



Final report: 11178
**Dynamic modelling of *Campylobacter* sources in the
Manawatu**

August 2009

prepared for Dr Donald Campbell
New Zealand Food Safety Authority

by
Professor Nigel French
Dr Jonathan Marshall

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



Final Report: 11178

Dynamic modelling of *Campylobacter* sources in the Manawatu

August 2009

1 Introduction

New Zealand has one of the highest per-capita incidence rates of campylobacteriosis in the world. Understanding the reasons for this is crucial to reducing the incidence in the future. A major part of this is having a clear idea of what the sources and pathways of infection are, and identifying which of these sources and pathways are most prevalent among reported cases.

Source attribution is the process of determining the proportions in which the various pathways and sources contribute to the total disease incidence. This information is critical in creating targeted intervention strategies, and for assessing the effectiveness of such strategies. Unfortunately source attribution of pathogens is rarely accomplished due to numerous difficulties. Problems include inconsistencies in the traditional methods of data collection from sporadic cases, the difficulty of detecting smaller outbreaks, and the difficulties of conducting laboratory analysis of both human and environmental (including food and water) samples.

The most straightforward way to quantify the effect of an exposure would be to estimate the numbers of cases that were caused by this exposure. However,

this number is not estimable from ordinary incidence data, as the observation of an exposed case does not reveal the mechanism that caused the disease [6]. In contrast to an outbreak situation, where the attributable risk fraction for an identified risk factor would be very high, the source of infection in sporadic cases is more difficult to identify [10].

Traditional approaches to source attribution include full risk assessments, analysis and extrapolation of surveillance or outbreak data, and analytical epidemiological studies [1, 3, 4, 5]. Recent approaches include the Proportional Similarity Index (PSI), as well as various statistical models (the Dutch, Hald and Island models).

The Proportional Similarity Index is a general technique for comparing the area of intersection between two probability distributions [11]. By comparing the distribution of sequence types observed among human isolates with the distribution of sequence types observed among food or environmental sources, a score can be derived, with a higher score indicating a closer similarity. Scores are given between 0 and 1, with 1 indicating a complete overlap, and 0 indicating no similarity.

The Dutch model uses the relative occurrence p_{ij} of each bacterial subtype i in each source j with the number of human cases caused by that subtype o_i in order to estimate the number of cases per source λ_j :

$$\lambda_j = \sum_i \frac{p_{ij}}{\sum_j p_{ij}} o_i.$$

The Hald model [7] generalises the Dutch model, by adding additional terms to the estimate for λ_{ij} , the number of cases of subtype i attributed to source j . Both type-specific and source-specific terms, in addition to relative consumption of the various sources are added into a Bayesian framework which allows for uncertainty surrounding each parameter to be explicitly included and quantified. This model has recently been modified for use on New Zealand *C. jejuni* data by P.Mullner et. al. [8] and is detailed further in Section 3.1.

Finally, the Island model [2] uses a genetics based approach to reconstruct the genealogy of each isolate. Based on the allelic profiles of each isolate, mutation and recombination rates are estimated for each source ‘island’, in addition to migration rates to the human ‘island’. The migration rates are then used to estimate the relative contribution from each source. This modelling technique has the advantage in that it can attribute human isolates that have not yet been observed in any of the source reservoirs - all other models must discard any such isolates. We give further details of this model in Section 3.2.

Each of these models are designed to give a single estimate for the proportion of cases attributed to each source over time - they do not allow for temporal variation of that attribution. Adding temporal variation to these models is the main focus of this report. We summarise the application of extensions to the Hald, Dutch, and Island models for source attribution to the isolates from a sentinel collection site over a 4 year period, focusing on the temporal variation in the attribution through time.

2 Data collection

Over the period of 2005 through to 2009, isolates from human cases of campylobacteriosis and environmental and food sources were collected and sequence typed using Multilocus Sequence Typing (MLST). MLST is based on a sequence of 7 housekeeping genes, which are relatively conserved in an evolutionary sense. This technique allows a high level of diversity between different sequence types, while rationalising this diversity into groups of isolates with related genotypes. By utilising the relative prevalence of sequence types within each of the environmental and food sources, one can obtain estimates on the likelihood of a human isolate arising from those sources.

