates, in particular the ones for poultry, are low. (Fox, Reinstein et al. 2008; Kegode, Doetkott et al. 2008; Little, Richardson et al. 2008; Snow, Davies et al. 2008). However both New Zealand studies are based on similar data sources and the results from these studies will need to be confirmed by the provision of more data to make international comparisons with more confidence. In addition the low exposure is in contrast with the moderate notification rates observed in New Zealand. The rate of human notification could be driven by non domestic food sources and alternate infection pathways including imported foods and travel however this hypothesis would need to be supported by further studies. A recent report on *Salmonella* exposure and pathways has explored this in some detail (New Zealand Food Safety Authority (NZFSA) 2007).

In this analysis prevalence data was pooled over several years, assuming that the occurrence of *Salmonella* is constant in time. However the occurrence of individual *Salmonella* subtypes is known to vary over time and changes in the prevalence of individual types would be expected. Since we assumed that these changes would not occur very suddenly, data from adjacent time periods were included and weighted accordingly. The number of collected and typed samples included in our analysis varied between sources and this was reflected by the precision of the prevalence estimates. Pork in New Zealand is not routinely monitored and a very small number of isolates have been typed from that source. This makes it very difficult to make reliable inferences about this source in disease transmission and this has also impacted on the performance of the risk attribution model present in Chapter 5 (Mullner, Jones et al. 2009). The inclusion of datasets was judged by expert opinion and moving weighting averages were used to allow for differences in sample size, representativeness as well as time period assessed. Making use of such 'imperfect' and fragmented data enabled us to identify data gaps as well as has given us the opportunity to get an overview of ongoing surveillance activities.

Since the food consumption data from the NNS was generated in 1997, it was considered necessary to update the data. Changes in consumption figures over time are to be expected and will be the result of factors such as changes in dietary habits and consumer perception, and/or a change in availability and price of certain food items. More current estimates of apparent consumption from Meat and Wool New Zealand, based on production data, were used to do this. Although apparent consumption figures differed from the NNS data it can be assumed that the underlying changes in time affect both estimation approaches equally.

6.8 Conclusions

I have provided a detailed descriptive epidemiological analysis of *Salmonella* serotypes isolated from human cases and animal food sources in New Zealand from 2000 to 2004 and presented results from a novel approach to prevalence estimation. These studies revealed differences in the proportion of travel and outbreak associated cases for the individual *Salmonella* subtypes, and the implication of these findings for disease control were discussed. The prevalence of *Salmonella* in New Zealand animal food sources was estimated to be low. This low exposure to *Salmonella* from domestic animals food sources is in contrast with the moderate notification rates observed in New Zealand and needs further exploration. The fraction of human cases caused by a multitude of minor ('other') serotypes indicate the existence of a variety of uncommon exposure pathways such as imported food or international travel. Data quality was a problem, in particular for pork sources, and further data and work will be necessary to support the hypotheses generated by this report.

6.9 Acknowledgements

I would like to thank Carol Kliem (ESR, Kenepuru) for providing the human case data and Carolyn Nicols (ESR, Kenepuru) for her help with the food source surveillance data and her general support. Additional intellectual input was provided by Terry Ryan, Roger Cook and Peter van der Logt (all NZFSA Science Group). I would also like to thank Stuart MacDiarmid (now MAF BNZ) for his help with the data

6.10 Appendices

Appendix D provides further details into the data and methods applied in this chapter.

Chapter 7

A comparative modeling approach to identify the origin of human campylobacteriosis in New Zealand

Petra Mullner, Simon E. F. Spencer, Daniel J. Wilson, Geoff Jones, Alasdair D. Noble, Anne C. Midwinter, Julie M. Collins-Emerson, Philip Carter, Steve Hathaway, Nigel P. French. A comparative modeling approach to identify the origin of human campylobacteriosis in New Zealand. Manuscript submitted to *Emerging Infectious Diseases* on 16th April 2009.

7.1 Abstract

Until recently New Zealand had extraordinarily high and increasing rates of notified human campylobacteriosis, and our limited understanding of the source of these infections was hindering efforts to control this disease. Genetic and epidemiological modeling of a three-year dataset comprising multilocus sequence typed isolates from human clinical cases, coupled with concurrent data on food and environmental sources, enabled us to estimate the relative importance of different sources of human disease. Our studies provided strong evidence that poultry was the leading cause of human campylobacteriosis in New Zealand, causing an estimated 58-74 % of cases with widely-varying contributions by individual poultry suppliers. In contrast environmental sources played a relatively minor role. These findings influenced national policy making and, after the implementation of poultry-industry-specific interventions, a dramatic decline in human notified cases was observed in 2008. The comparative-modeling framework proposed in this study can be extended to other pathogens and provides new opportunities for disease management and control.

7.2 Introduction

Knowledge of the proportion of human cases that are caused by a particular exposure source is critical for the prioritization of public health resources and the successful implementation of disease control strategies (Batz, Doyle et al. 2005). However, estimating the contribution of different pathogen reservoirs to human infection is challenging, particularly when there are many risk pathways and sources of infection. The need to identify the likely origin of human cases in such a complex environment led to the development of model-based tools using Bayesian inference applied to multi-source microbial subtyping (Hald, Vose et al. 2004). Subsequently these tools have been modified and applied to other pathogens (Mullner, Jones et al. 2009) and a new generation of genetic attribution tools have emerged (Wilson, Gabriel et al. 2008; Sheppard, Dallas et al. 2009; Strachan, Gormley et al. 2009).

Molecular tools are now being increasingly applied to study transmission patterns within populations at the strain level, and to evaluate host- and strain-specific risk factors. This is changing the way epidemiologists study infectious diseases (Murray 2002). The Hald model (Hald, Vose et al. 2004) is an associative simulation model that utilizes the distribution of bacterial subtypes in potential sources of disease to estimate the contribution of each source to the human disease burden, while accounting for factors including differences between the ability of sources to transmit disease. On the other hand population genetic models, such as those presented by Wilson and Sheppard (Wilson, Gabriel et al. 2008; Sheppard, Dallas et al. 2009), use the relative frequency and relatedness of isolates from different sources to infer attribution estimates.

The availability of molecular typing schemes that identify strong associations with animal hosts, combined with new modeling tools based on different underlying assumptions, has provided a novel platform for understanding the origin of human infections and informing public health policy. We have used this comparative modeling approach to study the origin of human campylobacteriosis in New Zealand.

Campylobacteriosis is a major public health problem worldwide and in New Zealand the number of notifications has increased markedly over the last decade. In 2005 and 2006 the incidence exceeded 300 cases per 100,000 people per annum (Baker, Wilson et al. 2006). As a consequence the relatively high prevalence of campylobacteriosis in New Zealand attracted considerable media attention, was regarded as a national epidemic, and raised a public demand for urgent action (Baker, Wilson et al. 2006). Until recently the complex epidemiology of campylobacteriosis, and lack of an appropriate subtyping scheme had hindered the development of successful measures to control this pathogen in New Zealand and elsewhere (Mullner, Jones et al. 2009). The situation in New Zealand is quite unique: the country is geographically remote, with extensive agricultural land use and human, animal and pathogen populations that are relatively isolated. Although there is relatively little importation of animals and animal products into the country as

a result of rigid border biosecurity measures, the country is exposed to a large amount of international travelers. The country's poultry supply is different to many developed countries in that suppliers are almost entirely focused on the domestic market and for biosecurity reasons no raw poultry products are imported into the country.

Environmental samples and food sources were gathered in a defined geographical area of New Zealand over a three year period (French and Molecular Epidemiology and Veterinary Public Health Group Hopkirk Institute 2008) and genotyped using multilocus sequence typing (MLST) (Dingle, Colles et al. 2002). The resulting dataset contained a total of 969 typed samples of which 502 were from human cases. Using data from this study, four different methods were applied and where necessary modified to assess the relative contribution of food and non-food sources to the burden of human disease. These were the Proportional Similarity Index (PSI) (Rosef, Kapperud et al. 1985), the Dutch model (Hald 2002), a modified Hald model (Mullner, Jones et al. 2009), and the asymmetric island model (Wilson, Gabriel et al. 2008). We focused our models on the main human pathogen *Campylobacter jejuni*, which is responsible for an estimated 90% of human campylobacteriosis cases and differs in its epidemiology and infection pathways from other *Campylobacter* species (Oporto, Esteban et al. 2007).

7.3 Methods

7.3.1 Data

Over a three year period from March 1st 2005 until February 29th 2008 a total of 2766 human, retail meat, on-farm and environmental samples were collected in the Manawatu region of New Zealand's North Island (French and Molecular Epidemiology and Veterinary Public Health Group Hopkirk Institute 2008). This included monthly sampling of retail fresh carcasses from different poultry suppliers in the region, which are dominated by two major suppliers named A[1] and B in this study. The dataset used for the attribution models consists of a subset of confirmed *C. jejuni* samples from the above mentioned study and wild bird samples from a separate study (French, Midwinter et al. 2008). The aim was to get a subset of approximately 100 isolates from each source spread over the whole study period (Table 7.1). Human surveillance information was linked to laboratory notified cases and duplicates, travelers and out of region cases were excluded from the dataset. Samples were cultured for *Campylobacter* spp, obtained isolates were speciated by PCR and if confirmed as *C. jejuni* MLST typed. Approval for this study was given by the Central Regional Ethics Committee.

[1]Supplier identification letters were assigned arbitrarily

7.3.2 Proportional similarity index

The proportional similarity index (PSI) or Czekanowski index is an objective and simple measure of the area of intersection between two frequency distributions (Rosef, Kapperud et al. 1985) that can estimate the similarity between the frequency distributions of bacterial subtypes from different sources. The PSI is calculated by: $PS = 1 - 0.5 \sum_i \mid p_1 - q_i \mid = \sum_i min(p_i, q_i)$, where p_i and q_i represent the proportion of strains belonging to type i out of all strains typed from sources P and Q (Feinsinger, Spears et al. 1981; Rosef, Kapperud et al. 1985). The value for PSI ranges between one for identical frequency distributions, to zero for distributions with no common types. Bootstrap confidence intervals for this measure were estimated (Garrett, Devane et al. 2007).

Table 7.1: Overview of samples collected in the Manawatu.

Source	Samples collected	Samples culture positive	Samples confirmed *C. jejuni*	Samples typed and included in analysis
Human cases	773	661	584	502
Fresh chicken carcasses	562	454	415	275
Retail meats beef	400	44	33	29
Retail meats lamb	418	112	103	87
On-farm cattle	145	119	80	66
On-farm sheep	133	92	61	49
Environmental water	335	140	82	70
Total samples	2766	1622	1215	1078

7.3.3 Dutch model

The principle behind the Dutch model (Hald 2002) is to compare 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:

$$\lambda_{ij} = \frac{p_{ij}}{\sum_j p_{ij}} x_i, \tag{7.1}$$

where p_{ij} is the relative occurrence of bacterial subtype i in source j, x_i is the estimated number of human cases of type i per year and λ_{ij} is the expected number

of cases per 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, Devane et al. (2007) was extended to provide bootstrap confidence intervals for the Dutch model.

7.3.4 Modified Hald model

Similar to the Dutch model, the modified Hald model 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 (Mullner, Jones et al. 2009). In the model o_i represents the number of human cases of type i and the expected number of cases of *Campylobacter* type i from source j is denoted by λ_{ij}. Assume that

$$o_i \sim Poisson(\sum_j \lambda_{ij}), \tag{7.2}$$

and that

$$\lambda_{ij} = p_{ij} q_i a_j, \tag{7.3}$$

where λ_{ij} is the expected number of cases/year of type i from source j, p_{ij} is the prevalence of type i in source j, q_i is the bacteria dependent factor for type i and a_j is the food source dependent factor for type j.

7.3.5 Island model

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 (Wilson, Gabriel et al. 2008). The technique models the zoonotic transmission of isolates between source populations using an evolutionary approach based on Wright's island model (Wright 1931). The source of human infection is inferred by assuming that each human isolate represents a sample of one of these source populations. By modeling the evolutionary processes of mutation and recombination, the approach allows for the occurrence of novel alleles, and novel combinations of alleles, in the human isolates not observed in samples from the source populations.

7.4 Results

7.4.1 Proportional similarity index

The PSI was used to compare the distribution of human genotypes with the distribution of genotypes in different disease sources (Table 7.2). Genotypes from Poultry Supplier A were significantly more similar to human genotypes than isolates from any other source (median estimate 0.58, 95% CI 0.48–0.64). The PSI was similar for the other two poultry suppliers and bovine sources, with median values between 0.32 and 0.34. The least similarity was observed between environmental and human sources (median estimate 0.18, 95% CI 0.12–0.22).

7.4.2 Dutch model

The Dutch model (Figure 7.1) estimated that the majority of human cases came from the three major poultry suppliers, with Supplier A associated with the most human cases: an estimated 34% (95% CI 28–41%). Bovine sources were estimated to cause 19% of cases (95% CI 14–23%) followed by environmental and ovine sources which were estimated to contribute 12% and 11% of cases respectively.

7.4.3 Modified Hald model

Poultry Supplier A was estimated by a modified Hald model (Figure 7.1) to cause the majority of human cases (58%, 95% CrI 25–78%). The contribution of the other poultry suppliers as well as ovine and bovine sources was estimated to be similar ranging from 7% to 11%, with cattle estimated to contribute more cases than sheep. The proportion of cases estimated to originate from environmental sources was only 1% (95% CrI 0–6%).

7.4.4 Island model

Poultry was identified by the Island model as the most important source of human infection, accounting for an estimated 76% of human cases (Figure 7.2). 63% of human case (95% CrI 55–72%) were attributable to Poultry Supplier A. Suppliers B and C were estimated to contribute only 8% and 4% respectively. The next highest contributor to the burden of human cases was estimated to be ruminants, with cattle estimated to contribute 18% of cases (95% CrI 10–26%).

7.4.5 Comparing the results from different approaches

A visual comparison of the output from all three attribution models is shown in (Figure 7.1). All three models lead to similar estimates, and these are consistent with the PSI index. However the level of precision varied between the individual models. All methods suggest that the majority of cases over the three-year period could be attributed to poultry, in particular Poultry Supplier A, providing further evidence of the importance of the contribution of this food source to the burden of campylobacteriosis in New Zealand. The next most important source was cattle, followed by sheep, with relatively minor contributions from the environment.

Table 7.2: 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 genotypes identified in the source and the human cases.

Source	PSI	Lower 95% CI	Upper 95% CI
Poultry Supplier A	0.58	0.48	0.64
Poultry Supplier B	0.32	0.26	0.36
Poultry Supplier C	0.32	0.24	0.35
Bovine sources	0.34	0.28	0.39
Ovine sources	0.28	0.22	0.32
Environmental sources	0.18	0.12	0.22

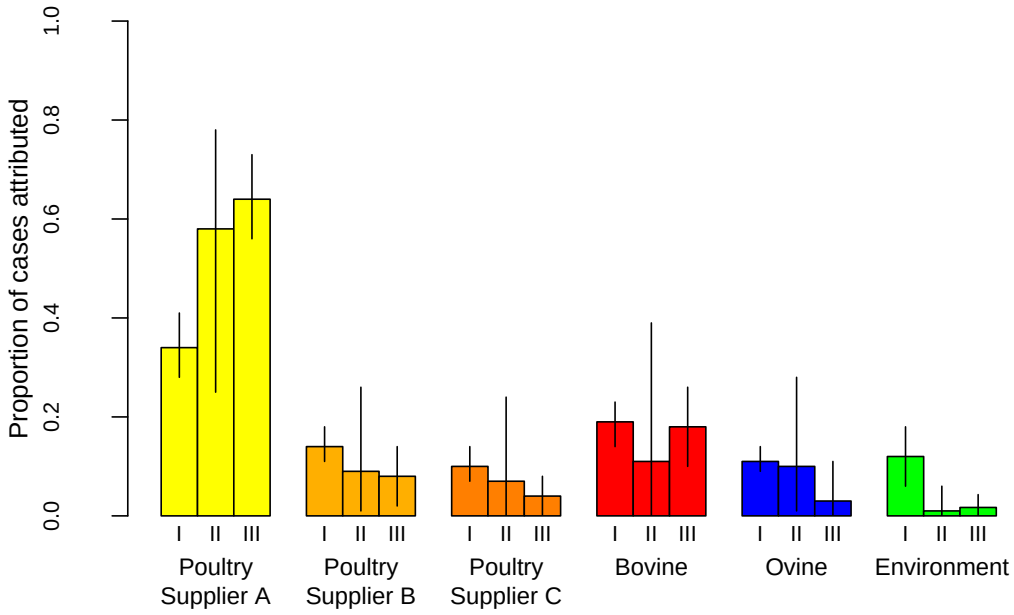

Figure 7.1: Proportion of human cases attributable to each source: comparing, from left to right, the Dutch (I), modified Hald (II) and asymmetric island model (III). Error bars represent 95% confidence / credible intervals.

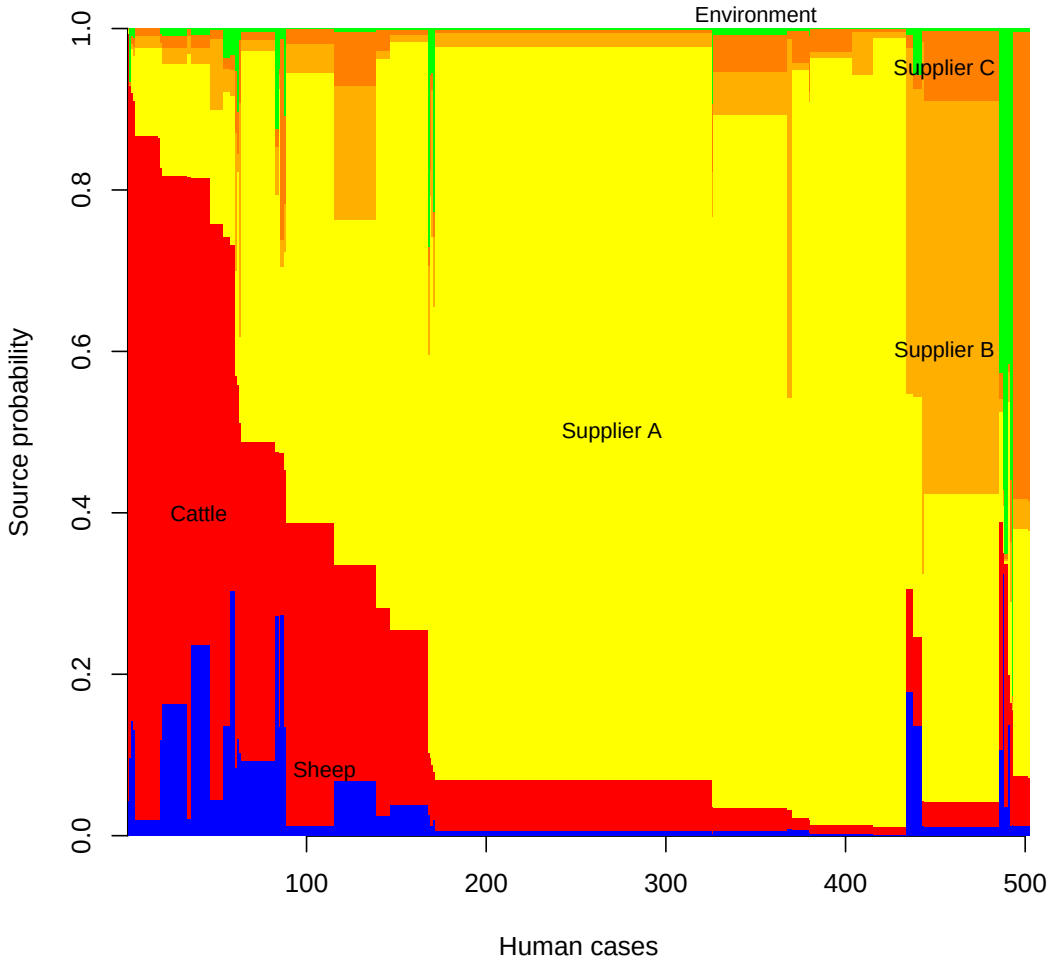

Figure 7.2: Output from asymmetric island model represented as a matrix plot. Each human case is a vertical column coloured according to the probability it came from each source.