A total of 1267 human samples that were identified as positive for *Campylobacter* by ELISA (ProSpecT R, Remel, USA) by MedLab Central, Palmer-

	2005*	2006	2007	2008	2009*	Total
All samples	263	397	359	182	66	1267
Primary samples	200	285	274	144	50	953
Confirmed by PCR	153	224	200	100	41	718
Complete MLST profiles	133	200	181	76	34	624

Table 1: Yearly number of human isolates, from 1st March 2005 through 30th April 2009.

ston North, were sent to the Hopkirk Molecular Epidemiology laboratory over the period from 1st March 2005 to 30th April 2009. Of these, 953 isolates were identified as primary cases, with the rest duplicate isolates from the same case, or not from the Manawatu region. 718 of these have been confirmed as *C. jejuni* and 624 have complete allelic profiles. Table 1 shows the number of isolates for each of the years.

In addition, 766 isolates from food and environmental sources have been identified as *C. jejuni* and have been sequence typed. The food samples were collected from meat sampled at retail stores in Palmerston North over the same time period. Fresh whole poultry carcasses from each of the four major poultry companies were sampled, as was fresh red meat and offal (beef and lamb mince and liver). Water was sampled from 6 popular river swimming locations in the Manawatu, and cattle and sheep faeces from farms adjacent to the catchments of these river sources were also sampled. The recreational swimming sites were Mangapapa stream, Woodville; Manawatu River, Hopelands picnic reserve, Hopelands; Oroua River, Timona Park, Feilding; Manawatu River, Albert Street, Palmerston North; Tokomaru River, Horseshoe Bend, Tokomaru and Kaikokopu Stream, Himatangi Beach.

The source isolates were then grouped into one of four sources: Poultry (consisting of isolates from the 4 poultry companies), Bovine (beef mince and liver, as well as cattle faecal samples), Ovine (lamb mince and liver, and sheep faecal samples), and Environmental (water samples). Table 2 shows the number of isolates in each source type.

Further details of the data collection and background to the project may

Source	Number of Isolates
Poultry (retail)	391
Bovine (retail and live animal)	116
Ovine (retail and live animal)	168
Environment (water)	91
Total	766

Table 2: Total number of isolates in each source.

be found in the report “Enhancing the surveillance of potentially foodborne enteric disease” [9].

3 Source attribution modelling

There have been several models proposed for source attribution, each of which assess the proportion of infections attributable to a particular source in various ways. These models, however, do not take into account temporal variation that may occur in the data.

Of the existing models, the Modified Hald and Island models are most amenable to the incorporation of temporal variation. Each of these models attribute a human infection to a source with a certain probability, based on sequence type information, such as from multilocus sequence typing (MLST).

We present the most suitable modification of the Hald model, along with two different uses of the Island model: Multiple runs of the Island model through time, and a combined Dutch/Island model.

3.1 The Modified Hald Model

The modified Hald model [8] is a bayesian model of source attribution that estimates the expected number of infections of each MLST sequence type attributable to each source. We proposed 4 further modifications to this model in order to incorporate time dependence of attribution.

Let λ_{ijt} , be the expected number of infections of sequence type i attributable to source j during time period t . Then the observed counts of a particular sequence type during that time period o_{it} may be given by

$$o_{it} \sim \text{Poisson}(\sum_j \lambda_{ijt}).$$

It is clear that λ_{ijt} will be dependent on the prevalence of each sequence type on each source, p_{ij} , as if a particular sequence type is only ever observed in a particular source, then one may assume that any human infection of that sequence type is most likely to originate from that source.

In addition, we assume that the λ_{ijt} is dependent on some bacteria dependent factor q_i , and a source dependent factor a_j , so that

$$\lambda_{ijt} = p_{ij}q_i a_j.$$

The source dependent factor a_j , then, may be thought of as summarising the ability of a source to act as a vehicle for infections. It thus incorporates any conditions for the introduction and sustaining of bacterial contamination throughout the farm to fork chain. The bacteria dependent factor q_i , on the other hand, is a measure of the ability of a particular sequence type to cause disease, combining information on the virulence, survivability and pathogenicity of the sequence type.

If there is good information available on the prevalence of particular sequence types on each of the sources being considered, then one could use point estimates for the p_{ij} directly from the data. This was used in the original application of the Hald model [7], where Denmark's extensive surveillance system meant that accurate point estimates of the p_{ij} were available.

In the New Zealand situation, however, we do not have enough data to ensure that such point estimates would be valid, having only 766 data points over a period of 4 years. Hence, Mullner et. al. modified the Hald model, by first modelling the p_{ij} using a separate bayesian scheme, thus introducing

variability in the prevalence estimates.

The time dependence of the λ_{ijt} 's may then be brought in by:

1. Adding an independent time term b_t .
2. Adding time dependence to the prevalence terms p_{ij} .
3. Adding time dependence to the bacteria factors q_i .
4. Adding time dependence to the source factors a_j .

One key problem with the Modified Hald model as presented is the lack of identifiability. Assuming there is no time dependence, that we have a total of I sequence types and J sources, and that estimates of p_{ij} are available, then both the q_i and a_j terms need to be estimated from the o_i data points. We cannot possibly identify the $I + J$ parameters from I data points, and thus any estimates of these parameters will be necessarily sensitive to the priors used.

When the model is extended to consider T time periods, then under each of the 4 models above we have

1. $I + J + T$ parameters from IT data points;
2. $I + J$ parameters from IT data points, with good estimates of p_{ijt} being required;
3. $IT + J$ parameters estimated from IT data points;
4. $I + JT$ parameters estimated from IT data points.

We note that model 1 is useful for modelling dynamic behaviour, however the independence means that it cannot be used to identify which parts of the epidemiology are likely to be the cause of the observed changes in attribution. The variability may have been induced by some change in processing of a

particular foodgroup (e.g. the use of chemical decontamination in the poultry industry) or may be due to a bacterial subtype becoming more or less virulent, or due to consumption of a particular foodgroup decreasing. It allows the total expected prevalence to change through time, while the distribution of prevalence across sequence types and source factors does not change. Model 1 is in fact a subset of each of the other models, where we assume that the time dependent behaviour is independent of source, bacteria and prevalence. The other models, therefore, will be more useful at explaining any temporal variation.

Model 2 is problematic as it requires prevalence estimates for each bacterial sequence type on each source at each time step. With just 766 data points over the 4 year period, split over up to 92 bacteria types and up to 10 sources, further subdividing of the time periods is unfeasible. Furthermore, a temporal analysis of the prevalence data indicates that much of the variation is likely due to the order in which data was collected, rather than there being underlying temporal changes in the prevalence of individual bacteria on different sources.

Figure 1 shows the number of isolates collected through time from each food source, showing that the data are clustered in time, with large numbers of isolates coming in from the same source at the same time, as evidenced by the early peaks in the poultry source. Some sources have also been sampled only in the latter years, which may bias the 2005 and 2006 attribution towards those sources that had more data collected. Any true temporal change in the prevalence, therefore, is likely swamped by variance due to the collection method and the likelihood of the sample yielding an isolate. It is suggested, therefore, that this model is not appropriate in the New Zealand context with the data available.

Of the final two models, model 4 is the preferred option. Identifiability is improved in comparison to model 3, given that the number of sequence types I is typically far greater than the number of sources J . The key assumption of Model 4 is that the relative prevalence between sequence types on a par-

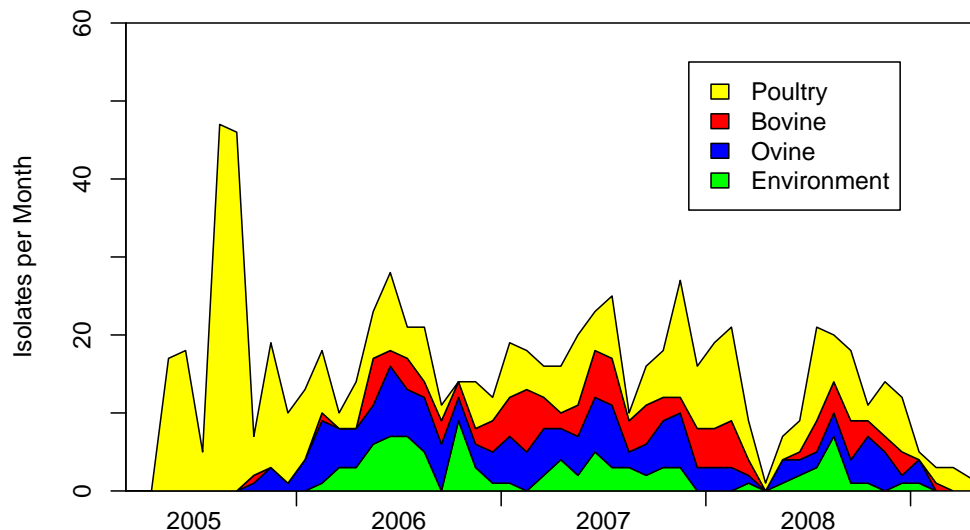


Figure 1: Number of isolates from each of the sources per month.

ticular source is constant through time: Only the level of infection caused by all sequence types may change. Any temporal variation, therefore, is attributable to changes in either the prevalence of positive isolates in the source, or the bacterial levels present on those positive isolates, or a combination of the two. Both a higher prevalence or higher counts of bacteria among the infected sources imply a larger potential to cause infection. Given that the source factor applies to all bacterial subtypes on a given source, one expects that any changes in these factors are likely to have a larger impact on source attribution than changes in particular bacterial factors. Improvements to food processing, for instance, are likely to reduce bacterial contamination (both in terms of prevalence and counts) across all sequence types, which may dramatically alter the attribution. Thus, from a regulatory point of view, changes in the source factors are likely of more interest. Model 4, therefore, has been chosen as the final version of the Modified Hald model.