7.5 Discussion

Estimating the relative contribution of individual sources to human infection helps public-health policy makers to decide upon the most appropriate control measures to implement. Our studies have provided such estimates for New Zealand: 58 to 76% of human cases were attributed to poultry sources, and differences in the contribution from the individual poultry suppliers were observed. The dominance of Poultry Supplier A was supported by the relatively large PSI when compared with human isolates. A major contributor to the PSI was the high frequency of a particular sequence type, namely ST-474, in this source. ST-474 accounted for 30.7% of human cases in our study and was predominantly found in samples from this supplier (French and Molecular Epidemiology and Veterinary Public Health Group Hopkirk Institute 2008). These results support previous evidence by microbiological and epidemiological studies (Wagenaar, Mevius et al. 2006; Gormley, MacRae et al. 2008; Sheppard, Dallas et al. 2009) that poultry is the major contributor to the disease burden in industrialized countries.

The large variation in the estimated contribution of individual poultry suppliers to the burden of human disease was a surprising observation. Differences in contamination levels of poultry have been found in a recent study in Belgium (Habib, Sampers et al. 2008), but to date no other study has investigated differences in risk from individual domestic suppliers by considering them as separate sources of infection. The ability to differentiate the contribution of individual suppliers was aided by the relatively simple structure of the New Zealand poultry industry, which could facilitate targeted surveillance and intervention in the future.

We chose to model the following sources: bovine, ovine, poultry (divided into the major suppliers) and environment. However, cases assigned to a source might be acquired by a variety of pathways such as retail meats, occupational exposure and environmental fecal contamination. Hence this framework cannot ascertain the contribution of particular transmission pathways, but attempts to estimate the relative contributions from amplifying animal hosts. We included wild bird and water samples to assess the contribution of potential environmental pathways to the disease burden (French, Midwinter et al. 2008). However including the environment in these types of models is problematic. Cases from the non-wildlife sources may also be acquired via environmental pathways (such as run-off from dairy or poultry farms). We found a distinctive and diverse distribution of genotypes in our environmental water samples, containing many previously unreported MLST profiles. Their allelic profiles indicate that these are likely to have originated in so far unidentified wildlife sources that are contributing to the presence of *Campylobacter* spp. in environmental water, but are seldom identified in human cases (French and Molecular Epidemiology and Veterinary Public Health Group Hopkirk Institute 2008; Mullner, Jones et al. 2009). Therefore, although environmental water cannot be considered a single amplifying host, it represents a vehicle that delivers

a unique exposure primarily from putative wildlife sources.

The importance of bovine sources in disease transmission has been highlighted with an estimated 11 to 18% of human cases attributed to this source. Given the relatively low prevalence and pathogen counts in ruminant meat samples (Wong, Hollis et al. 2007; French and Molecular Epidemiology and Veterinary Public Health Group Hopkirk Institute 2008), bovine-derived cases are commonly assumed to be the result of direct environmental and occupational, rather than food-borne exposures (Strachan, Gormley et al. 2009). This highlights the importance of combining attribution approaches with a detailed understanding of the epidemiology of disease, including aspects such as the spatial distribution of cases, or age-related risk factors (Strachan, Gormley et al. 2009). A combination of environmental, occupational and food safety interventions will be necessary to reduce the number of cases from sources where multiple transmission pathways exist. This has also been recognized for other zoonoses, such as *E.* coli O157 (Strachan, Dunn et al. 2006).

Rather than focus on a single approach, we have applied and adapted epidemiological and population-genetic models for ascertaining the source of infection. As the techniques are based on different underlying assumptions and applied to the same dataset the outputs from each model can be compared and contrasted. In this study all models provided broadly similar estimates of the contribution of each source to the burden of human infection. Further, the results are in general agreement with the findings of earlier studies conducted in New Zealand (Eberhart-Phillips, Walker et al. 1997) and elsewhere (Pearson 2000). There are some notable exceptions; for example the estimated contribution from environmental sources varies from 1% (modified Hald model) to 2% (asymetric island model) to 12% (Dutch model). The divergence of the estimate from the Dutch model might be explained by the models assumption of equal impact between the sources, which results in more evenly distributed attribution estimates and thereby in an overestimate of the contribution of environmental water sources. This is also supported by the low PSI between human and environmental sources.

In its application the PSI is the most straightforward measure to calculate, and possibly the easiest to interpret. Although this technique doesn't estimate the number or proportion of human cases attributable to each source, it may be considered a useful first-step in the assessment of source attribution, by providing a measure of the correlation between source distributions (Garrett, Devane et al. 2007). The Dutch model assumes that all strains are considered to have an equal probability of causing human disease, at a given dose in a given source, and all sources contaminated with *C. jejuni* are assumed to have an equal probability of causing disease in humans. However it is highly likely that some STs are more virulent or pathogenic than others (Pope, Wilson et al. 2007), 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, as some foods are more likely to be undercooked than others (Uyttendaele, De Troy et al. 1999; Cornelius, Nicol et al. 2005). Since the Dutch model assumes equal impact between sources, the number of cases tends to be more equally distributed among sources as compared to the other models. In contrast to the Dutch model, the modified Hald model does explicitly consider both bacteria and food-type dependent factors, and incorporates both sampling and parameter uncertainty using Bayesian inference. The estimates from our modified Hald model are associated with large uncertainty. This lack of precision is partially overcome when pooling individual poultry suppliers into one group (Mullner, Jones et al. 2009), and is the result of decreased heterogeneity of genotypes between sources when poultry suppliers are modelled individually. This also slightly changes the overall attribution results (Chapter 5). By taking a population genetics approach, the asymmetric island model (Wilson, Gabriel et al. 2008) is substantially different from the above models. The relatedness of genotypes is considered explicitly, and all human isolates can probabilistically be assigned to individual sources. The credible intervals for the island model are narrower than for the modified Hald model, most likely because more information (the allelic profiles), is used to fit the model.

In this study, a 3-year structured longitudinal study was used to apportion cases of disease to a source, which is in contrast to the cross-sectional surveys, using data from different geographical regions and time periods, reported in other studies (McCarthy, Gillespie et al. 2007; McCarthy, Shepherd et al. 2007; Wilson, Gabriel et al. 2008; Strachan, Gormley et al. 2009). In our sentinel site study 2766 samples were collected over a three year period. By simultaneously collecting data from human clinical cases and potential disease sources in a defined region we aimed to capture human exposure over a three year period. A random sample of a total of 969 samples were typed (80% of all *C. jejuni* isolates), consisting a subset of *C. jejuni* isolated in the sentinel surveillance site during the study period which were selected to be representative of each source and each month of the study. Comparative studies conducted in two other regions show a similar distribution of genotypes in both humans and poultry (French and Molecular Epidemiology and Veterinary Public Health Group Hopkirk Institute 2008). This suggests that the results from the study site are likely to be similar to that in other regions and allow for extrapolation to the whole of New Zealand.

Our attribution estimates can be compared with those of a study conducted in Lancashire, England (Wilson, Gabriel et al. 2008), the major difference being the relatively higher contribution from cattle (35% in Lancashire, 11-19% in the Manawatu) and a lower contribution from poultry sources (57% in Lancashire, 58-76% in the Manawatu). This may be explained by a higher relative consumption of poultry meat in New Zealand

and the presence of unique poultry-associated genotypes (Baker, Wilson et al. 2006). However estimates from our study are very similar to those from a study conducted in Scotland (Sheppard, Dallas et al. 2009), where poultry sources were associated with 58 - 78% of infections.

The estimated low importance of non-livestock sources estimated by these recent studies contrasts with the finding by Champion et al. (Champion, Gaunt et al. 2005). Using a whole genome micro-array, they concluded that non-livestock, non-agricultural sources may be the most important for human infection. The striking difference in the conclusion may be attributed to the selection of isolates for inclusion in the analysis and model specification (Wilson, Gabriel et al. 2008).

The evidence provided by our approach has provided an important contribution to the NZFSA *Campylobacter* Risk Management Strategy (New Zealand Food Safety Authority (NZFSA) 2007), which has subsequently included mandatory targets for limiting contamination with *Campylobacter* spp. of chilled poultry carcasses. The introduction of these interventions has coincided with a dramatic decrease in human campylobacteriosis notifications to a 16-year low. In 2008 some 6689 human cases were reported in New Zealand compared to 15,873 cases in 2006; the year before the announcement and implementation of control measures. A logical next step is to monitor changes in source attribution post-intervention to ascertain whether the relative importance of poultry has diminished. In addition further refinement of the sampling scheme and source attribution models should enable finer-scale temporal changes in attribution to be described. This will be valuable for determining seasonal changes in the relative contribution of sources (Nylen, Dunstan et al. 2002) and provide a dynamic tool for assessing the effectiveness of interventions.

In conclusion, a combination of epidemiology, molecular biology and population genetics has provided a valuable risk-based framework to inform surveillance and public health policy for the control of enteric zoonotic pathogens. Given the use of temporally and spatially appropriate sampling strategies these tools have the potential to be applied to inform decision making for the control of other zoonoses, such as cryptosporidiosis and salmonellosis. Furthermore, the risk attribution framework presented here can be modified to enable continuous, dynamic assessment of source attribution, enabling seasonal variation in infection sources to be identified, and facilitate an assessment of the effectiveness of control programs.

7.6 Acknowledgements

I would like to thank all members of the Hopkirk Molecular Epidemiology Team, Environmental Science and Research (ESR, in particular P. Carter and C. Kliem), MidCentral Health, Public Health Services, MedLab Central and the New Zealand Food Safety Authority (NZFSA, in particular the Science Group and S. Wagener) for their help

with this work. I would like to in particular acknowledge Sarah Moore (EpiCentre), Lynn Rogers (MedLab Central) and Tui Shadbolt (MidCentral Health) for their help with the data collection. The island model results for this manuscript were derived by Jonathan Marshall (Massey University), who also provided the main contribution to its graphical output.

7.7 Appendices

Appendix E provides further details into the data and methods applied in this study.

Chapter 8

General Discussion

8.1 Summary

This thesis used data from a sentinel surveillance site (the Manawatu region of New Zealand) to examine the epidemiology of human campylobacteriosis and estimate the contribution of different sources to the burden of human disease. By combining molecular epidemiology and novel approaches to source attribution this work has shed new light on the epidemiology of New Zealand's most important zoonosis and has overcome some of the challenges posed by its complex epidemiology.

In the opening chapter I reviewed subtyping methods for *Campylobacter* spp. and risk assessment approaches, and highlighted the importance of including molecular data in studies of risk. Source attribution approaches were presented and discussed. A variety of subtyping techniques are available for the identification of *Campylobacter* spp. and each was described and advantages and drawbacks discussed. MLST was identified as the most promising approach for (non-outbreak) epidemiological studies and assessment of risk (Chapter 2).

The epidemiology of human cases in the Manawatu was studied in detail and risk factors for individual *C. jejuni* genotypes could be identified. Surprisingly a dominant human clone (ST - 474) was found, which accounted for 30.7 % of human cases. The analysis revealed relative differences in the risk of infection with poultry versus non-poultry strains in rural and urban areas and identified farm animal contact as a risk factor for infection with strains not associated with poultry (Chapter 3).

This was complemented by a study of *C. jejuni* in poultry and human cases, where a high similarity of genotypes between poultry and human isolates was detected using descriptive and analytical approaches. This study also provided new knowledge about the distribution of *C. jejuni* within the New Zealand poultry industry. Probability of contamination and levels of contamination varied between individual producers and company specific MLST STs, such as ST - 474 and ST - 3609, could be identified (Chapter 4).

Using MLST data from all sampled sources their relative contribution to the num-

ber of human cases was apportioned and poultry was identified by this model to cause the majority of human cases (80 %), followed by bovine sources (10 %). An existing source attribution model was modified for this estimation and simultaneously applied to salmonellosis using routine surveillance data. This modified Hald model has many advantages over the original model, such as the potential for application to a wide range of zoonotic pathogens, an improved identifiability, and the potential to include environmental sources as well as the availability of a novel approach to estimate prevalence from fragmented datasets (Chapter 5).

Chapter 6 described in detail the data set used in Chapter 5 to apportion human cases of salmonellosis. The molecular epidemiology of human cases and food sources of salmonellosis was descriptively analysed. The large proportion of human cases caused by a multitude of minor serotypes indicated the existence of a variety of uncommon exposure pathways, such as imported food or international travel. While human case data was of high quality and quantity, data gaps for the food sources, in particular for pork, were identified.

In the next step (Chapter 7) a novel risk attribution framework was applied to the data from the Manawatu study. This approach included four different models namely, the modified Hald model (Chapter 5) (Mullner, Jones et al. 2009), the population genetics-based island model (Wilson, Gabriel et al. 2008), the frequentist Dutch model (Hald 2002) and the proportional similarity index (Rosef, Kapperud et al. 1985). The availability of a set of models and approaches allowed them to be used as cross-references. Although no gold standard was available, consistent results from different models, with different underlying assumptions, strengthened our understanding and confidence in the results and inferences (Chapter 7).

The key elements of this thesis were the application of molecular epidemiology, the use of novel source attribution approaches and an interdisciplinary framework. These concepts will be discussed below.

8.2 Molecular epidemiology

The field of molecular epidemiology has recently emerged from the integration of molecular biology into traditional epidemiologic research. The analysis presented in this thesis has demonstrated the value of investigating pathogens beyond the species level. The rapid development of new techniques now allows scientists to study small genetic differences between microorganisms at a level that would be impossible using conventional approaches. This has led to several advances in veterinary epidemiology such as the identification of previously undetected infections, the identification of risk factors for diseases and the application of new diagnostic approaches (Thrusfield 2005). In addition molecular tools are now becoming increasingly available to study transmission patterns within populations at the strain level and to evaluate host- and strain-specific

risk factors. This is changing the way epidemiologists study infectious diseases (Murray 2002).

Foxman and Riley (2001) conclude that the application of molecular techniques to the study of heterogeneous organisms enhances epidemiologic studies by improving our ability to sub-classify these organisms into meaningful groups. However molecular techniques do not substitute for conventional methods. The authors identify a clear need to train individuals who are capable of developing new theories and methods to address the questions that will arise from the interface of molecular biology and epidemiology for infectious diseases. They argue that molecular epidemiology training requires practical application of both the laboratory and epidemiologic techniques to address a real-world infectious disease problem and that molecular epidemiologists will need to interface with clinicians, statisticians, epidemiologists, molecular biologists, computer scientists, engineers, and practitioners in the new field of bioinformatics and computational biology.

Overall the field of molecular epidemiology provides exciting new opportunities for disease control, which could not have been achieved using traditional approaches. This is supported by recent advances in our understanding of the epidemiology of a variety of pathogens (Colles, Jones et al. 2003; Feil, Li et al. 2004; Heir, Lindstedt et al. 2004; Acik and Cetinkaya 2006) and is likely to contribute to an improved prevention and control of infectious diseases, assisted by the on-going advancements of molecular and epidemiological tools (Riley 2004).

8.3 Source attribution modelling

Source attribution has recently received considerable attention in the field of epidemiology, in particular since the development of the Hald model (2004). To facilitate risk management strategies, novel approaches to source attribution try to identify major sources of human infection by combining information on the occurrence of individual pathogen subtypes in human cases and animal food sources. Other approaches include full risk assessments (McBride, Meleason et al. 2005), analysis and extrapolation of surveillance or outbreak data or analytical epidemiological studies (Denno, Keene et al. 2007; Gormley, MacRae et al. 2008), as well as methods based on population genetics (Wilson, Gabriel et al. 2008; Strachan, Gormley et al. 2009). Overall, attribution tools now make a vital contribution to the prioritisation of hazards and the design of interventions in food systems (Doyle and Erickson 2006). The value of this approach is undeniable as it facilitates the ability to set informed priorities and the identification of the importance of individual disease reservoirs by decision makers. A variety of novel techniques are now becoming available, many using the tools of molecular epidemiology (McCarthy, Gillespie et al. 2007; Wilson, Gabriel et al. 2008; Strachan, Gormley et al. 2009).

Source attribution is not new to epidemiology. Concepts such as population attributable risk have been applied successfully and have provided valuable information for disease management (Stafford, Schluter et al. 2008). However when considering a component cause model, these methods often fail to quantify the relative contribution of different sources (Dohoo, Martin et al. 2003) and although applied to campylobacteriosis (Stafford, Schluter et al. 2008) the results from these methods have been insufficient to provide regulators with the insight needed to implement disease control strategies. Epidemiologists have clearly sharpened their tools recently to quantify the relative importance of different disease reservoirs, by including concepts such as molecular techniques (Murray 2002) and through the increasing availability of statistical techniques such as Bayesian inference (Spiegelhalter, Myles et al. 1999; Van Boekel, Stein et al. 2003). This has lead to new insight where traditional methods have failed to untangle the complex epidemiology of disease.

However source attribution approaches can merely be a first step in the control of zoonotic pathogens. These approaches are focused on the major food-animal reservoirs and therefore cannot identify responsible foods at the point of consumption or at other points along the farm-to-fork continuum (Batz, Doyle et al. 2005). Therefore when using these methods it cannot be ascertained if, for example, the contribution of bovine sources to the human disease burden is a result of occupational exposures, environmental faecal contamination, or consumption of contaminated foods (Mullner, Jones et al. 2009). In order to implement successful control strategies, individual pathways of transmission need to be well understood and detailed risk assessments or studies of transmission pathways will need to be conducted to achieve this goal.