As with previous versions of the Hald model, the prevalence parameters are found using a separate Bayesian process, by assuming that that we can model p_{ij} as

$$p_{ij} = \pi_j r_{ij}$$

where π_j is the prevalence over all types in source j , and r_{ij} is the relative occurrence of type i in source j . Assuming independent priors $r_{ij} \sim \text{Dirichlet}(1, 1, \dots, 1)$ and $\pi_j \sim \text{Beta}(1, 1)$, a bayesian analysis shows that we obtain independent Dirichlet and Beta posteriors for the r_{ij} and π_j .

We then use the posterior means and standard deviations for the prevalence p_{ij} to estimate the parameters α_{ij}, β_{ij} of a Beta distribution, and use this Beta distribution as the prior for the p_{ij} in the source attribution model. Due to convergence problems for very small α_{ij} values, we limit α_{ij} to be at least 1, and adjust β_{ij} so that the means match. This split modelling approach allows for some variation in the p_{ij} around the actual data values, thus accounting for the lack of prevalence data.

The final model is then:

- Model the p_{ij} , using independent Dirichlet and Beta priors, and obtain posterior means and variances.
- Estimate α_{ij} and β_{ij} via the method of moments, so that the posterior distribution $p_{ij} \sim \text{Beta}(\alpha_{ij}, \beta_{ij})$.
- Model λ_{ijt} using the Beta prior on p_{ij} and suitable priors on the q_i and a_{jt} parameters.

Lastly, we need priors for the q_i and a_{jt} parameters. The sequence type factors q_i are modelled as random observations from a lognormal distribution $\log(q_i) \sim N(0, \tau)$, where the hyperparameter τ (with prior $\text{Gamma}(0.01, 0.01)$) controls the variation in characteristics between types. Similarly, the source factors a_{jt} are given priors of $a_{jt} \sim \text{Exp}(\lambda)$. Thus, we consider each source factor at each time period to be independent to the previous ones, preferring

the data to show any correlation through time rather than attempting to model this correlation directly. Sensitivity analysis showed that λ could take a fairly large range without altering the output significantly, and a value of $\lambda = 0.002$ was used for the final model.

3.2 The Island Model

The Island Model [2] uses an evolutionary model to assign sequence types to a particular ‘island’ or source population. Sequence types of known origin (i.e. isolates from food and environmental sources) are first used to estimate the evolutionary parameters for each island. The posterior distribution of the evolutionary parameters are then used infer the origin of the human isolates, and in doing so, estimate the proportion of human cases attributable to each source population. This modelling strategy allows sequence types previously unobserved in any of the sources to be assigned to a source, based on its similarity to other sequence types from that source.

In addition to being able to assign previously unobserved isolates, the second major benefit of the Island model is that it is relatively fast to run, and can be used to obtain reasonably robust estimates in a matter of minutes, in comparison with the Hald model that may take several hours to run.

We can incorporate dynamic time behaviour in two separate ways. The first and most obvious approach is to divide the data up into separate intervals of time, and run the model separately on each interval. We need to ensure that this is done in a way that allows sufficient data resolution in each interval in order to obtain robust estimates. With the sparsity of the New Zealand data, it is suggested that a sliding window approach may be the most useful, where we infer for the current time period using data from that time period combined with data from previous time periods.

As with the modifications to the Hald model, it is thought that the prevalence of isolates within the sources is unlikely to change as rapidly as the result-

ing human isolates, as the human isolates are more significantly affected by improvements in food handling and industry processing. Thus, we consider a wide window of 12 months for the source isolates, and a smaller window of 2 months for the human isolates. The much wider smoothing window of the source isolates helps to reduce the effect of clustering in time due to the data collection regime.

The second use of the Island model is to use a single run based on all source and human isolates to obtain the probabilities that each particular human isolate has come from a particular source. Once found, we then run a subsequent analysis on the human isolates alone that takes these probabilities in addition to the number of each human sequence type at each time interval o_{it} to obtain the total source attribution for that time period. Let s_{jt} be the attribution to source j at time period t . Then

$$s_{jt} = \frac{\sum_i o_{it} b_{ij}}{\sum_i o_{it}}, \quad (1)$$

where b_{ij} is the probability that sequence type i is attributable to source j , as given by the Island Model.

This modelling approach allows for very quick analysis of newly collected human isolates as they come in, but does have the same assumptions as the Modified Hald model in terms of assuming that the attributions of each sequence type are constant in time. This assumption is equivalent to assuming that sequence types are likely adapted to a particular source, and that any further adaptation to a new source is likely to coincide with a change in the biology, and thus the introduction of a new sequence type. We consider this a valid assumption for the time frame that we are dealing with, with the major benefit being the use of all available source isolates to determine the attribution.

4 Results

All models were run over the period from 1st March 2005 through 30th April 2009, utilising 766 source isolates grouped into 4 sources: Poultry, Ovine, Bovine, and Environmental. The environmental source includes isolates found in environmental water, or attributed to wildbirds such as ducks and geese. A total of 624 human isolates over this same time period are also examined.

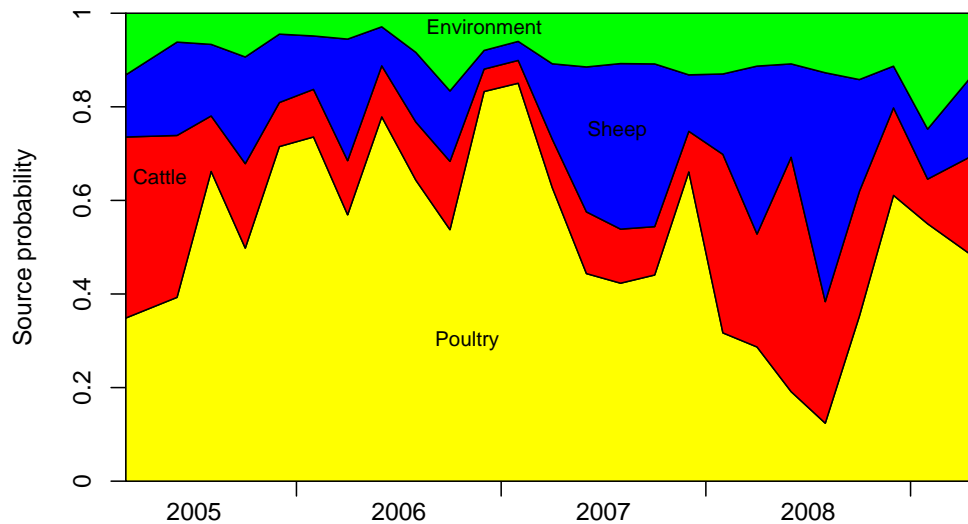


Figure 2: Source attribution for human isolates for 1st March 2005 through 30th April 2009, using bi-monthly intervals using the modified Hald Model.

Figure 2 shows the modified Hald model run using bi-monthly time intervals, with 5 chains of 10000 iterations following a burn-in period of 2000 iterations. The bi-monthly time period was chosen based on obtaining reasonably accurate estimates whilst still allowing sufficient temporal resolution to easily see variation in the attribution through time. Once the 27 human isolates that

were not present in the source isolates were excluded, this gave an average of 24 isolates per time period. As can be seen, poultry isolates dominate, particularly in the summer peaks. Of note is the reduced attribution to poultry during 2008, which corresponds to increased measures at reducing poultry contamination within the poultry industry. However, there is an increase again in the 2008/2009 summer period.

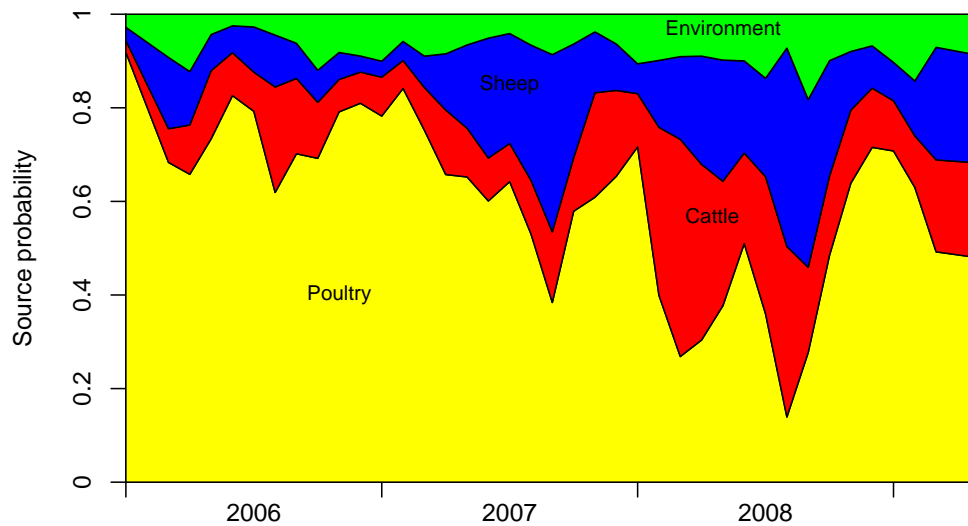


Figure 3: Source attribution for human isolates for 1st March 2005 through 30th April 2009, based on runs of the Island model a sliding window of 12 months for source isolates and 2 months for human isolates.