8.4 Interdisciplinarity

The topic of 'One world, one medicine' has gained a huge amount of popularity recently and the call has gone out for a better coordination between the animal and public health sectors (FAO 2008). This is further enhanced by an increasing awareness that infectious diseases are dominated by zoonoses and have increased significantly over time (Jones, Patel et al. 2008), thereby putting the spotlight on the medical and veterinary professions for closer cooperation. However, while this appears to be a very rational approach to disease control and prevention, as far as history is concerned this is a challenging task. Science used to be generally orientated in its beginnings. From Plato (428 BC - 348 BC) to Leonardo Da Vinci (1452 - 1519) and up until Goethe (1749 - 1832) scientists were polymaths, whose knowledge was not restricted to one subject area. This was followed by the development of specialisation, and since then science has been divided by disciplines, professionalised and institutionalised. One of the biggest challenges of the 21st century is going to be to break down traditional hierarchical thinking and take a cross-functional approach. Existing interfaces need

to be re-evaluated and with intelligent interaction of different disciplines, coherences and new knowledge will be retrieved. This will increase our understanding of problems and processes from which solutions have been hidden when viewed through the eyes of highly specialised researchers. While this might sound intoxicating on paper, the reality of interdisciplinary research is often quite a different one and involves a steep learning curve (Mullner, Pleydell et al. 2008). In addition the nature of science is changing dramatically (Bell. G., Hey et al. 2009; Science Pod Cast 2009) and this will further intensify the need for working without boundaries. The first scientific paradigm was 'empirical science', which involved simply describing nature. This was followed in time by the analytical paradigm - capturing nature in equations. The solving of complex problems was then tried by simulation in the 'computational' paradigm, as they were too complicated to be handled in equations. 'Data-intensive science' is an emerging field driven by the need for handling data-intensive projects delivering petabytes of data (Bell. G., Hey et al. 2009; Science Pod Cast 2009).

In industry, much important goal-orientated research and development is successfully conducted by interdisciplinary teams. However, interdisciplinary collaboration in academic research was comparatively rare traditionally (Abelson 1994), but this is rapidly changing. According to Van Mulligen et al. (2008), successful education programmes should include not only traditional teaching activities, but also the development of skills necessary for working as part of a interdisciplinary team (Van Mulligen, Cases et al. 2008). Practical research considerations in integrated research show some of the advances and challenges of this concept (Tanaka, Karn et al. 2008). Interdisciplinary research involves scientists of diverse training and experience with a range of expertise in order to address different aspects of a research problem, but this can require substantial financial resources. Team research can encounter some difficulties as individuals must sometimes set aside personal priorities for the benefit of the group. Other compromises might need to be made such as moving out of the 'comfort zone' of a special area of knowledge in order to solve multifaceted problems. Funding and communication can be the greatest deterrents to adopting a team approach and often new and non-standard statistical analysis may be needed due to the complexity of the approach. However, although team research may be more difficult to conduct and interpret, it may well provide information on the interaction of different aspects of a problem rather than just one isolated component of it.

The work that is presented in this thesis was enabled by a team of specialists from a variety of disciplines including human and veterinary medicine, and the environmental and mathematical sciences. I combined and utilised animal, food source and human surveillance data and demonstrated how veterinarians and health protection officers can work together to provide a holistic approach to a public health issue. In addition, the laboratory team contributed expertise on the population structure and dynamics

of *Campylobacter* that, when combined with mathematical modelling, improved our understanding of the epidemiology of this important pathogen. Figure 8.1 illustrates the role of the different disciplines in the project.

In conclusion, setting up a multidisciplinary team can be a challenging task. In my experience the establishment of a collegial level of communication between team members who respect each other's expertise and contributions was essential. This allowed for the development of an understanding of the capabilities, requirements and potential problems of each of the speciality areas. In addition, factors such as a shared vision for the collaborative venture, leadership, sufficient resources as well as good planning with a regular review mechanism are of primary importance. Key factors to successful collaborative research are illustrated and compiled in the Wilder Collaborations Factors Inventory[1]. Despite the steep learning curve, being part of an interdisciplinary research team was an extremely rewarding exercise. By taking this approach exciting research could be conducted, which would not have been achieved without the collaboration of a team of specialists from different fields.

[1] http://www.wilder.org/pubs

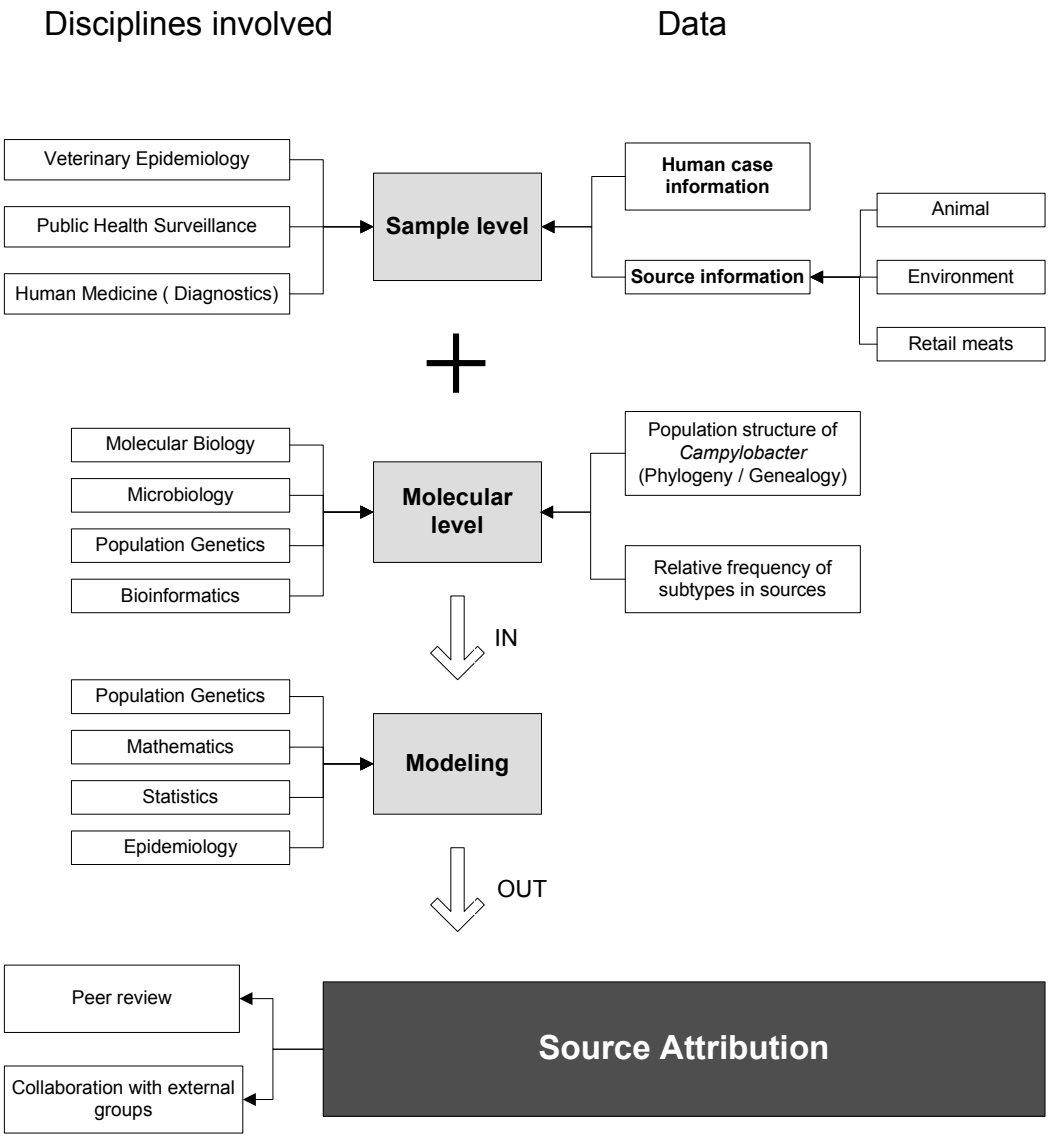

Figure 8.1: Framework of multidisciplinary approach for attributing cases of campylobacteriosis to a source.

8.5 Conclusions

The molecular epidemiological framework developed for this study has provided new knowledge for the control of the most important zoonoses in New Zealand. It can be extended to other pathogens, and can be modified to enable continuous, dynamic assessment of source attribution and an assessment of the effectiveness of control programs, thereby providing exciting new opportunities for disease control. Novel modelling and source attribution approaches have been developed and applied in this thesis and assisted in furthering our knowledge of the (molecular) epidemiology of campylobacteriosis and salmonellosis. The availability of longitudinal data was instrumental to this approach and this underlines the importance of data quality in epidemiological research. Interdisciplinary in nature, the project involved researchers from a variety of disciplines, working on different aspects of the project. This interdisciplinarity has been central to finding a solution for the complex problems that were faced at the beginning of this study. This emphasizes the importance of taking a cross-disciplinary approach and the value of training without boundaries, when working at the human and animal health interface.

The findings from this thesis have provided the scientific basis for the development of disease control strategies within NZFSA's *Campylobacter* in Poultry Risk Management Strategy. As a likely consequence of industry specific interventions, a more than 50% drop of human notifications in 2008 to a 16 year low was observed.

Appendix A

A.1 Summary

This material provides further background information for the review provided in Chapter 1 and is based on the paper by Strulens et al. (Struelens and Members of the European Study Group on Epidemiological Markers (ESGEM) of the European Society for Clinical Microbiology and Infectious Diseases (ESCMID) 1996)

A.2 Selected performance criteria for evaluation of typing systems

A.2.1 Typeability

The typeability is the proportion of strains that are assigned a type by the typing system. The formula reads:

$$T = \frac{N_t}{N}$$

Where: N_t is the number of isolates assigned a type and N the number of isolates tested. For a marker to be useful, T should be as close to 1 as possible.

A.2.2 Reproducibility

The reproducibility is the ability of a typing system to assign the same type to a strain tested on independent separate assays. The formula reads:

$$R = \frac{N_r}{N}$$

Where N_r is the number of isolates assigned the same type on repeat testing and N the number of isolates tested.

A.2.3 Stability

The stability of epidemiological markers conditions the ability of a typing system to recognise the clonal relatedness of strains derived in vitro or in vivo from a common ancestor strain, despite the phenotypic or genomic variation that may occur during laboratory storage and replication, or during clonal dissemination in nature, especially over prolonged periods or large-scale epidemics. Mutations and intra- and intergenomic recombination related to integration or mobilisation of plasmid, phage and transposable DNA occur at frequencies depending on species, strain and environmental conditions and the stability of markers needs to be evaluated for every species and ecosystem under study. In vitro stability is assessed by comparing strains tested before and after storage and passage. The formula reads:

$$S = \frac{N_s}{N}$$

Where N_s is the number of tests in which the same strains were correctly assigned the same type on repeat testing and N is the total number of tests.

The in vivo stability can be estimated by comparing strains tested e.g. before and after passage in a suitable animal method or along the course of epidemiologically well documented outbreaks. Clonal variants that may be recognised to arise in such settings should be taken into consideration for adjusting the working definition (level of similarity) of a clone as applied to future studies of the organism over similar time-space frames. Such variation must be distinguished from the fortuitous occurrence of different strains in these situations in vivo.

A.2.4 Discriminatory power

The discriminatory power is the average probability that the typing system will assign a different type to two unrelated strains randomly sampled in the microbial population of a given taxon. This probability can in general be expressed by the formula of the Simpson index of diversity, which reads:

$$D = 1 - \frac{1}{N(N-1)} \sum j = 1^N n_j(n_j - 1)$$

Where D is the index of discriminatory power, N the number of unrelated strains tested and nj the number of strains belonging to the jth type, assuming that strains will be classified into mutual exclusive categories.

Appendix B

B.1 Summary

This material about the epidemiology of campylobacteriosis cases in the Manawatu appears as an appendix to Chapter 3. During the process of preparing the manuscript for publication, this material was removed to avoid pulling the manuscript in too many different directions. However the results are discussed in the manuscript. For completeness, I include this extra material since it further illustrates the findings from the work presented in Chapter 3.

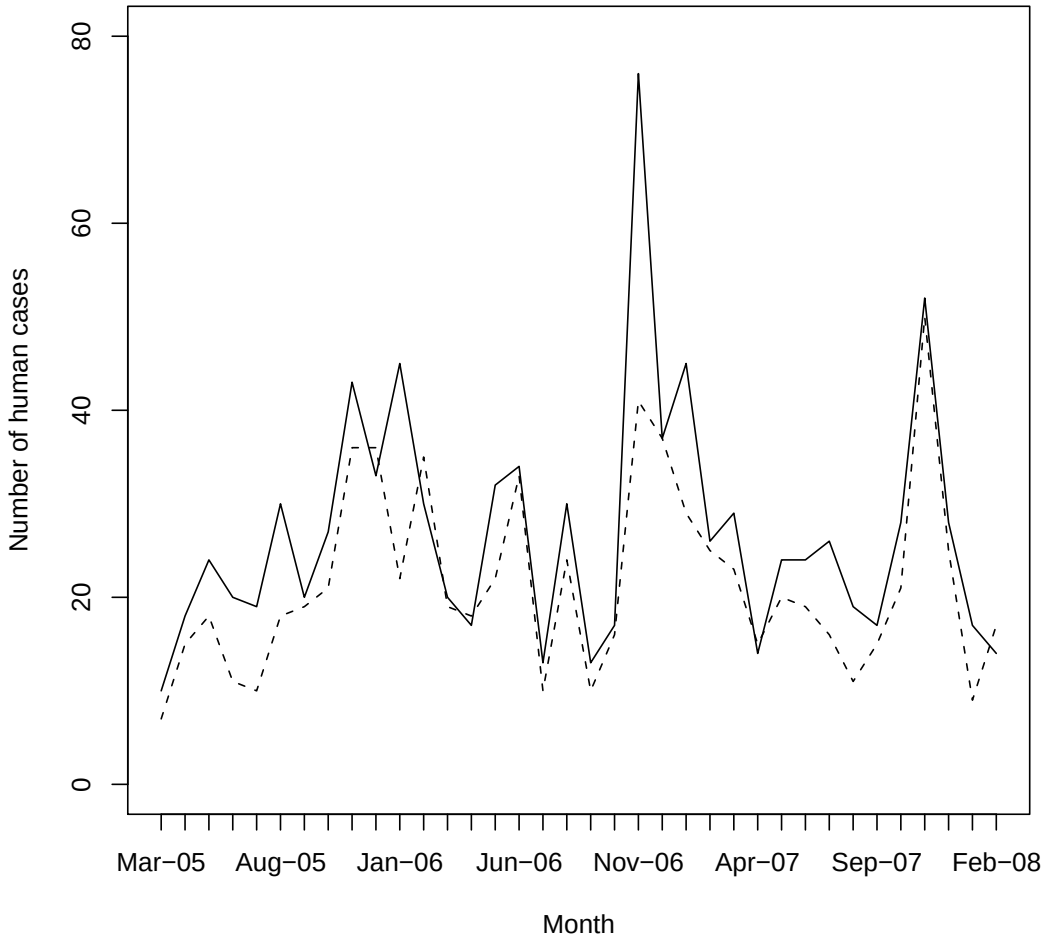

Figure B.1: The number of notified cases (solid line) and samples recruited via the diagnostic laboratory (dashed line) from the Manawatu surveillance site in each month between March 2005 and February 2008.

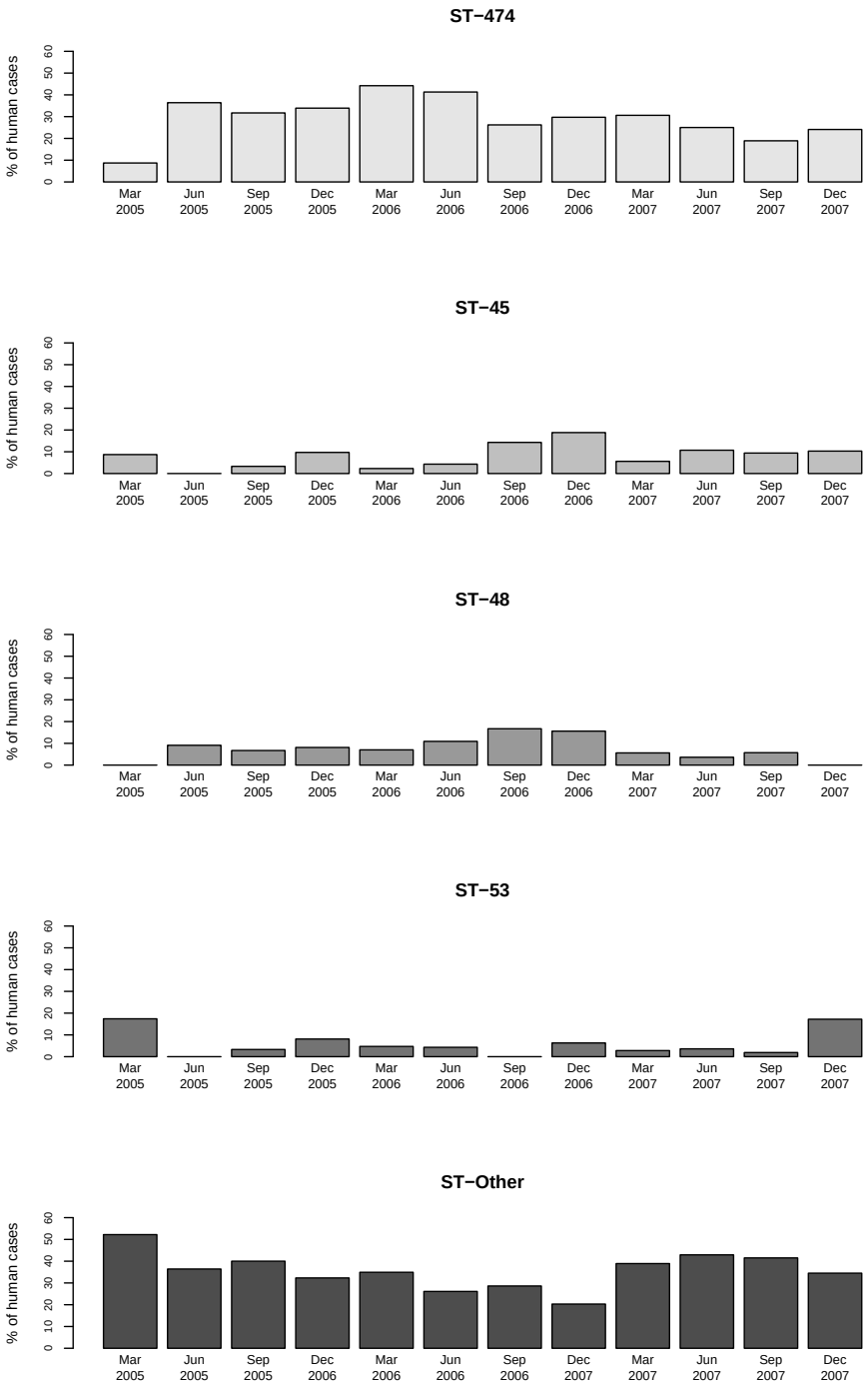

Figure B.2: Relative frequency of human *C. jejuni* MLST genotypes that occurred in more than 5 % of clinical cases from the Manawatu surveillance site in each month between March 2005 and February 2008. The frequency of the remaining ("Other") STs is added for comparison.