Figure 3 shows the Island model run over the period from 2005 through 2009, where the model has been run monthly, using a historical sliding window of 12 months for the source isolates and 2 months for the human isolates. Thus, the estimate for April uses source isolates going back to May of the previous year and human isolates from both March and April. We see a similar picture as with the Hald model, with peaks of poultry attributed infection in the summer months, and a large dip in poultry attributed isolates during 2008.

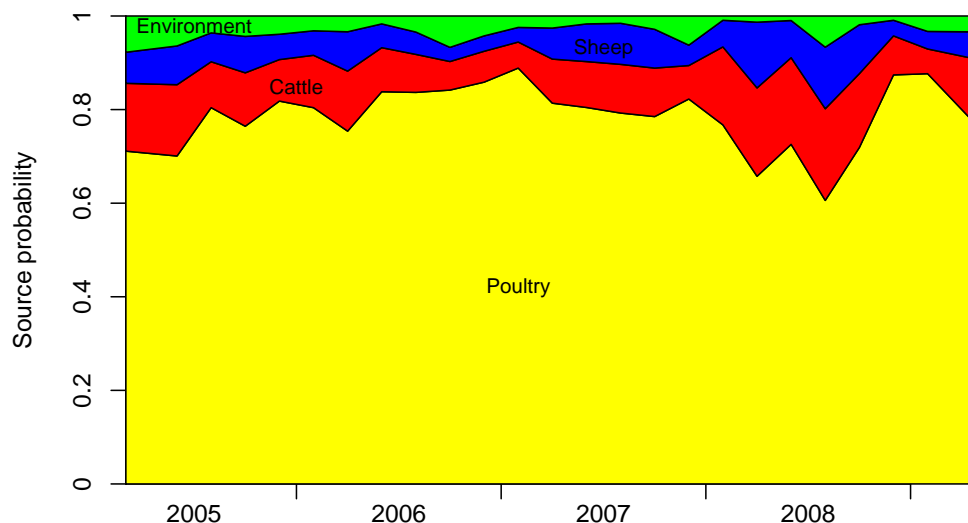


Figure 4: Source attribution for human isolates for 1st March 2005 through 30th April 2009, using bi-monthly intervals with the Dutch/Island model.

Lastly, Figure 4 shows the results from the Dutch/Island model run, where the attribution of each human isolate is computed first using the Island model, and then final attribution is computed at each time period using equation 1. The time periods were bi-monthly, as used for the Hald model. Here we see the effect of the summer peaks seems to have vanished, possibly due to the averaging effect of computing the attribution across the entire time period at the same time. The dip in poultry attribution during 2008, however, is still present, though is of a much lower magnitude. The results of this modelling approach appear to give a far broader attribution to the poultry sources than the other two models. Even sequence types that are traditionally thought of as being ruminant types such as ST-61, which appears only 4 times in the poultry sources compared to 27 times in the ruminant sources, is equally attributed to the poultry and ruminant sources by the Island model. Further research is needed to pin point why this is occurring.

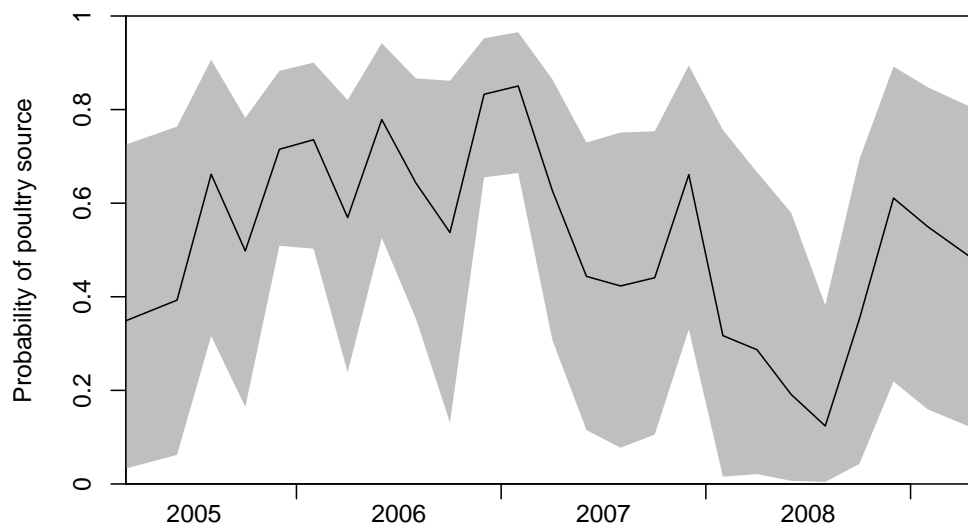


Figure 5: Attribution to poultry of human isolates by year, using the Hald model, with 95% confidence levels.

The results in Figures 2 to 4 are mean results of the attribution estimates – they do not show how much variability there is about those estimates. Given that the time periods considered are short at just two months, we have only a small number of isolates available each time period (between 7 and 52, with a mean of 24), and hence variation around these estimates are large. Figure 5 shows a 95% confidence envelope for the poultry attribution based on the Hald model. Longer time periods reduce this considerably, at the disadvantage of removing some of the interesting temporal behaviour. Figure 6 for instance, shows the attribution for each year, along with 95% confidence bars, which are much reduced.

We must also consider the results in Figures 2 to 4 in light of the total number of human isolates over time. A reduction in the proportion attributed to a source may not imply a reduction in the number of cases attributed to that

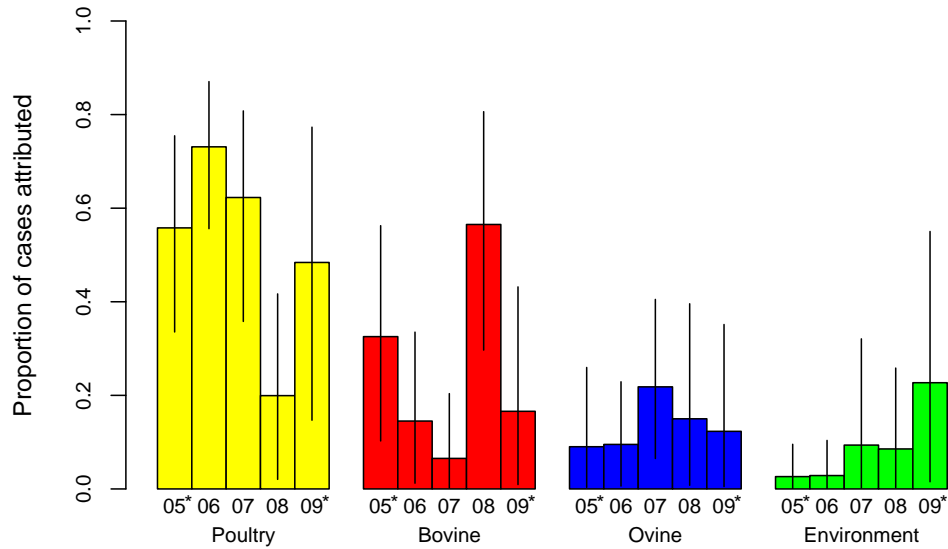


Figure 6: Source attribution with 95% confidence intervals of human isolates by year, using the Hald model. *Note that 2005 precludes January and February, and 2009 includes January through April only.

source – the proportion may be decreasing due to increases in attribution to others sources. Similarly, if the number cases attributed to one source reduces, the proportion of cases allocated to a second source may increase, even though there is no increase in the number of cases.

Assuming that any isolates that have not yet been sequence typed may be attributed to sources in the same proportions as those that have been typed, we may scale the proportions from each of our models by the number of notified cases reported to our lab that were identified as primary cases in order to estimate the expected number of cases attributable to each source. Figures 7 and 8 show these results for the modified Hald and Island models respectively. The summer peaks, most of which are attributed to poultry sources, can be clearly seen, with the peak for summer 2008/2009 appearing