Appendix C

C.1 Summary

This material provides further detail about the study described in Chapter 4. During the process of preparing the manuscript for publication material was abbreviated or not included. For completeness, I included this extra material since it further illustrates the findings from the work presented in Chapter 4.

C.2 Material and methods

C.2.1 Summary

This section describes in more detail the molecular methods applied in the study.

C.2.2 Detailed methods

***Campylobacter* isolates.** Human specimens submitted to MedLab Central, Palmerston North that were positive for *Campylobacter* by ELISA (ProSpecT R, Remel, USA) were sent to the Hopkirk Molecular Epidemiology laboratory over a three year period from 1st March 2005 to 29th February 2008. Faecal swabs were made using Amies Charcoal transport swabs (Copan, Italy). Over the same period 12-18 fresh whole poultry carcasses were sampled each month from retail outlets in Palmerston North, representing the different poultry suppliers in the region with the number of samples collected per supplier reflecting market share. The Manawatu region is predominantly supplied by three companies, the regional plants of Supplier A, and through the nationwide distribution of products from Supplier B, which are dominating the supply with fresh poultry carcasses. In addition a localised company (Supplier C) also distributes fresh whole carcasses product in the region.

Bacterial culture and identification. 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 microaerophilic atmosphere (85% N_2, 10% CO_2, 5% O_2) for two days. A single colony resembling *Campylobacter* species was subcultured to Blood Agar (BA) (Fort Richard,

Auckland) and incubated microaerophilically at 42 °C for two days before DNA preparations were made. In addition in a small exploratory study multiple colonies (2-5 colonies per sample) were subcultured from individual samples. Cultures were frozen at −80 °C in Glycerol Broth (Difco, USA). Chickens were washed and massaged in 200 ml of Buffered Peptone Water (BPW) (Difco, USA). The chicken wash was centrifuged (16,264 RCF (g), 6 °C, 35 mins, Sorvall RC5B) and the resultant pellet resuspended in 5ml of BPW. Approximately 3 ml of the resuspended pellet was added to 90 ml of Bolton Broth, which was incubated at 42 °C microaerophilically for two days. After incubation, the broth was subcultured onto mCCDA and incubated microaerophilically at 42 °C for two days. Single colonies resembling *Campylobacter* species were subcultured to BA and incubated microaerophilically at 42 °C for two days before DNA preparations were made. Cultures were frozen at −80 °C.

Species confirmation was done by PCR. DNA was extracted from freshly grown cultures by boiling for ten minutes 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 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* (Stucki et al, 1995), 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)(Denis et al., 2001). 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 one 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 $MgCl_2$. 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 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 603 bp product indicated *C. jejuni* while a 462 bp one indicated *C. coli*.

Sequence typing. After speciation, MLST of *C. jejuni* isolates was performed using seven house-keeping genes: *aspA* (aspartase A), *glnA* (glutamine synthase), emphgltA (citrate synthase), emphglyA (serine hydroxymethyltransferase), emphpgm (phosphoglucomutase), emphtkt (transketolase) and emphuncA (ATP synthase alpha subunit) based on the method as outlined by Dingle et al., 2001. Each amplification reaction comprised 2 μl of the DNA preparation, 5 pmoles of both forward and reverse

amplification primers, 12.5 μl of ABI 2x AmpliTaq Gold PCR Mastermix and water to make up to a total volume of 25 μl. 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 25 μl 20% PEGS / 2.5 M NaCl solution, washed with 80% EtOH, dried and taken up in 12 μl H2O and screened on agarose gels. Sequencing reactions were performed using 2 μl of PCR product, 3.2 pmoles primer, 2 μl ABI BigDye, 2 μl of x 5 BigDye buffer and water to a total volume of 10 μl. Reactions were performed under the following conditions: Initial denaturation at 96C for 3 mins, then 25 cycles of 96 °C for 15 sec, 50 °C for 15 sec, and 60 °C for 4 mins. Sequenced products were precipitated with 0.1 M Na acetate / 78% EtOH solution, washed with 70% EtOH, dried and taken up in 12 μl H2O 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. P. 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)using the same protocol as above.

C.3 Genotypes detected in human cases and poultry samples

C.3.1 Summary

This section provides more information on the genotypes isolated from human cases and poultry samples. Table C.1 shows the clonal complex assignment and allelic profile of all sequence types isolated from human cases and poultry suppliers in the Manawatu region of New Zealand.

Table C.1: Clonal complex assignment and allelic profile of sequence types isolated from human cases and different poultry suppliers in the Manawatu region.

		Allelic profile						
ST	CC	*aspA*	*glnA*	*glt*	*gly*	*pgm*	*tkt*	*uncA*
21	21	2	1	1	3	2	1	5
25	45	4	7	10	1	1	7	1
38	48	2	4	2	2	6	1	5
42	42	1	2	3	4	5	9	3
45	45	4	7	10	4	1	7	1
48	48	2	4	1	2	7	1	5
50	21	2	1	12	3	2	1	5
52	52	9	25	2	10	22	3	6
53	21	2	1	21	3	2	1	5
61	61	1	4	2	2	6	3	17
81	61	2	4	2	2	6	3	17
137	45	4	7	10	4	42	7	1
190	21	2	1	5	3	2	3	5
219	61	1	4	2	2	6	3	1
227	206	2	4	5	2	2	1	5
257	257	9	2	4	62	4	5	6
354	354	8	10	2	2	11	12	6
403	403	10	27	16	19	10	5	7
422	21	2	1	5	3	2	5	5
436	-	7	21	5	62	4	61	44
451	21	2	1	2	3	2	3	5
459	42	1	2	3	3	5	9	3
474	48	2	4	1	2	2	1	5
520	21	2	1	12	88	2	1	5
578	61	1	4	2	2	2	3	17
583	45	4	7	10	4	42	51	1
658	658	2	4	2	4	19	3	6

Continued on next page

Table C.1 – *continued from previous page*

		Allelic profile						
ST	CC	*aspA*	*glnA*	*glt*	*gly*	*pgm*	*tkt*	*uncA*
677	677	10	81	50	99	120	76	52
694	1034	2	59	4	105	126	25	23
829	828	33	39	30	82	113	43	17
1457	-	2	165	73	147	220	190	104
1517	354	8	10	149	2	11	12	6
1581	-	129	66	30	82	189	47	17
1707	607	9	2	5	2	11	3	1
1818	45	4	7	10	4	2	7	1
1911	-	7	84	5	10	119	178	26
2026	403	10	1	16	19	10	5	7
2219	45	10	7	10	4	1	7	1
2343	48	2	4	5	2	10	1	5
2345	206	2	4	5	2	2	5	5
2350	48	2	4	1	2	2	5	5
2391	1034	2	15	4	48	360	25	23
2397	828	184	39	30	82	113	43	17
2535	-	10	2	107	62	120	76	1
3072	828	33	39	30	82	104	173	68
3222	-	33	283	44	82	189	44	17
3230	828	33	39	30	322	104	85	17
3538	-	47	2	4	2	6	5	17
3609	48	2	4	12	2	7	1	5
3676	42	1	307	3	4	5	9	3
3711	257	1	2	4	62	4	5	17
3712	362	1	2	215	4	90	24	8
3715	21	2	1	12	3	357	1	5
3717	21	2	1	21	3	62	1	5
3718	48	2	4	1	4	1	1	5
3719	48	2	4	12	2	2	1	5

Continued on next page

Table C.1 – *continued from previous page*

		Allelic profile						
ST	CC	*aspA*	*glnA*	*glt*	*gly*	*pgm*	*tkt*	*uncA*
3720	49	3	1	5	10	11	11	6
3721	354	8	10	149	2	1	12	6
3725	692	37	1	57	26	107	29	35
3726	-	37	253	4	48	126	25	3
3727	45	48	7	10	4	183	7	1
3784	354	8	315	2	2	11	12	6
3792	257	9	316	4	62	4	5	6

Appendix D

D.1 Summary

This material provides further detail about the salmonellosis attribution model and data described in Chapter 5 and 6.

D.2 Bayesian statistics and simulation modelling

Bayesian methods differ from 'classic' frequentist and likelihood approaches in several aspects (Carlin and Louis 1996). One of the most important differences to traditional techniques is that these methods combine information available beforehand, captured in the prior distribution, and the likelihood function (Van Dongen 2006). It is a major advantage of this approach to allows for direct probability statements about the parameters (Bolstad 2004), while it also facilitates a common-sense interpretation for statistical conclusions (Gelman, Carlin et al. 1995). According to Gelman et al. (1995), the process of Bayesian data analysis can be divided into three steps: Setting up a full probability model, calculating and interpreting the appropriate posterior distribution and evaluating the fit of the model and the implications of the resulting posterior distribution. Bayesian inference is the process of fitting a probability model to a set of data summarizing the result by a probability distribution on the parameters of the model and on unobserved quantities such as predictions for new observations (Gelman, Carlin et al. 1995). When using Bayesian procedures the estimator is considered a random variable having a probability (sampling) distribution (Bolstad 2004). This sampling distribution of the estimator is used to look at how the estimator is distributed around the parameter value, a technique called sample space averaging. By drawing a large numbers of random samples, Monte Carlo simulations are a useful way to perform sample space averaging. The empirical distribution of the statistics (over the large number of random samples) approximates its sampling distribution (over all possible random samples) (Bolstad 2004). For example a mean calculated on this Monte Carlo sample approximates the mean of the sampling distribution. The use of Bayesian methods was hampered by the need for conjugate distributions where the prior together with the observed data provide a posterior distribution from which estimates can be ob-

tained analytically (Toft, Innocent et al. 2007). To overcome these problems Monte Carlo Markov Chain (MCMC) methods have been developed (Toft, Innocent et al. 2007) which draw random samples from the posterior distribution, without completely evaluating it. Taking a large enough sample from the posterior distribution then allows to approximate it to any accuracy (Bolstad 2004). Bayesian statistics have now become widely applicable thanks to cheap and massive computational power as well as reliable and fairly flexible software (WinBUGS) (Mira 2005). Modern Bayesian computation techniques (principally MCMC methods) have now very few constraints on the nature of models that can be considered, and therefore modellers are free to specify models as realistically as possible without worrying about their computational feasibility (Gustafson, Hossain et al. 2005). Overall the Bayesian paradigm provides a conceptually simple method for coping with complex problems like multiple parameters (Gelman, Carlin et al. 1995).

The Monte Carlo Method is essentially a numerical integration which uses a random element to obtain a measure of interest. A simple example is to consider a (random) variable X with a density function $g(x)$. The mean $E(x)$ is then defined as the integral (Toft, Innocent et al. 2007):

$$E(x) = \int_{\infty}^{-\infty} xg(x)dx \tag{D.1}$$

The Monte Carlo estimate of the mean $E_{MC}(X)$ is obtained by sampling a sequence of random variable $X1$, $X2$, $X3$,Xn all having the density function $g(x)$ and then calculate the arithmetic mean of these i.e.,

$$E_{MC}(x) = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{D.2}$$

However in Bayesian analysis the target distribution is typically multidimensional and rather complicated and therefore samples cannot be drawn directly from this distribution (Toft, Innocent et al. 2007). To overcome this problem MCMC methods provide a unifying framework within which many complex problems can be analysed using generic software (Gilks, Richardson et al. 1996). MCMC are a powerful tool to address the computational challenges posed by the Bayesian paradigm and has become the approximation method of choice for a wide range of Bayesian problems thanks to its flexibility for high-dimensional posteriors (Mira 2005). A Markov chain is a sequence of random variables that can be thought of as evolving over time, with the probability of transition depending on the particular set in which the chain is in (Robert and Casella 2004). The basic idea of MCMC is to simulate a Markov chain which has the desired

posterior distribution as its stationary distribution (Gustafson, Hossain et al. 2005). The essence is that we are sampling from a series of random variables that eventually converges to a Markov Chain where the stationary distribution samples form the target distribution (Toft, Innocent et al. 2007). MCMC methods are methods of iterative nature for generating a sample from which expectations of functions of X can reliably be estimated, which can easily be customised to very diverse and difficult problems (Givens and Hoeting 2005). The most widely used Markov chain simulation methods are the generalized Metropolis algorithm and the Gibbs sampler (Gelman, Carlin et al. 1995). While Bayesian statistics have many advantages (Gelman, Carlin et al. 1995; Carlin and Louis 1996), there are two major difficulties, namely the convergence of the MCMC simulation (Brooks and Gelman 1998; Brooks and Roberts 1998) and the prior specification (Van Dongen 2006). To reduce the possibility of inferential bias caused by the effect of starting values, iterates within a initial transient phase of the process (burn-in period) are discarded (Brooks 1998). Unfortunately there is no certain way of establishing whether a chain had converged to the distribution we wish to sample from (Toft, Innocent et al. 2007) and the assessment of convergence has been discussed intensively within the scientific community (Brooks 1998; Brooks and Gelman 1998; Brooks and Roberts 1998). Time series plots, Gelman Rubin convergence diagnostics, autocorrelation plots as well as Monte Carlo standard errors are ways of assessing the performance of MCMC chains in WinBUGS (Toft, Innocent et al. 2007). In any practical situation it is necessary to determine how long the simulations need to run, in particular as computer time and storage limits may also be of consideration (Brooks 1998). The choice between running one longer chain or several shorter chains has been discussed by Brooks (Brooks 1998). The author argues that when using a single chain it is closer to the target distribution whereas on the other multiple chains can provide the opportunity to guard against a single chain leaving a "significant proportion" of the sampling space unexplored as well as explore to what extends the chains have mixed. In conclusion multiple chains protect against bias by ensuring that the entire sample space is covered, whereas one longer run provides less variable estimates, which are based on a larger sample. Bayesian analysis is driven by the prior distribution and its source and use present many challenges, for example the turning of informally expressed opinions into a mathematical distribution (Spiegelhalter, Abrams et al. 2004). Van Dongen (Van Dongen 2006) gives examples on why the specification of the prior distribution should be done with great care, which is supported by Spiegelhalter et al (2004) when using default priors intended to be minimally informative. Generally prior distributions are not necessarily pre-specified, unique, known or important. The use of priors is based on judgement and hence a degree of subjectivity cannot be avoided and a range of options should be examined in a sensitivity analysis (Spiegelhalter, Abrams et al. 2004). In practice therefore when fitting the model using a Markov Chain Monte

Carlo model problems might occur when (1) the priors specify too strongly that prevalence is very small, when in fact it is not and (2) the model is highly non-identifiable from a frequentist perspective. In the later case the posterior distribution for many parameters may be multimodal or very dispersed.

D.3 Assignment of travel and outbreak association of cases using the EpiSurv questionnaire

D.3.1 Outbreak association

Outbreak association was determined using the outcome of the question labelled 'Outbreak details': *'Is this case part of an outbreak (i.e., known to be linked to one or more other cases of the same disease)?'*. 'Yes' or 'No' were the possible answers to this question. Cases that answered yes to this question were assigned to the category outbreak associated, cases that answered 'No' were categorized as not outbreak associated.

D.3.2 Travel association

Travel association was determined using the outcome of the question labelled 'Overseas travel': *"Was the case overseas during the incubation period for this disease?'* This questions could be answered with 'Yes', 'No' or 'unknown'. Cases that answered 'Yes' were considered to be travel associated. Cases that answered 'No' were considered to be not travel associated. If the question was answered with 'unknown' the status was categorized as not available.

D.4 'Other *Salmonella types*' detected in human cases from 2002-2004

Table D.1: 'Other *Salmonella types*' detected in human cases from 2002-2004.

Salmonella type	2002	2003	2004	2000-2004
Salmonella Aberdeen	0	1	3	4
Salmonella Agona	7	3	5	15
Salmonella Alachua	1	0	0	1
Salmonella Albany	2	1	0	3
Salmonella Altona	0	0	1	1
Salmonella Anatum	4	4	3	11
Salmonella Augustenborg	0	1	0	1
Salmonella Baildon	0	1	0	1
Salmonella Bareilly	5	1	5	11
Salmonella Birkenhead	1	2	1	4
Salmonella Blockley	0	2	0	2
Salmonella Bovismorbificans	5	1	6	12
Salmonella Braenderup	5	1	4	10
Salmonella Butantan	0	0	1	1
Salmonella Choleraesuis var Kunzendorf	1	0	0	1
Salmonella Corvallis	0	5	8	13
Salmonella Derby	3	2	1	6
Salmonella Djugu	1	0	0	1
Salmonella Ealing	1	0	0	1
Salmonella Eastbourne	0	0	1	1
Salmonella Emek	1	0	0	1
Salmonella Give 15+	0	1	0	1
Salmonella Group 1	0	1	0	1
Salmonella Group B	0	3	1	4
Salmonella Group C	2	1	1	4
Salmonella Group D	0	1	0	1

Continued on next page

Table D.1 – *continued from previous page*

Salmonella type	2002	2003	2004	2000-2004
Salmonella Group E 3,10 : r :-	1	0	0	1
Salmonella group Q	1	0	0	1
Salmonella Haardt	0	0	1	1
Salmonella Hadar	14	5	2	21
Salmonella Haifa	0	1	0	1
Salmonella Havana	0	1	1	2
Salmonella Hindmarsh	3	3	3	9
Salmonella Hvittingfoss	7	0	1	8
Salmonella Irumu	0	1	0	1
Salmonella Javiana	8	4	3	15
Salmonella Kentucky	2	2	0	4
Salmonella Kottbus	0	1	0	1
Salmonella Krefeld	0	1	0	1
Salmonella Lansing	0	1	1	2
Salmonella Lexington 15+	0	2	1	3
Salmonella Liverpool	0	0	1	1
Salmonella Livingstone	1	0	0	1
Salmonella London	1	1	1	3
Salmonella Mbandaka	6	6	8	20
Salmonella Mgulani	0	0	1	1
Salmonella Mississippi	9	14	9	32
Salmonella Muenchen	0	1	0	1
Salmonella Newport	12	4	10	26
Salmonella Ohio	0	2	2	4
Salmonella Oranienburg	5	5	4	14
Salmonella Orion 15+	2	1	0	3
Salmonella Oslo	2	2	1	5
Salmonella Othmarschen	0	1	0	1
Salmonella Panama	3	3	4	10
Salmonella Papuana	1	1	0	2

Continued on next page

Table D.1 – *continued from previous page*

Salmonella type	2002	2003	2004	2000-2004
Salmonella Pensacola	3	2	5	10
Salmonella Poona	2	1	1	4
Salmonella Potsdam	0	1	0	1
Salmonella Reading	5	1	3	9
Salmonella Riffen	4	0	0	4
Salmonella Rissen	0	0	3	3
Salmonella Rubislaw	1	0	0	1
Salmonella Ruiru	0	0	1	1
Salmonella Sandiego	0	2	2	4
Salmonella Schwarzengrund	1	3	2	6
Salmonella Senftenberg	2	2	4	8
Salmonella Singapore	0	1	0	1
Salmonella Species	1	0	0	1
Salmonella species 1, 4,12 : eh : -	1	0	0	1
Salmonella species 3	1	0	0	1
Salmonella species 3,10 : r : -	4	8	9	21
Salmonella species 3,15 : r : -	0	2	0	2
Salmonella species 4 12	2	0	0	2
Salmonella species 4,12 : - : 1,2	2	6	3	11
Salmonella species 4,5,12 : d : -	17	17	15	49
Salmonella species 6,12 :-:- 1 2	0	0	1	1
Salmonella species 6,7 : - : - (non-motile)	3	0	0	3
Salmonella species 6,7 : k : -	27	5	1	33
Salmonella species 6,8 : eh : -	0	0	2	2
Salmonella species 6,8,20 : e,h : -	1	0	1	2
Salmonella species 9	0	0	1	1
Salmonella species 9,12 : - : 1,5	0	0	1	1
Salmonella species Rough : r : 1,2	1	1	0	2
Salmonella Stanley	1	1	2	4
Salmonella subspecies I Rough : - : - (non motile)	1	1	0	2

Continued on next page

Table D.1 – *continued from previous page*

Salmonella type	2002	2003	2004	2000-2004
Salmonella subspecies II 21 : z10 : z6	0	1	0	1
Salmonella Tennessee	1	1	0	2
Salmonella Uganda	0	2	1	3
Salmonella Victoria	1	1	0	2
Salmonella Virchow	15	12	25	52
Salmonella Wandsworth	0	0	1	1
Salmonella Waycross	1	0	0	1
Salmonella Weltevreden	14	13	14	41
Salmonella Zanzibar	0	1	0	1
Total	229	180	193	602

Appendix E

E.1 Summary

This material provides further detail about the dataset used for the campylobacteriosis attribution models described in Chapter 7.

Table E.1: Relative frequency of sequence types (ST) in % and total number of isolates from human cases and disease sources in the Manawatu dataset.

	Source Type						
ST	Human	Supplier A	Supplier B	Supplier C	Bovine	Ovine	Environment
5	0	0	0	0	0	0	0
21	1.4	0	0.8	0	4	0.7	0
25	0.2	1.5	0	2.5	0	0	1.1
38	2.6	0	0	0	1	0	0
42	3.8	3.1	0	0	11.1	12.9	1.1
45	8.2	27.5	12.6	28.4	2	1.4	17.8
48	8.4	0.8	22.7	8.6	0	0	0
50	4.6	3.8	17.6	16	12.1	15.7	1.1
51	0	0	0	0	0	0	0
52	3.4	4.6	0	0	0	0	0
53	5.4	7.6	3.4	3.7	24.2	0.7	0
61	2.8	0	0	0	8.1	10	1.1
81	0.2	0	0	0	0	0	0
137	0.2	0	0	0	0	0	2.2
177	0	0	0	0	0	0	2.2
190	4.2	6.1	0	0	7.1	5.7	0

Continued on next page

Table E.1 – *continued from previous page*

	Source Type						
ST	Human	Supplier A	Supplier B	Supplier C	Bovine	Ovine	Environment
219	0.2	0	0	0	0	0	0
227	0	0.8	0	0	0	0	0
257	2.4	7.6	3.4	0	0	0	0
354	4.6	2.3	0	1.2	0	0	0
393	0	0	0	0	0	0.7	0
403	0.2	0	0	0	0	0	0
422	0.6	0	0	0	2	13.6	3.3
436	0.8	0	0	0	2	2.1	2.2
451	1.6	0	0	24.7	0	0	0
459	0.2	0	0	0	0	0	0
474	30.7	20.6	1.7	0	5.1	1.4	2.2
520	1.6	3.8	0	2.5	6.1	2.1	0
526	0	0	0	0	0	0	1.1
578	0.2	0	0	0	0	0	0
583	1.8	6.1	0	4.9	0	0.7	1.1
618	0	0	0	0	0	0.7	0
658	0.4	0	0	0	0	0	0
677	1	0	0.8	0	0	0.7	1.1
694	0	0	0	0	0	0	1.1
829	0.2	0	0	0	0	0	0
1030	0	0	0	0	0	0	1.1
1115	0	0	0	0	1	0	0
1191	0	0	0	0	0	0.7	0
1223	0	0	0	0	0	0	1.1
1225	0	0	0	0	0	0	7.8
1243	0	0	0	0	0	0	1.1
1457	0.2	0	0	0	0	0	0
1517	0.8	0	5	0	0	5	0
1581	0.2	0	4.2	0	0	0	0

Continued on next page

Table E.1 – *continued from previous page*

	Source Type						
ST	Human	Supplier A	Supplier B	Supplier C	Bovine	Ovine	Environment
1707	0.2	0	0	0	0	0	0
1818	0	0	0.8	0	0	0	0
1911	0	0	0.8	0	0	0	0
2026	2.2	0	0	0	5.1	12.9	0
2140	0	0	0	0	0	0	0
2219	0.2	0	0	0	0	0	0
2343	0.2	0	0	0	0	0	0
2345	0.6	0.8	9.2	0	0	0	0
2347	0	0	0	0	0	0	1.1
2350	0.4	0	0	0	1	0	0
2354	0	0	0	0	0	0	1.1
2381	0	0	0	0	0	0	17.8
2391	0.2	0.8	0	0	0	0	0
2392	0	0	0	0	0	0.7	0
2397	0	0	2.5	0	0	0	0
2535	0	0	0	1.2	0	0	0
2584	0	0	0	0	0	0	1.1
2619	0	0	0	0	0	0	1.1
3072	0.2	0	0	0	2	0	0
3222	0.2	0	0	0	0	0	0
3230	0	0	0	1.2	0	0	0
3232	0	0	0	0	0	1.4	0
3301	0	0	0	0	0	0	1.1
3538	0.2	0	0	0	0	0	1.1
3609	0	0	12.6	0	0	0	0
3610	0	0	0	0	0	1.4	1.1
3640	0	0	0	0	0	0	1.1
3655	0	0	0	0	0	0	2.2
3656	0	0	0	0	0	0	1.1

Continued on next page

Table E.1 – *continued from previous page*

	Source Type						
ST	Human	Supplier A	Supplier B	Supplier C	Bovine	Ovine	Environment
3657	0	0	0	0	0	0	1.1
3658	0	0	0	0	0	0	1.1
3659	0	0	0	0	0	0	2.2
3660	0	0	0	0	0	0	1.1
3661	0	0	0	0	0	0	1.1
3662	0	0	0	0	0	0	2.2
3663	0	0	0	0	0	0	2.2
3664	0	0	0	0	0	0	1.1
3672	0	0	0	0	0	0	1.1
3673	0	0	0	0	0	0	1.1
3674	0	0	0	0	0	0	1.1
3675	0	0	0	0	0	0	1.1
3676	0.4	0	0	0	0	0.7	1.1
3711	0.4	0	0	1.2	0	1.4	0
3712	0.4	0	0	0	0	0	0
3713	0	0	0	0	0	0	0
3714	0	0	0	0	1	0	0
3715	0.2	0	0	0	0	0	0
3716	0	0	0	0	1	0	0
3717	0.2	0	0	3.7	0	0	0
3718	0.2	0	0	0	0	0	0
3719	0	0	0.8	0	0	0.7	0
3720	0.2	0	0	0	0	0	0
3721	0	0.8	0	0	0	0	0
3722	0	0	0	0	0	0.7	0
3723	0	0	0	0	1	0	0
3724	0	0	0	0	0	0.7	0
3725	0	0	0.8	0	0	0	0
3726	0	1.5	0	0	0	0	0

Continued on next page

Table E.1 – *continued from previous page*

	Source Type						
ST	Human	Supplier A	Supplier B	Supplier C	Bovine	Ovine	Environment
3727	0.2	0	0	0	0	0	0
3728	0	0	0	0	0	0	0
3784	0.2	0	0	0	0	0	0
3792	0.2	0	0	0	0	0	0
3793	0	0	0	0	0	1.4	0
3794	0	0	0	0	0	0.7	0
3795	0	0	0	0	0	0.7	0
3797	0	0	0	0	0	1.4	0
3798	0	0	0	0	1	0	0
3799	0	0	0	0	2	0	0
3800	0	0	0	0	0	0	1.1
3802	0	0	0	0	0	0	1.1
Isolates	502	131	119	81	99	140	90

Bibliography

Abelson, P. H. (1994). 'Multidisciplinary Research.' *Science* **266**: 951.

Acik, M. N. and Cetinkaya, B. (2006). 'Heterogeneity of *Campylobacter jejuni* and *Campylobacter coli* strains from healthy sheep.' *Veterinary Microbiology* **115**(4): 370-375.

Adak, G. K., Cowden J. M., Nicholas S. and Evan H. S. (1995). 'The Public-Health Laboratory Service National Case-Control Study of Primary Indigenous Sporadic Cases of *Campylobacter* Infection.' *Epidemiology and Infection* **115**(1): 15-22.

Advisory Committee on the Microbiological Safety of Food (ACMSF) (2005). *Second Report on Campylobacter*, Food Standards Agency UK.

Asakura, M., Samosornsuk W., Hinenoya A., Misawa N., Nishimura K., Matsuhisa A. and Yamasaki S. (2008). 'Development of a cytolethal distending toxin (cdt) gene-based species-specific multiplex PCR assay for the detection and identification of *Campylobacter jejuni*, *Campylobacter coli* and *Campylobacter fetus*.' *FEMS Immunology and Medical Microbiology* **52**(2): 260-266.

Australian Government. Department of Health and Ageing (2008). 'Australia's notifiable diseases staus, 2006: Annual report of the National Notifable Diseases Surveillance System.' *Communicable Diseases Intelligence* **32**(2).

Baker, M. G., Sneyd E. and Wilson N. A. (2007). 'Is the major increase in notified campylobacteriosis in New Zealand real?' *Epidemiology and Infection* **135**(1): 163-170.

Baker, M., Wilson N., Ikram R., Chambers S., Shoemack P. and Cook G. (2006). 'Regulation of chicken contamination urgently needed to control New Zealand's serious campylobacteriosis epidemic.' *The New Zealand Medical Journal* **119**(1243).

Barker, G. C., Talbot N. L. C. and Peck M. W.(2002). 'Risk assessment for *Clostridium botulinum*: a network approach.' *International Biodeterioration & Biodegradation* **50**(3-4): 167-175.

Batz, M. B., Doyle M. P., Morris J. G., Painter J., Singh R., Tauxe R. V., Taylor M. R. and Wong D. (2005). 'Attributing illness to food.' *Emerging Infectious Diseases* **11**(7): 993-999.

Bell. G., Hey T. and Szalay A. (2009). 'Beyond the Data Deluge.' *Science* **323**(5919): 1297.

Bemrah, N., Bergis H., Colmin C., Beaufort A., Millemann Y., Dufour B., Benet J. J., Cerf O. and Sanaa M. (1998). 'Quantitative risk assessment of human listeriosis from consumption of soft cheese made from raw milk.' *Preventive Veterinary Medicine* **37**(1-4): 129-145.

Ben Aissa, R., Al-Gallas N., Troudi H., Bethadj N. and Belhadj A. (2007). 'Trends in *Salmonella enterica* serotypes isolated from human, food, animal, and environment in Tunisia, 1994-2004.' *Journal of Infection* **55**(4): 324-339.

Best, E. L., Fox A. J., Owen R. J., Cheesbrough J. and Bolton F. J. (2007). 'Specific detection of *Campylobacter jejuni* from faeces using single nucleotide polymorphisms.' *Epidemiology and Infection* **135**(5): 839-846.

Birk, T., Rosenquist H., Brondsted L., Ingmer H., Bysted A. and Christensen B. B. (2006). 'A comparative study of two food model systems to test the survival of *Campylobacter jejuni* at -18 degrees C.' *Journal of Food Protection* **69**(11): 2635-2639.

Black, R. E., Levine M. M., Clements M. L., Hughes T. P. and Blaser M. J. (1988). 'Experimental *Campylobacter jejuni* infection in humans.' *Journal of Infectious Disease* **157**: 472-479.

Bolstad, W. M. (2004). *Introduction to Bayesian Statistics*, John Wiley & Sons.

Brooks, S. P. (1998). 'Markov chain Monte Carlo method and its application.' *Journal of the Royal Statistical Society Series D-The Statistician* **47**(1): 69-100.

Brooks, S. P. and Gelman, A. (1998). 'General methods for monitoring convergence of iterative simulations.' *Journal of Computational and Graphical Statistics* **7**(4): 434-455.

Brooks, S. P. and Roberts, G. O. (1998). 'Convergence assessment techniques for Markov chain Monte Carlo.' *Statistics and Computing* **8**(4): 319-335.

Cahill, S. M. (2005). 'Risk assessment and campylobacteriosis.' In: *Risk Management Strategies: Monitoring and Surveillance.* F. J. M. Smulders and J. D. Collins, Wageningen Academic Publishers.

Carlin, B. S. and Louis, T. A. (1996). *Bayes and Empirical Bayes Methods for Data Analysis*, Chapmann & Hall.

Carter, P. E. (2007). 'Multilocus Sequence Typing of New Zealand *Campylobacter jejuni* isolates: Comparison with Pulse Field Gel Electrophoreses and Penner Serotyping'. *Institute of Environmental Science & Research Limited (ESR).*

Cassin, M. H., Lammerding A. M., Todd E. C. D., Ross W. and McColl R. S. (1998). 'Quantitative risk assessment for *Escherichia coli* O157:H7 in ground beef hamburgers.' *International Journal of Food Microbiology* **41**(1): 21-44.

Center for Disease Control and Prevention 'FoodNet Report 2006.' (http://www.cdc.gov/foodnet/reports.htm), Accessed on January 25, 2009.

Champion, O. L., Gaunt M. W., Gundogdu O., Elmi A., Witney A. A., Hinds J., Dorrell N. and Wren B. (2005). 'Comparative phylogenomics of the food-borne pathogen *Campylobacter jejuni* reveals genetic markers predictive of infection source.' *Proceedings of the National Academy of Sciences of the United States of America* **102**(44): 16043-16048.

Chan, K. F., Tran H. L., Kanenaka R. Y. and Kathariou S. (2001). 'Survival of clinical and poultry-derived isolates of *Campylobacter jejuni* at a low temperature (4 degrees C).' *Applied and Environmental Microbiology* **67**(9): 4186-4191.

Chrystal, N. D., Hargraves S. J., Boa A. C. and Ironside C. J. (2008). 'Counts of *Campylobacter* spp. and Prevalence of *Salmonella* Associated with New Zealand Broiler Carcasses.' *Journal of Food Protection* **71**(12): 2526-2532.

Clark, C. G., Bryden L., Cuff W. R., Johnson P. L., Jamieson F., Ciebin B. and Wang G. H. (2005). 'Use of the Oxford multilocus sequence typing protocol and sequencing of the flagellin short variable region to characterize isolates from a large outbreak of waterborne *Campylobacter* sp strains in Walkerton, Ontario, Canada.' *Journal of Clinical Microbiology* **43**(5): 2080-2091.

Clough, H. E., Clancy D. and French N. P. (2006). 'Vero-cytotoxigenic *Escherichia coli* O157 in pasteurized milk containers at the point of retail: A qualitative approach to exposure assessment.' *Risk Analysis* **26**(5): 1291-1309.

Clough, H. E., Clancy D. and French N. P. (2009). 'Quantifying exposure to vero-cytotoxigenic *Escherichia coli* O157 in milk sold as pasteurized: a model-based approach.' *International Journal of Food Microbiology* **131**(2-3): 95-105.

Codex Alimentarius Commission (CAC) (1999). 'Principles and Guidelines for the Conduct of Microbiological Risk Assessment.' Joint FAO/WHO food standards programme. *Codex Commitee on Food Hygiene*, 32th session, Rome, Italy, 28 June - 3 July 1999.

Coleman, M. E. and Marks, H. (1998). 'Topics in Dose-Response Modeling.' *Journal of Food Protection* **61**(11): 1150-1159.

Coleman, M. E. and Marks, H. M. (1999). 'Qualitative and quantitative risk assessment.' *Food Control* **10**(4-5): 289-297.

Coleman, M. E., Marks H. M., Golden N. J. and Latimer H. K. (2004). 'Discerning strain effects in microbial dose-response data.' *Journal of Toxicology and Environmental Health-Part A-Current Issues* **67**(8-10): 667-685.

Colles, F. M., Jones K., Harding R. M. and Maiden M. C. (2003). 'Genetic diversity of *Campylobacter jejuni* isolates from farm animals and the farm environment.' *Applied and Environmental Microbiology* **69**(12): 7409-13.

Colles, F. M., Jones T. A., McCarthy N. D., Sheppard S. K., Cody A. J., Dingle K. E., Dawkins M. S. (2008). '*Campylobacter* infection of broiler chickens in a free-range environment.' *Environmental Microbiology* **10**(8): 2042-2050.

Cook, R. (2008). Principal Advisor Microbiology, New Zealand Food Safety Authority: *Personal communication.*

Cornelius, A. J., Nicol C. and Hudson J. A. (2005). '*Campylobacter* spp. in New Zealand raw sheep liver and human campylobacteriosis cases.' *International Journal of Food Microbiology* **99**(1): 99-105.

Covello, V. and Merkhofer, M. (1993). 'Risk assessment methods: approaches for assessing health and environmental risks.' New York: Plenum Press.

Cowden, J. (1992). '*Campylobacter*-Epidemiologic Paradoxes.' *British Medical Journal* **305**(6846): 132-133.

Crump, J. A., Murdoch D. R. and Baker M. G. (2001). 'Emerging infectious diseases in an island ecosystem: The New Zealand perspective.' *Emerging Infectious Diseases* **7**(5): 767-772.

De Boer, P., Duim B., Rigter A., van Der Plas J., Jacobs-Reitsma W. F. and Wagenaar J. A. (2000). 'Computer-assisted analysis and epidemiological value of genotyping methods for *Campylobacter jejuni* and *Campylobacter coli*.' *Journal of Clinical Microbiology* **38**(5): 1940-6.

DeJesus, A. J. and Whiting R. C. (2003). 'Thermal Inactivation, Growth, and Survival Studies of *Listeria monocytogenes* Strains Belonging to Three Distinct Genotypic Lineages.' *Journal of Food Protection* **66**(9): 1611-1617.

Delignette-Muller, M. L., Cornu M., Pouillot R. and Denis J. B. (2006). 'Use of Bayesian modelling in risk assessment: Application to growth of *Listeria monocytogenes* and food flora in cold-smoked salmon.' *International Journal of Food Microbiology* **106**(2): 195-208.

Denis, M. and Refregier Petton, J.(2001). '*Campylobacter* contamination in french chicken production from farm to consumers. Use of a PCR assay for detection of *Campylobacter jejuni* and *C. coli*.' *Journal of Applied Microbiology* **91**(2) :255-267

Denno, D. M., Keene W. E., Hutter C. M., Koepsell J. K., Patnode M., Flodin-Hursh D., Stewart L. K., Duchin J. S., Rasmussen L., Jones R. and Tarr P. I.(2007). 'Tri-County Comprehensive Assessment of Risk Factors for Sporadic Reportable Bacterial Enteric Infection in Children'. *Annual Meeting of the Pediatric-Academic-Societies/Society-of-Pediatric-Research*, Toronto, CANADA.

Didelot, X. and Falush, D. (2007). 'Inference of Bacterial Microevolution using Multilocus Sequence Data.' *Genetics* **175**: 1251-1266.

Dingle, K. E., Colles F. M., Falush D. and Maiden M. C. (2005). 'Sequence typing and comparison of population biology of *Campylobacter coli* and *Campylobacter jejuni*.' *Journal of Clinical Microbiology* **43**(1): 340-7.

Dingle, K. E., Colles F. M., Ure R., Wagenaar J. A., Duim B., Bolton F. J., Fox A. J., Wareing D. R. and Maiden M. C. (2002). 'Molecular characterization of *Campylobacter jejuni* clones: a basis for epidemiologic investigation.' *Emerging Infectious Diseases* **8**(9): 949-55.

Dingle, K. E., Colles F. M., Wareing D. R., Ure R., Fox A. J., Bolton F. E., Bootsma H. J., Willems R. J., Urwin R. and Maiden M. C. (2001). 'Multilocus sequence typing system for *Campylobacter jejuni*.' *Journal of Clinical Microbiology* **39**(1): 14-23.

Dohoo, I. A., Martin, W., et al. (2003). Introduction and causal concepts. *Veterinary Epidemiologic Research*, ACV Inc.

Donnison, A. M. and Ross, C. M. (2004). Microbiological Safety of Meat; Thermotolerant *Campylobacter*. *Encyclopedia of Meat Sciences* W. K. Jensen, Elsevier. Two: 798 - 804.

Dopfer, D., Buist W., Soyer Y., Munoz M. A., Zadoks R. N., Geue L. and Engel B. (2008). 'Assessing genetic heterogeneity within bacterial species isolated from gastrointestinal and environmental samples: How many isolates does it take?' *Applied and Environmental Microbiology* **74**(11): 3490-3496.

Doyle, M. P. and Erickson, M. C. (2006). 'Emerging microbiological food safety issues related to meat.' *Meat Science* **74**(1): 98-112.

Duim, B., Godschalk P. C., van den Braak N., Dingle K. E., Dijkstra J. R., Leyde E., van der Plas J., Colles F. M., Endtz H. P., Wagenaar J. A., Maiden M. C. and van Belkum A. (2003). 'Molecular evidence for dissemination of unique *Campylobacter jejuni* clones in Curacao, Netherlands Antilles.' *Journal of Clinical Microbiology* **41**(12): 5593-5597.

Duim, B., Wassenaar T. M., Rigter A. and Wagenaar J. (1999). 'High-resolution genotyping of *Campylobacter* strains isolated from poultry and humans with amplified fragment length polymorphism fingerprinting.' *Applied and Environmental Microbiology* **65**(6): 2369-75.

Duke, L. A., Breathnach, A. S., et al. (1996). 'A mixed outbreak of *cryptosporidium* and *campylobacter* infection associated with private water supplies in England and Wales 1970 - 2000.' *Epidemiology and Infection* **116**(3): 303-308.

Duncan, G. (2008). Programm Manager Laboratory and Microbiological Programmes, New Zealand Food Safety Authority. *Personal communication.*

Eaker, S., Bergstrom R., Bergstrom A., Adami H. O. and Nyren O. (1998). 'Response rate to mailed epidemiologic questionnaires: A population-based randomized trial of variations in design and mailing routines.' *American Journal of Epidemiology* **147**(1): 74-82.

Eberhart-Phillips, J., Walker N., Garrett N., Bell D., Sinclair D., Rainger W. and Bates M. (1997). 'Campylobacteriosis in New Zealand: results of a case-control study.' *Journal of Epidemiology and Community Health* **51**(6): 686-691.

Enteric Reference Laboratory (1995 - 2004). Isolates typed at ERL. C. Nicols, ESR, Kenepuru. *Personal communication.*

European Food Safety Authority (2006). 'Community Summary Report on Trends and Sources of Zoonoses, Zoonotic Angents, Antimicrobial resistance and Foodborne outbreaks in the European Union in 2006.' (http://www.efsa.europa.eu/efsa-locale-1178620753812_1178671312912.htm), Accessed on January 25, 2009.

European Food Safety Authority (2008). 'A quantitative microbiological risk assessment on Salmonella in meat: Source attribution for human salmonellosis from meat.' *EFSA Panel on Biological Hazards.*

Evers, E. G., Van Der Fels-Klerx H. J., Nauta M., Schijven J. F. and Havelaar A. H. (2008). 'Campylobacter source attribution by exposure assessment.' *International Journal of Risk Assessment and Management* **8** (1-2): 174 - 190.

Excoffier, L., Laval G. and Schneider S. (2005). 'Arlequin ver. 3.0: An integrated software package for population genetics data analysis.' *Evolutionary Bioinformatics Online* **1**: 47 - 50.

Excoffier, L., Smouse P. E. and Quattro J. M. (1992). 'Analysis of molcular variance inferred from metric distances among DNA haplotypes - application to human mitochondrial-DNA restriction data ' *Genetics* **131**(2): 479-491.

Fahey, T. and Morgan, D. (1995). 'An Outbreak of *Campylobacter-Jejuni* Enteritis Associated with Failed Milk Pasteurization.' *Journal of Infection* **31**(2): 137-143.

Falush, D., Stephens M. and Pritchard J. (2003). 'Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies.' *Genetics* **164**: 1567 - 1587.

Falush, D., Stephens M. and Pritchard J. (2007). 'Inference of population structure using multilocus genotype data: dominant markers and null alleles.' *Molecular Ecology Notes* **7**(4): 574-578.

FAO/WHO (2002). 'Risk assessments of *Salmonella* in eggs and broiler chickens.' *Microbiological Risk Assessment Series*.

FAO/WHO (2006). 'A Risk Assessment of Campylobacter spp. in Broiler Chickens - Full Report.'

FAO/WHO Global Forum of Food Safety Regulators - GF CRD Denmark-1 (2002). 'The Campylobacter Risk Analysis Initiative.' Marrakech, Morocco.

FAO (2008). Situation Update 56. *FAOAIDEnews*.

Fearnhead, P., Smith N., Barrigas M., Fox A. and French N. P. (2005). 'Analysis of Recombination in *Campylobacter jejuni* from MLST Population data.' *Journal of Molecular Evolution* **60**: 1-9.

Federal Institute for Risk Assessment BFR (2005). 'Campylobacteriose durch Haehnchenfleisch Eine quantitative Risikoschaetzung.' *Bundesinstutut fuer Risikobewertung (BFR)*, Berlin, Germany.

Feil, E. J., Li B. C., Aanensen D. M., Hanage W. P. and Spratt B. G. (2004). 'eBURST: Inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data.' *Journal of Bacteriology* **186**(5): 1518-1530.

Feinsinger, P., Spears E. E. and Poole R. W. (1981). 'A Simple Measure of Niche Breadth.' *Ecology* **62**(1): 27-32.

Fitzgerald, C., Helsel L. O., Nicholson M. A., Olsen S. J., Swerdlow D. L., Flahart R., Sexton J. and Fields P. I. (2001). 'Evaluation of methods for subtyping *Campylobacter jejuni* during an outbreak involving a food handler.' *Journal of Clinical Microbiology* **39**(7): 2386-90.

Fitzgerald, C. and Stanley, K. (2001). 'Use of pulsed-field gel electrophoresis and flagellin gene typing in identifying clonal groups of *Campylobacter jejuni* and *Campylobacter coli* in farm and clinical environments.' *Applied and Environmental Microbiology* **67**(4): 1429-1436.

Fox, J. T., Reinstein S., Jacob M. E. and Nagaraja T. G. (2008). 'Niche Marketing Production Practices for Beef Cattle in the United States and Prevalence of Foodborne Pathogens.' *Foodborne Pathogens and Disease* **5**(5): 559-569.

Foxman, B. and Riley, L. (2001). 'Molecular epidemiology: Focus on infection.' *American Journal of Epidemiology* **153**(12): 1135-1141.

French, N. (2009). 'Binary genomotyping of *Campylobacter jejuni* using genes involved in cell surface-exposed macromolecule biosynthesis predicts infection source'. *Personal communication.*

French, N., Barrigas M., Brown P., Ribiero P., Williams N., Leatherborrow H., Birtles R., Bolton E., Fearnhead P. and Fox A. (2005). 'Spatial epidemiology and natural population structure of *Campylobacter jejuni* colonizing a farmland ecosystem.' *Environmental Microbiology* **7**(8): 1116-1126.

French, N., Midwinter A., Holland B., Collins-Emerson J., Pattison R., Colles F. M. and Carter P. (2008). 'Molecular epidemiology of *Campylobacter jejuni* isolated from wild bird faecal material in children's playgrounds.' *Applied and Environmental Microbiology* **75**(3): 779-83.

French, N. and Molecular Epidemiology and Veterinary Public Health Group Hopkirk Institute. (2008). 'Enhancing Surveillance of Potentially Foodborne Enteric Diseases in New Zealand: Human Campylobacteriosis in the Manawatu.' Final report: FDI / 236 /2005 http://www.nzfsa.govt.nz/science/research-projects/Campy_Attribution_Manawatu.pdf., Accessed on January 16th, 2009.

French, N. and Mullner, P. (2008). 'Risk-Based Tools to Inform Decision Making in Food Safety and Public Health - Applications in Europe and New Zealand' *FoodFrenz 2008.* Budapest, Hungary.

Frost, J. A., Kramer J. M. and Gillanders S. A. (1999). 'Phage typing of *Campylobacter jejuni* and *Campylobacter coli* and its use as an adjunct to serotyping.' *Epidemiology and Infection* **123**(1): 47-55.

Garrett, N., Devane M. L., Hudson J. A., Nicol C., Ball A., Klena J. D., Scholes P., Baker M. G., Gilpin B. J. and Savill M. G. (2007). 'Statistical comparison of *Campylobacter jejuni* subtypes from human cases and environmental sources.' *Journal of Applied Microbiology* **103**(6): 2113-2121.

Gelman, A. B., Carlin J. S., Stern H. S. and Rubin D. B. (1995). *Bayesian Data Analysis*, Chapman & Hall/CRC.

Gilbert, S., Lake R., Cressey P. and Hudson A. (2008). Risk Profile: Salmonella (non-typhoidal) in pork and pork products, Institute of Environmental Science & Research Limited (ESR).

Gilks, W. R., Richardson S. and Spiegelhalter D. J. (1996). *Markov Chain Monte Carlo in Practice*, Chapman & Hall.

Gillespie, I. A. (2003). 'Milkborne general outbreaks of infectious intestinal disease, England and Wales, 1992-2000.' *Epidemiology and Infection* **130**(3): 461-68.

Gillespie, I. A., O'Brien S. J., Frost J. A., Adak G. K., Horby P. and Svan A. V. (2002). 'A Case-Case comparison of *Campylobacter coli* and *Campylobacter jejuni* infection: A Tool for generating Hypotheses.' *Epidemiology and Infection* **128**(2): 111-118.

Gillespie, I. A., O'Brien S. J., Penman C., Tompkins D., Cowden J. and Humphrey T. J. (2008). 'Demographic determinants for *Campylobacter* infection in England and Wales: implications for future epidemiological studies.' *Epidemiology and Infection* **136**(12): 1717-1725.

Givens, G. H. and J. A. Hoeting (2005). *Computational Statistics*, John Wiley & Sons.

Godschalk, P. C. R., Bergman M. P., Gorkink R. F. J., Simons G., van den Braak N., Lastovica A. J., Endtz H. P., Verbrugh H. A. and Van Belkum A. (2006). 'Identification of DNA sequence variation in *Campylobacter jejuni* strains associated with the Guillain-Barre syndrome by high-throughput AFLP analysis.' *Bmc Microbiology* **6**.

Gormley, F. J., MacRae M., Forbes K. J., Ogden I. D., Dallas J. F. and Strachan N. J. C. (2008). 'Has retail chicken played a role in the decline of human campylobacteriosis?' *Applied and Environmental Microbiology* **74**(2): 383-390.

Greenland, S. and K. J. Rothman (1998). Measure of Effect and Measures of Association. *Modern Epidemiology*, Lippincott Williams & Wilikins: 54-58.

Grenfell, B. T., Pybus O. G., Gog J. R., Wood J. L. N., Daly J. M., Mumford J. A. and Holmes E. C. (2004). 'Unifying the epidemiological and evolutionary dynamics of pathogens.' *Science* **303**(5656): 327-332.

Guan, J., Grenier C. and Brooks B. W. (2006). 'In vitro study of *Salmonella* enteritidis and *Salmonella* typhimurium definitive type 104: Survival in egg albumen and penetration through the vitelline membrane.' *Poultry Science* **85**(9): 1678-1681.

Gustafson, P., Hossain S. and McCandless L. (2005). Innovative Bayesian Methods for Biostatistics and Epidemiology. *Bayesian Thinking: Modelling and Computation.* D. K. Dey and C. R. Rao, Elsevier. 25.

Haas, C. N., Rose J. B. and Gerba C. P. (1999). Risk Assessment Paradigms. *Quantitative Microbial Risk Assessment.* N. C. Haas, John Wiley & Sons, Inc.

Haas, N. G., Rose J. B. and Gerba C. P. (1999). *Quantitative microbial risk assessment*, John Wiley & Sons, INC.

Habib, I., Sampers I., Uyttendaele M., Berkvens D. and De Zutter L.(2008). 'Baseline data from a Belgium-wide survey of *Campylobacter* species contamination in chicken meat preparations and considerations for a reliable monitoring program.' *Applied and Environmental Microbiology* **74**(17): 5483-5489.

Habib, I., Sampers I., Uyttendaele M., De Zutter L. and Berkvens D. (2008). 'A Bayesian modelling framework to estimate *Campylobacter* prevalence and culture methods sensitivity: application to a chicken meat survey in Belgium.' *Journal of Applied Microbiology* **105**(6): 2002-2008.

Hald, T. (2002). 'Quantifying sources of human salmonellosis in the Netherlands and in Denmark: Two approaches, same result?' Danish Zoonosis Centre, Danish Veterinary Institute.

Hald, T., Vose D., Wegener H. C. and Koupeev T. (2004). 'A Bayesian approach to quantify the contribution of animal-food sources to human salmonellosis.' *Risk Analysis* **24**(1): 255-269.

Hald, T., Wegener H. C., Borck B. and Lo Fo Wong D. M. A. (2005). The integrated surveillance of *Salmonella* in Denmark and the effect on public health. *Food Safety Assurance and Veterinary Public Health.* F. J. M. Smulders and J. D. Collins, Wageningen Academic Publishers.

Hald, T., Wingstrand A., Brondsted T. and Wong D. (2006). 'Human health impact of Salmonella contamination in imported soybean products: A semiquantitative risk assessment.' *Foodborne Pathogens and Disease* **3**(4): 422-431.

Harrington, C. S., Thomson-Carter F. M. and Carter P. E. (1997). 'Evidence for recombination in the flagellin locus of *Campylobacter jejuni*: implications for the flagellin gene typing scheme.' *Journal of Clinical Microbiology* **35**(9): 2386-92.

Harris, N. V.and Weiss, N. S. (1986). 'The Role of Poultry and Meats in the Etiology of *Campylobacter- Jejuni Coli* Enteritis.' *American Journal of Public Health* **76**(4): 407-411.

Hartnett, E., Kelly L., Newell D., Wooldridge M. and Gettinby G. (2001). 'A quantitative risk assessment for the occurrence of *Campylobacter* in chickens at the point of slaughter.' *Epidemiology and Infection* **127**(2): 195-206.

Hartnett, E., Kelly L. A., Gettinby G. and Wooldridge M. (2002). 'A quantitative risk assessment for *campylobacters* in broilers: work in progress.' *International Biodeterioration & Biodegradation* **50**(3-4): 161-165.

Havelaar, A., Braeunig J., Christiansen K., Cornu M., Hald T., Mangen M.-J. J., Molbak K., Pielaat A., Snary E., Van Pelt W., Velthuis A. and Wahlstroem H. (2007). 'Towards an Integrated Approach on Supporting Microbiological Food Safety Decisions.' *Zoonosis and Public Health*(**54**): 103 -117.

Hearnden, M., Skelly C., Eyles R. and Weinstein P. (2003). 'The regionality of campylobacteriosis seasonality in New Zealand.' International *Journal of Environmental Health Research* **13**(4): 337-348.

Heir, E., B. A. Lindstedt, et al. (2004). 'Molecular epidemiology and disinfectant susceptibility of Listeria monocytogenes from meat processing plants and human infections.' *International Journal of Food Microbiology* **96**(1): 85-96.

Heir, E., Lindstedt B. A., Rotterud O. J., Vardund T., Kapperud G. and Nesbakken T. (2004). 'Molecular epidemiology and disinfectant susceptibility of *Listeria monocytogenes* from meat processing plants and human infections.' *International Journal of Food Microbiology* **96**(1): 85-96.

Hilton, A. C., Mortiboy D., Banks J. G. and Penn C. W. (1997). 'RAPD analysis of environmental, food and clinical isolates of *Campylobacter* spp.' *FEMS Immunology and Medical Microbiology* **18**(2): 119-24.

Hof, H. and J. Rocourt (1992). 'Is any strain of *Listeria Monocytogenes* detected in food a health risk?' *International Journal of Food Microbiology* **16**: 173-182.

Horvath, P., Romero D. A., Cot-Monvoisin A. C., Richards M., Deveau H., Moineau S., Boyaval P., Fremaux C. and Barrangou R. (2008). 'Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*.' *Journal of Bacteriology* **190**(4): 1401-12.

Hudson, R. R. (1998). 'Island models and the coalescent process.' *Molecular Ecology* **7**: 413 - 418.

Institute of Environmental Science and Research Limited (ESR) (2003/2004). MOAS study.

Institute of Environmental Science and Research Limited (ESR) (2004). Risk Profile: 'Salmonella in and on eggs'.

Institute of Environmental Science and Research Limited (ESR) 'Public Health Surveillance - EpiSurv.' Available at: http://www.surv.esrcri.nz/episurv/index/php, Accessed on January 16, 2009.

Isenberg, H. D. (1998). *Essential Procedures for Clinical Microbiology.* American Society for Microbiology.

Jackson, C. J., Fox A. J., Wareing D. R., Hutchinson D. N. and Jones D. M. (1996). 'The application of genotyping techniques to the epidemiological analysis of *Campylobacter jejuni*.' *Epidemiology and Infection* **117**(2): 233-44.

Jones, K. E., Patel N. G., Levy M. A., Storygard A., Balk D., Gittleman J. L. and Daszak P. (2008). 'Global trends in emerging infectious diseases.' *Nature* **451**(7181): 990 - U4.

Karenlampi, R., Rautelin H., Schonberg-Norio D., Paulin L. and Hanninen M. L. (2007). 'Longitudinal study of Finnish *Campylobacter jejuni* and *C. coli* isolates from humans, using multilocus sequence typing, including comparison with epidemiological data and isolates from poultry and cattle.' *Applied and Environmental Microbiology* **73**(1): 148-155.

Kegode, R. B., Doetkott D. K., Khaitsa M. L. and Wesley I. V. (2008). 'Occurrence of *Campylobacter* species, *Salmonella* species and generic *Escherichia coli* in meat products from retail outlets in the Fargo metropolitan area.' *Journal of Food Safety* **28**(1): 111-125.

Kelly, L. A., Hartnett E., Gettinby G., Fazil A., Snary E. and Wooldridge M. (2003). 'Microbiological safety of poultry meat: risk assessment as a way forward.' *World Poultry Science Journal* **59**(4): 495-508.

Kemp, R., Leatherbarrow A. J. H., Williams L. J., Hart C. A., Clough H. E., Turner J., Wright E. J. and French N. P. (2005). 'Prevalence and genetic diversity of *Campylobacter* spp. in environmental water samples from a 100-square-kilometer predominantly dairy farming area.' *Applied and Environmental Microbiology* **71**(4): 1876-1882.

Klena, J. (2001). A Survey of Phenotypic and Genetic Methods Used to Identify and Differntiate Thermotolerant *Campylobacter* spp. strains. A report to the Ministry of Health, Ministry of Health, New Zealand.

Kramer, J. M. (2000). '*Campylobacter* contamination of raw meat and poultry at retail sale: Identification of multiple types and comparison with isolates from humans infection.' *Jounal of Food Protection* **63**(12): 1654-1659.

Kwan, P. S., Barrigas M., Bolton F. J., French N. P., Gowland P., Kemp R., Leatherbarrow H., Upton M. and A. J. Fox A. J. (2008). 'The Molecular Epidemiology of *Campylobacter jejuni* Populations in Dairy Cattle, Wildlife and the Environment in a Farmland Area.' *Applied and Environmental Microbiology* **74**(16): 5130 - 5138.

Kwan, P. S. L., Birtles A., Bolton F. J., French N. P., Robinson S. E., Newbold L. S., Upton M. and Fox A. J. (2008). 'Longitudinal study of the molecular epidemiology of *Campylobacter jejuni* in cattle on dairy farms.' *Applied and Environmental Microbiology* **74**(12): 3626-3633.

Lake, R., Adlam B. and Perera S. (2007). Acute Gastrointestinal Ilness (AGI) study: Final study report, Institute of Environmental Science & Research Limited (ESR).

Lake, R. J., Baker M. G., Garrett N., Scott W. G. and Scott H. M. (2000). 'Estimated number of cases of foodborne infectious disease in New Zealand.' *New Zealand Medical Journal* **113**(1113): 278-281.

Lake, R. (2006). Transmission routes for Campylobacteriosis in New Zealand, Institute of Environmental Science and Research Limited (ESR).

Lake, R., Hudson A., Cressey P. and Bayne G. (2007). Quantitative Risk Model: *Campylobacter* spp. in the poultry food chain, Institute of Environmental Science & Research Limited (ESR).

Lake, R., Hudson A., Cressey P. and Gilbert S. (2007). Risk Profile: *Campylobacter jejuni / coli* in mammalian and poultry offals, Institute of Environmental Science & Research Limited (ESR).

Lake, R., Hudson A., Cressey P. and Gilbert S. (2007). Risk Profile: *Campylobacter jejuni / coli* in red meat, Institute of Environmental Science & Research Limited (ESR).

Lake, R., Hudson A., Cressey P. and Nortje G. (2003). Risk Profile: *Campylobacter jejuni / coli* in poultry (whole and pieces), Institute of Environmental Science & Research Limited (ESR).

Lake, R., Hudson A., Cressey P., Wong T. and Gilbert S. (2004). Risk Profile: *Salmonella* (non thyphoidal) on Poultry (whole and pieces), ESR.

Lake, R., Whyte R. and Kliem C. (2005). 'Evaluation of foodborne disease outbreaks / human health surveillance interface.' Retrieved Available at: http://www.nzfsa.govt.nz/science/research-projects/foodborne-surveillance/foodbornesurveillanceevaluation.pdf, Accessed on January 27, 2009.

Last, J. M., Ed. (2001). *A Dictionary of Epidemiology*, Oxford University Press.

Latimer, H. K., Jaykus L.-A., Morales R. A., Cowev P. and Crawford-Brown D. (2001). 'A Weighted Composite Dose-Response Model For Human Salmonellosis.' *Risk Analysis* **21**(2): 295-305.

Leatherbarrow, A. J. H., Hart C. A., Kemp R., Williams N. J., Ridley A., Sharma M., Diggle P. J., Wright E. J., Sutherst J. and French N. P. (2004). 'Genotypic and antibiotic susceptibility characteristics of a *Campylobacter coli* population isolated from dairy farmland in the United Kingdom.' *Applied and Environmental Microbiology* **70**(2): 822-830.

Leonard, E. E., Takata T., Blaser M. J., Falkow S., Tompkins L. S. and Gaynor E. C. (2003). 'Use of an open-reading frame-specific *Campylobacter jejuni* DNA microarray as a new genotyping tool for studying epidemiologically related isolates.' *Journal of Infectious Diseases* **187**(4): 691-4.

Lindqvist, R. and M. Lindblad (2008). 'Quantitative risk assessment of thermophilic *Campylobacter* spp. and cross-contamination during handling of raw broiler chickens evaluating strategies at the producer level to reduce human campylobacteriosis in Sweden.' *International Journal of Food Microbiology* **121**: 41-52.

Little, C. L., Richardson J. F., Owen R. J., de Pinna E. and Threlfall E. J. (2008). '*Campylobacter* and *Salmonella* in raw red meats in the United Kingdom: Prevalence, characterization and antimicrobial resistance pattern, 2003-2005.' *Food Microbiology* **25**(3): 538-543.

Luber, P., Brynestad S., Topsch D., Scherer K. and Bartelt E. (2006). 'Quantification of *Campylobacter* species cross-contamination during handling of contaminated fresh chicken parts in the kitchen.' *Applied and Environmental Microbiology* **72**(1): 66 - 70.

Madden, R. H., Moran L. and Scates P. (1996). 'Sub-typing of animal and human *Campylobacter* spp. using RAPD.' *Letters in Applied Microbiology* **23**(3): 167-70.

Manfreda, G. and A. De Cesare (2005). '*Campylobacter* and *Salmonella* in poultry and poultry products: hows and whys of molecular typing.' *World Poultry Science Journal* **61**(2): 185-197.

McBride, G., Meleason M., Skelly C., Lake R., VanDerLodgt P. and Collins R. (2005). Preliminary Relative Risk Assessment for *Campylobacter* Exposure in New Zealand. NIWA Client Report HAM2005-094.

McBride, G., Till D., Ryan T., Ball A., Lewis G., Palmer S. and Weinstein P. (2002). Freshwater Microbiology Research Programme Report: Pathogen Occurrence and Human Health Risk Assessment Analysis. Ministry for the Environment, New Zealand.

McCarthy, N. D., Gillespie I. A., Lawson A. J., Cody A. J., Neal K. R., Hawtin P. R., Owen R. J., Maiden M. C. and O'Brien S. (2007). 'A Three-Year Population Based Investigation of Human *Campylobacter jejuni* Epidemiology using Sequence Typing and Patient Survey Data.' *Zoonoses and Public Health* **54**(1): 37.

McCarthy, N. D., Shepherd S. K., Richardson J. F., Colles F. M., O'Brien S. J., Owen R. J. and Maiden M. C. (2007). 'Developing an epidemiologically robust typing approach to identify, describe and trace outbreaks of *Campylobacter jejuni* infection.' *Zoonoses and Public Health* **54**(1): 153-154.

McTavish, S. M., Pope C. E., Nicol C., Sexton K., French N. and Carter P. E.(2008). 'Wide geographical distribution of internationally rare *Campylobacter* clones within New Zealand.' *Epidemiology and Infection* **136**(9): 1244-1252.

Meat & Wool New Zealand (2006). Domestic meat consumption 1997-2004.

Medema, G. J., Teunis P. F. M., Havelaar A. H. and Haas C. N. (1996). 'Assessment of the dose-response relationship of *Campylobacter jejuni*.' International Journal of Food Microbiology **30**(1-2): 101-111.

Mellmann, A., Mosters J., Bartelt E., Roggentin P., Ammon A., Friedrich A. W., Karch H. and Harmsen D. (2004). 'Sequence-based typing of flaB is a more stable screening tool than typing of flaA for monitoring of *Campylobacter* populations.' *Journal of Clinical Microbiology* **42**(10): 4840-2.

Mickan, L., Doyle R., Valcanis M., Dingle K. E., Unicomb L. and Lanser J. (2007). 'Multilocus sequence typing of *Campylobacter jejuni* isolates from New South Wales, Australia.' *Journal of Applied Microbiology* **102**(1): 144-152.

Miller, W. G., On S. L. W., Wang G. L., Fontanoz S., Lastovica A. J. and Mandrell R. E. (2005). 'Extended multilocus sequence typing system for *Campylobacter coli*, *C. lari*, *C. upsaliensis*, and *C. helveticus*.' *Journal of Clinical Microbiology* **43**(5): 2315-2329.

Miller, W. G. and Pearson, B. M. 'Diversity within the *Campylobacter jejuni* type i restriction-modification loci.' *Microbiology*, **151** (Pt2) :337-51, February 2005.

Ministry of Health (2007). Direct Laboratory Notification of Communicable Diseases National Guidelines.

Ministry of Health (2007). Metadata Database, http://www.moh.govt.nz/PHMetadata, Accessed on April 23rd, 2009.

Mira, A. (2005). *MCMC Methods to Estimate Bayesian Parametric Models*. Bayesian Thinking: Modeling and Computation, Elsevier.

Moser, I., Lentzsch P., Rieksneuwoehner B., Schwerk P. and Wieler L. H. (2002). 'High resolution genotyping of C*ampylobacter jejuni* strains by macrorestriction analysis with XhoI and polymerase chain reaction targeting enterobacterial repetitive intergenic consensus sequences: can we predict the zoonotic potential of strains?' *Epidemiology and Infection* **129**(3): 435-43.

Mullner, P., Carter P., Midwinter A., Collins-Emerson J., Pattison R. and French N.(2007). 'Source attribution of *Campylobacter jejuni* using molecular epidemiology.' *Proceedings of the Epidemiology & Animal Health Management Branch of the NZVA* **69**.

Mullner, P., Jones, G., Noble A., Spencer, S., Hathaway, S. and French, N. (2009). 'Source Attribution of Food Borne Zoonosis in New Zealand: A Modified Hald Model.' *Risk Analysis* in press.

Mullner, P., Pleydell, E., Shadbolt, T., Collins-Emerson, J., Midwinter, A., Spencer, S. and French, N. 2008. Working at the human and animal health interface a multidisciplinary team approach towards reducing New Zealands campylobacteriosis problem. *Proceedings of the Epidemiology & Animal Health Management Branch of the NZVA*.

Mullner, P., Spencer S., Collins-Emerson J., Midwinter A., Jones G., Noble A. and French N. (2008). 'Molecular and Modeling Tools for *Campylobacter* Source Attribution.' *Proceedings of the Epidemiology & Animal Health Management Branch of the NZVA*.

Murray, M. (2002). 'Determinants of cluster distribution in the molecular epidemiology of tuberculosis.' *Proceedings of the National Academy of Sciences* **99**(3): 1538 - 1543.

Murray, N. (2002). *Import Risk Analysis Animals and Animal Products*. Wellington, New Zealand, MAF Biosecurity Authorithy.

Nauta, M., van der Fels-Klerx I. and Havelaar A. (2005). 'A poultry-processing model for quantitative microbiological risk assessment.' *Risk Analysis* **25**(1): 85-98.

Nauta, M. J. (2001). A modular process risk model structure for quantitative microbiological risk assessment and its application in an exposure assessment of Bacillus cereus in a REPFRD. RIVPM report 149106007. Bilthoven, RIVPM.

Nauta, M. J., Evers E. G., Takumi K. and Havelaar A. H. (2001). Risk assessment of Shiga-toxin producing *Escherichia coli* O157 in steak tartare in the Netherlands. RIVM report 257851003 / 2001. Bilthoven, RIVM.

New Zealand Food Safety Authority (1997). National Nutritional Survey.

New Zealand Food Safety Authority (NZFSA). (2007). '*Campylobacter* risk management strategy.' Retrieved 20 March 2009, from http://www.nzfsa.govt.nz/consumers/food-safety-topics/foodborne-illnesses/campylobacter/strategy/index.htm.

New Zealand Food Safety Authority (NZFSA). (2007). 'Modelling of exposure of New Zealand consumers to *Salmonella* risk management strategy.' Retrieved 28 April 2009, from http://www.nzfsa.govt.nz/science/research-projects/salmonella-pathways-final-270407.pdf

Newell, D. G., Frost J. A., Duim B., Wagenaar J. A., Madden R. H., van der Plas J. and On S. L. W. (2000). New Deveolpments in the subtyping of *Campylobacter* species. Campylobacter. I. Nachamkin and M. J. Blaser, American Society for Microbiology.

Nielson, B. (2004). Microbiological Safety of Meat; *Salmonella* spp. *Encyclopedia of Meat Sciences* W. K. Jensen, Elsevier. 779 - 781.

Nielsen, E. M., Engberg J., Fussing V., Petersen L., Brogren C. H. and On S. L. (2000). 'Evaluation of phenotypic and genotypic methods for subtyping *Campylobacter jejuni* isolates from humans, poultry, and cattle.' *Journal of Clinical Microbiology* **38**(10): 3800-10.

Nylen, G., Dunstan F., Palmer S. R., Andersson Y., Bager F., Cowden J., Feierl G., Galloway Y., Kapperud G., Megraud F., Molbak K., Petersen L. R. and Ruutu P. (2002). 'The seasonal distribution of *Campylobacter* infection in nine European countries and New Zealand.' *Epidemiology and Infection* **128**(3): 383-390.

Oberhelman, R. A. and D. N. Taylor (2000). *Campylobacter* infections in developing countries. *Campylobacter*. I. Nachamkin and M. J. Blaser.

Office Internationale des Epizooties (OIE) (2005). Terrestrial Animal Health Code. Paris, OIE.

Ono, K., Kurazono T., Niwa H. and Itoh K. (2003). 'Comparison of three methods for epidemiological typing of *Campylobacter jejuni* and *C. coli*.' *Current Microbiology* **47**(5): 364-71.

Oporto, B., Esteban J. I., Aduriz G., Juste R. A. and Hurtado A. (2007). 'Prevalence and strain diversity of thermophilic *campylobacters* in cattle sheep and swine farms.' *Journal of Applied Microbiology* **103**(4): 977 - 984.

Oscar, T. (2004). 'Dose-Response models for 13 strains of *Salmonella*.' *Risk Analysis* **24**(1): 41-49.

Pan American Health Organisation (2001). Zoonosis and communicable diseases common to man and animals.

Pearson, A. D. (2000). 'Continuous source outbreak of campylobacteriosis traced to chicken.' *Journal of Food Protection* **63**(3): 309 - 314.

Petersen, L. and S. L. On (2000). 'Efficacy of flagellin gene typing for epidemiological studies of *Campylobacter jejuni* in poultry estimated by comparison with macrorestriction profiling.' *Letters in Applied Microbiology* **31**(1): 14-9.

Pires, S. M., Nichols G., Whalstrom H., Kaesbohrer A., David J., Spitznagel H., Van Pelt W., Baumann A. and Hald T. (2008). *Salmonella* source attribution in different European countries. *Food Micro Conference*, Aberdeen.

Pope, C., Wilson J., Taboada E. N., MacKinnon J., Alves C. A. F., Nash J. H. E., Rahn K. and Tannock G. W. (2007). 'Epidemiology, relative invasive ability, molecular characterization, and competitive performance of *Campylobacter jejuni* strains in the chicken gut.' *Applied and Environmental Microbiology* **73**(24): 7959-7966.

Pritchard, J., Stephens M. and Donnelly P. (2000). 'Inference of Population Structure using Multilocus Genotype Data.' *Genetics* **155**: 945-959.

Ravva, S. V., Sarreal C. Z., Duffy B. and Stanker L. H. (2006). 'Survival of *Escherichia coli* O157 : H7 in wastewater from dairy lagoons.' *Journal of Applied Microbiology* **101**(4): 891-902.

Riley, L. W. (2004). *Molecular Epidemiology of Infectious Diseases - Principles and Practices*. Washington, DC, ASM Press.

Robert, P. R. and G. Casella (2004). *Monte Carlo Statistical Methods*. Springer.

Robinson, S. E., Brown P. E., Wright E. J., Hart C. A. and French N. P. (2009). 'Quantifying within- and between-animal variation and uncertainty associated with counts of *Escherichia coli* O157 occurring in naturally infected cattle faeces.' *Journal of the Royal Society Interface* **6**(31): 169-177.

Rosef, O., Kapperud G., Lauwers S. and Gondrosen B. (1985). 'Serotyping of *Campylobacter jejuni*, *Campylobacter coli* and *Campylobacter lardis* from domestic and wild animals.' *Applied and Environmental Microbiology* **49**(6): 1507-1510.

Rosenberg, N. A. (2004). 'Distruct: a program for the graphical display of population structure.' *Molecular Ecology Notes* **4**: 137-138.

Rosenquist, H., Nielsen N. L., Sommer H. M., Norrung B. and Christensen B. B. (2003). 'Quantitative risk assessment of human campylobacteriosis associated with thermophilic Campylobacter species in chickens.' *International Journal of Food Microbiology* **83**(1): 87-103.

Rothman, K. J. and S. Greenland (1998). *Modern Epidemiology*. Lippincott Williams & Wilkins.

Said, B. and Wright, F. (2003). 'Outbreaks of infectious disease associated with private drinking water supplies in England and Wales 1970-2000.' *Epidemiology and Infection* **130**(3): 469-479.

Sails, A. D., Swaminathan B. and Fields P. I. (2003). 'Clonal complexes of *Campylobacter jejuni* identified by multilocus sequence typing correlate with strain associations identified by multilocus enzyme electrophoresis.' *Journal of Clinical Microbiology* **41**(9): 4058-67.

Savill, M., Hudson A., Devane M., Garrett N., Gilpin B. and Ball A. (2003). 'Elucidation of potential transmission routes of *Campylobacter* in New Zealand.' *Water Science And Technology* **47**(3): 33-38.

Savill, M. G., Hudson J. A., Ball A., Klena J. D., Scholes P., Whyte R. J., McCormick R. E. and Jankovic D. (2001). 'Enumeration of *Campylobacter* in New Zealand recreational and drinking waters.' *Journal of Applied Microbiology* **91**(1): 38-46.

Schouls, L. M., Reulen S., Duim B., Wagenaar J. A., Willems R. J., Dingle K. E., Colles F. M. and Van Embden J. D. (2003). 'Comparative genotyping of *Campylobacter jejuni* by amplified fragment length polymorphism, multilocus sequence typing, and short repeat sequencing: strain diversity, host range, and recombination.' *Journal of Clinical Microbiology* **41**(1): 15-26.

Schwabe, C. W. (1994). Epidemiological intelligence. Veterinary medicine and human health. Baltimore, Williams & Wilkins: 392 - 429.

Schwartz D. C. and C. C. R. (1984). 'Separation of yeast chromosome sized DNAs by pulsed field gradient gel electrophoresis.' *Cell* **37**(1): 65 - 75.

Science Pod Cast (2009). 'Transcript from the 6th March 2009 pod cast.' *Science*.

Scott, W. G., Scott H. M., Lake R. J. and Baker M. G. (2000). 'Economic cost to New Zealand of foodborne infectious disease.' *New Zealand Medical Journal* **113**(1113): 281-284.

Sheppard, S. K., Dallas J. F., Strachan N. J. C., Macrae M., McCarthy,, N.D. F., D. O., I.D.,, Maiden M. C. J. and Forbes K. J. (2009). '*Campylobacter* genotyping to determine the source of human infection.' *Clinical Infectious Diseases* **48**(8): 1072 - 8.

Sheppard, S. K., Dallas J. F., Macrae M., McCarthy N. D., Sproston E. L., Gormley F. J., Strachan N. J., Ogden I. D., Maiden M. C. and Ken J. F. (2009). '*Campylobacter genotypes* from food animals, environmental sources and clinical disease in Scotland 2005/6.' *International Journal of Food Microbiology* [Epub ahead of print]

Sheppard, S. K., McCarthy N. D., Colles F. M., Richardson J. F., Cody A. J., Brick G., Meldrum R. J., Elson R., O'Brien S., Owen R. J. and Maiden M. C. J. (2007). '*Campylobacter* from retail poultry: MLST analysis and the origin of human infection.' *Zoonoses and Public Health* **54**: 36-36.

Sheppard, S. K., McCarthy N. D., Falush D. and Maiden M. C. J. (2008). 'Convergence of *Campylobacter* species: Implications for bacterial evolution.' *Science* **320**(5873): 237-239.

Shi, Z. Y., Liu P. Y., Lau Y. J., Lin Y. H., Hu B. S. and Tsai H. N. (1996). 'Comparison of polymerase chain reaction and pulsed-field gel electrophoresis for the epidemiological typing of *Campylobacter jejuni*.' *Diagnostic Microbiology and Infectious Disease* **26**(3-4): 103-8.

Skirrow, M. B. and M. J. Blaser (2000). Clinical Aspects of *Campylobacter* Infection. *Campylobacter*. I. Nachamkin and M. J. Blaser, American Society for Microbiology.

Slorach, S. A., Maijala R. and Belveze H. (2002). Examples of comprehensive and integrated approaches to risk analysis in the food chain experiences and lessons learned. Pan-European conference on Fodd Safety and Quality. Budapest, Hungary.

Smith, E. M., Green L. E., Medley G. F., Bird H. E., Fox L. K., Schukken Y. H., Kruze J. V., Bradley A. J., Zadoks R. N. and Dowson C. G. (2005). 'Multilocus sequence typing of intercontinental bovine *Staphylococcus aureus* isolates.' *Journal of Clinical Microbiology* **43**(9): 4737-4743.

Snow, L. C., Davies R. H., Christiansen K. H., Carrique-Mas J. J., Cook A. J. C., Teale C. J. and Evans S. J. (2008). 'Survey of the prevalence of *Salmonella* on commercial broiler farms in the United Kingdom, 2005/06.' *Veterinary Record* **163**(22): 649-654.

Sofos, J. N. (2002). Approaches to pre-harvest food safety assurance. *Food safety assurance in the pre-harvest phase.* F. J. M. Smulders and J. D. Collins.

Sopwith, W., Ashton M., Frost J. A., Tocque K., O'Brien S., Regan M. and Syed Q. (2003). 'Enhanced Surveillance of *Campylobacter* Infection in the North West of England 1997-1999.' *Journal of Infection* **46**: 35-45.

Sopwith, W., Birtles A., Matthews M., Fox A., Gee S., Painter M., Regan M., Syed Q. and Bolton E.(2006). '*Campylobacter jejuni* multilocus sequence types in humans, Northwest England, 2003-2004.' *Emerging Infectious Diseases* **12**(10): 1500-1507.

Sopwith, W., Birtles A., Matthews M., Fox A., Gee S., Painter M., Regan M., Syed Q. and Bolton E. (2008). 'Identification of Potential Environmentally Adapted *Campylobacter jejuni* Strain, United Kingdom.' *Emerging Infectious Diseases* **14**(11): 1769-1773.

Sorek, R., Kunin V. and Hugenholtz P. (2008). 'CRISPR a widespread system that provides acquired resistance against phages in bacteria and archaea.' *Nature Reviews Microbiology* **6**(3): 181-6.

Spiegelhalter, D. J., K. R. Abrams, et al. (2004). *Bayesian Approaches to Clincal Trials and Health-Care Evaluation.* John Wiley & Sons, Ltd.

Spiegelhalter, D. J., Myles J. P., Jones D. R. and Abrams K. R. (1999). 'Methods in health service research - An introduction to bayesian methods in health technology assessment.' *British Medical Journal* **319**(7208): 508-512.

Spratt, B. G., Hanage W. P., Li B., Aanensen D. M. and Feil E. J. (2004). 'Displaying the relatedness among isolates of bacterial species - the eBURST approach.' *FEMS Microbiology Letters* **241**(2): 129-134.

Stafford, R. J., Schluter P. J., Wilson A. J., Kirk M. D., Hall G. and Unicomb L. (2008). 'Population-attributable risk estimates for risk factors associated with *Campylobacter* infection, Australia.' *Emerging Infectious Diseases* **14**(6): 895-901.

Stanley, T., Wareing D. R., Murphy P. G., Millar B. C., Wilson T. and Moore J. E. (2004). 'Multilocus enzyme electrophoresis typing of clinical *campylobacters* from outbreak and sporadic sources.' *British Journal of Biomedical Science* **61**(1): 26-31.

Steele, M., McNab B., Fruhner L., DeGrandis S., Woodward D. and Odumeru J. A. (1998). 'Epidemiological typing of *ampylobacter* isolates from meat processing plants by pulsed-field gel electrophoresis, fatty acid profile typing, serotyping, and biotyping.' *Applied and Environmental Microbiology* **64**(7): 2346-9.

Strachan, N. J. C., Dunn G. M., Locking M. E., Reid T. M. S. and Ogden I. D. (2006). '*Escherichia coli* O157: Burger bug or environmental pathogen?' *International Journal of Food Microbiology* **112**(2): 129-137.

Strachan, N. J. C., Gormley F. J., Rotariu O., Ogden I. D., Miller G., Dunn G. M., Sheppard S. K., Dallas J. F., Reid T. M. S., Howie H., Maiden M. C. J. and Forbes K. J. (2009). 'Attribution of *Campylobacter* Infections in Northeast Scotland to Specific Sources by Use of Multilocus Sequence Typing.' *Journal of Infectious Diseases* **199**: 1 - 4.

Struelens, M. J. and Members of the European Study Group on Epidemiological Markers (ESGEM) of the European Society for Clinical Microbiology and Infectious Diseases (ESCMID) (1996). 'Consensus guidelines for appropriate use and evaluation of microbial epidemiologic typing systems.' *Clinical Microbiology and Infection* **2**(1): 2-11.

Stucki, U. and Frey, J. (1995) 'Identification of *Campylobacter jejuni* on the basis of a species-specific gene that encodes a membrane protein.' Journal of Clinical Microbiology, **33**(4) : 855-9.

Sullivan, C. B., Diggle M. A. and Clarke S. C. (2005). 'Multilocus sequence typing - Data analysis in clinical microbiology and public health.' *Molecular Biotechnology* **29**(3): 245 - 254.

Taboada, E. N., MacKinnon J. M., Luebbert C. C., Gannon V. P. J., Nash J. H. E. and Rahn K. (2008). 'Comparative genomic assessment of Multi-Locus Sequence Typing: rapid accumulation of genomic heterogeneity among clonal isolates of *Campylobacter jejuni*.' *Bmc Evolutionary Biology* **8**.

Tanaka, D. L., Karn J. F. and Scholljegerdes E. J. (2008). 'Integrated crop/livestock systems research: Practical research considerations.' *Renewable Agriculture and Food Systems* **23**(1): 80 - 86.

Tenover, F. C., Arbeit R. D., Goering R. V., Mickelsen P. A., Murray B. E., Persing D. H. and Swaminath (1995). 'Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing.' *Journal of Clinical Microbiology* **33**: 2233-2239.

Thakur, S., Morgan Morrow M. E., Funk J. A., Bahnson P. B. and Gebreyes W. A. (2006). 'Molecular Epidemiologic Investigation of *Campylobacter coli* in Swine Production Systems, Using Multilocus Sequence Typing.' *Applied and Environmental Microbiology* **72**(8): 5666-5669.

Thompson, R. C. A. (2000). *Molecular Epidemiology of Infectious Diseases*. Oxford University Press.

Thrusfield, M. (2005). *Veterinary Epidemiology*. Blackwell Science.

Titus, S. M. (2007). 'A novel model developed for Quantitative Microbial Risk Assessment in the pork food chain.' Institute of Veterinary, Animal and Biomedical Sciences. Massey University.

Toft, N., Innocent G. T., Gettinby G. and Reid S. W. J. (2007). 'Assessing the convergence of Markov Chain Monte Carlo methods: An example from evaluation of diagnostic tests in absence of a gold standard.' *Preventive Veterinary Medicine* **79**: 244-256.

Thrusfield, M. (2005). The scope of epidemiology. *Veterinary Epidemiology*. Blackwell Publishing: 23 - 33.

Urwin, R. and M. C. J. Maiden (2003). ' Multilocus sequence typing: a tool for global epidemiology.' *Trends in Microbiology* **11**(10): 479 - 487.

Uyttendaele, M., De Troy P. and Debevere J. (1999). 'Incidence of *Salmonella*, *Campylobacter jejuni*, *Campylobacter coli*, and *Listeria monocytogenes* in poultry carcasses and different types of poultry products for sale on the Belgian retail market.' *Journal of Food Protection* **62**(7): 735-740.

Van Belkum, A. (2003). 'High-throughput epidemiologic typing in clinical microbiology.' *Clinical Microbiology and Infection* **9**(2): 86-100.

Van Boekel, M. A. J. S., Stein A. and Van Bruggen A. H. C. (2003). Bayesian statistics and quality modelling in the agro-food production chain. Frontis Workshop on Bayesian Statistics and Quality Modelling in the Agro-Food Production Chain, Wageningen, Kluwer Academic Publishers.

Van Dongen, S. (2006). 'Prior specification in Bayesian statistics: Three cautionary tales.' *Journal of Theoretical Biology* **242**(1): 90-100.

Van Mulligen, E. M., Cases M. and Hettne K.(2008). 'Training Multidisciplinary Biomedical Informatics Students: Three Years of Experience.' *Journal of the American Medical Informatics Association* **15**(2): 246 - 254.

Vora, P., Senecal A. and Schaffner D. W. (2003). 'Survival of *Staphylococcus aureus* ATCC 13565 in intermediate moisture foods is highly variable.' *Risk Analysis* **23**(1): 229-236.

Vos, P., Hogers R., Bleeker M., Reijans M., Lee T., van der Hornes M., Frijters A., Pot J., Peleman J., Kuiper M. and Zabeau M. (1995). 'AFLP: a new technique for DNA fingerprinting.' *Nucleic Acids Research* **23**(21): 4407-4414.

Vose, D. (2000). Risk Analysis A quantitative guide, John Wiley & Sons.

Vose, D., Acar J., Anthony F., Franklin A., Gupta R., Nicholls T., Tamura Y., Thompson S., Threlfall E. J., Van Vuuren M., White D. G., Wegener H. C. and Costarrica M. L. (2001). 'Antimicrobial resistance: risk analysis methodology for the potential impact on public health of antimicrobial resistant bacteria of animal origin.' *Revue Scientifique et Technique Office International Des Epizooties* **20**(3): 811-827.

Wagenaar, J. A., Mevius D. J. and Havelaar A. H. (2006). '*Campylobacter* in primary animal production and control strategies to reduce the burden of human campylobacteriosis.' *Revue Scientifique Et Technique Office International Des Epizooties* **25**(2): 581-594.

Wahlstroem, Anderson H., Plym-Forsell Y. and Pires L. S. M. (2008). Source Attribution of Reported Human Salmonella Cases in Sweden. *FoodMicro 2008*. Aberdeen.

Walls, I. and V. N. Scott (1997). 'Use of predictive microbiology in microbial food safety risk assessment.' *International Journal of Food Microbiology* **36**: 97-102.

Wareing, D. R. A., Bolton F. J., Fox A. J., Wright P. A. and Greenway D. L. A. (2002). 'Phenotypic diversity of Campylobacter isolates from sporadic cases of human enteritis in the UK.' *Journal of Applied Microbiology* **92**(3): 502-509.

Whiting, R. and M. H. Golden (2002). 'Variation among *Escherichia coli* 0157:H7 strains relative to their growth, survival, thermal inactivation and toxin production in broth.' *International Journal of Food Microbiology* **75**: 127-133.

Whiting, R. C. and R. L. Buchanan (1997). 'Development of a quantitative risk assessment model for *Salmonella* enteritidis in pasteurized liquid eggs.' *International Journal of Food Microbiology* **36**(2-3): 111-125.

Wieland, B., Wittwer M., Regula G., Wassenaar T. M., Burnens A. P., Keller J. and Stark K. D. C. (2006). 'Phenon cluster analysis as a method to investigate epidemiological relatedness between sources of *Campylobacter jejuni*.' *Journal of Applied Microbiology* **100**(2): 316-324.

Wilsom, M. W. (2007). Survey of Retail Eggs For *Salmonella*, Institute of Environmental Science & Research Limited (ESR).

Wilson, D. J., Gabriel E., Leatherbarrow A. J., Cheesbrough J., Gee S., Bolton E., Fox A., Fearnhead P., Hart C. A. and Diggle P. J. (2008). 'Tracing the source of campylobacteriosis.' *PLoS Genetics* 4:e1000203.

Wilson, D. J., Gabriel E., Leatherbarrow A. J., Cheesbrough J., Gee S., Bolton E., Fox A., Hart C. A., Diggle P. J. and Fearnhead P. (2009). 'Rapid Evolution and the Importance of Recombination to the Gastroenteric Pathogen *Campylobacter jejuni*.' *Molecular Biology and Evolution* **26**(2): 385-397.

Wirth, T., Falush D., Lan R., Colles F., Mensa P., Wieler L. H., Karch H., Reeves P. R., Maiden M. C., Ochman H. and Achtman M. (2006). 'Sex and virulence in *Escherichia coli*: an evolutionary perspective.' *Molecular Microbiology* **60**(5): 1136 - 1151.

Wong, T. L. (2005). Microbiological Survey of Imported and New Zealand pork: *Salmonella*, generic *Escherichia coli* and *E. coli* 0157:H7, Institute of Environmental Science & Research Limited (ESR).

Wong, T. L., Hollis L., Cornelius A., Nicol C., Cook R. and Hudson J. A. (2007). 'Prevalence, numbers, and subtypes of *Campylobacter jejuni* and *Campylobacter coli* in uncooked retail meat samples.' *Journal of Food Protection* **70**(3): 566-573.

Wong, T. L., Nicol C., Cook R. and MacDiarmid S. (2007). '*Salmonella* in uncooked retail meats in New Zealand.' *Journal of Food Protection* **70**(6): 1360-1365.

World Health Organisation. (2007). 'About Risk Analysis in Food.' from http://www.who.int/foodsafety/micro/riskanalysis/en/.

Wright, S. (1931). 'Evolution in Mendelian Populations.' *Genetics* **16**(2): 97 - 159.

Xiao, Y., Clancy D., French N. P. and Bowers R. G. (2006). 'A semi-stochastic model for *Salmonella* infection in a multi-group herd.' *Mathematical Biosciences* **200**(2): 214-233.

Yamazaki-Matsune, W., Taguchi M., Seto K., Kawahara R., Kawatsu K., Kumeda Y., Kitazato M., Nukina M., Misawa N. and Tsukamoto T. (2007). 'Development of a multiplex PCR assay for identification of *Campylobacter coli*, *Campylobacter fetus*, *Campylobacter hyointestinalis subsp hyointestinalis*, *Campylobacter jejuni*, *Campylobacter lari* and *Campylobacter upsaliensis*.' *Journal of Medical Microbiology* **56**(11): 1467-1473.

Zadoks, R. N., Schukken Y. H. and Wiedmann M.(2005). 'Multilocus sequence typing of *Streptococcus uberis* provides sensitive and epidemiologically relevant subtype information and reveals positive selection in the virulence gene pauA.' *Journal of Clinical Microbiology* **43**(5): 2407-2417.