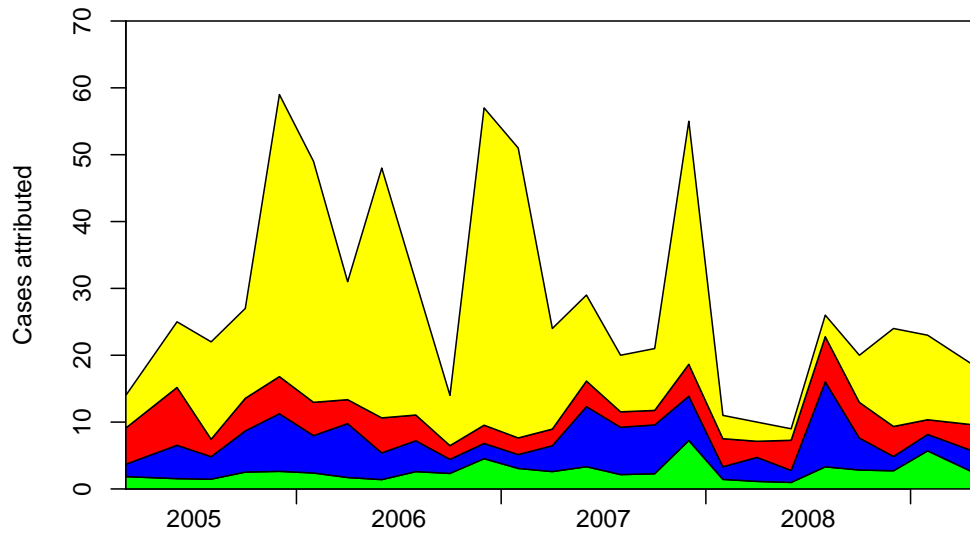


Figure 7: Estimated number of human cases per month attributed to each source from the modified Hald model.

to reach about half the level of previous years. The dramatic reduction in poultry attribution for the first half of 2008 can be seen to correspond to a reduction in isolates from all sources, although the reduction in poultry is of a higher magnitude. Ruminant strains on the whole appear relatively stable at between 10 and 20 cases per month, whilst attribution to the environmental source is between 2 and 8 cases per month.

5 Model Comparison

Each of the modelling approaches for dynamic source attribution have provided consistent estimates for the relative contribution of each of the food and environmental sources to the burden of human campylobacteriosis in the Manawatu. The temporal component of the models have clearly high-

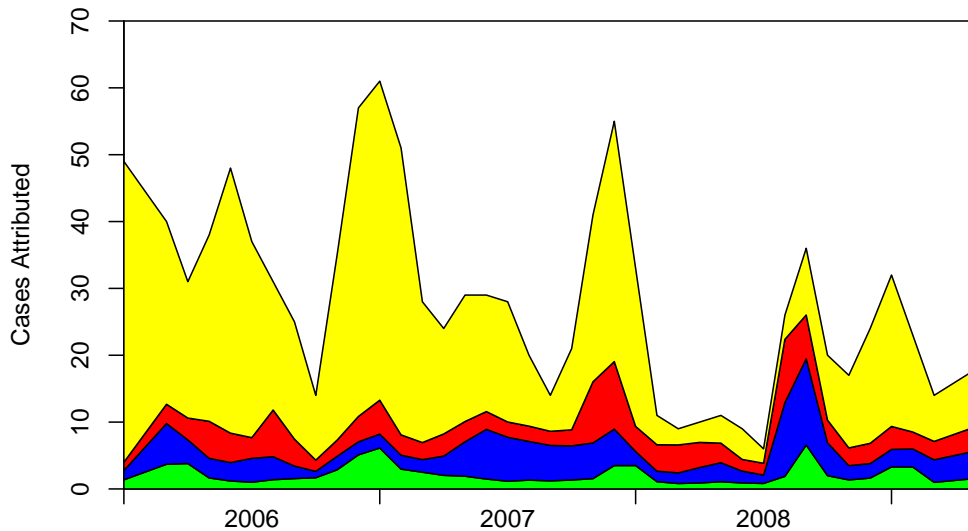


Figure 8: Estimated number of human cases per month attributed to each source from the Island model.

lighted the seasonal summer peak of campylobacteriosis notifications, which has been attributed mainly to the poultry food source.

Of the three modelling approaches, it is clear that the Modified Hald model and the dynamic version of the Island model are of more use than the combined Dutch/Island model. The latter has to over-smoothed the temporal variation, making key features such as the summer peaks in poultry less discernable. We thus focus on the Modified Hald and Dynamic Island models.

Table 3 gives a summary of these two models. The ability to assign human isolates that have yet to be observed in any of the sources is the key advantage of the Island model over the Hald model. There are 27 human isolates that have yet to be observed in the source isolates, so the absense of these isolates in the analysis by the Hald model must be considered. Furthermore, the attributions that the Island model makes based on genetic similarity, are

Model	Pros	Cons
Hald	Explicit modelling of expected counts.	Slower computation time.
	Bayesian framework allows variation about unknown parameters.	Unable to model human isolates that don't appear in the sources.
	Confidence Intervals are easy to obtain.	No genetic similarity used.
	Model modifications are easily implemented.	
Island	Faster computation time.	Confidence Intervals are hard to obtain.
	Utilises genetic similarity based on allelic profiles.	Modifying the model is difficult.
	Allows attribution of human isolates that don't appear in the sources.	

Table 3: Comparison of models

also useful in other analyses. For example, by allocating isolates based on the Island model output to a most likely source, one can then look to modelling the likelihood of getting a ruminant related strain in comparison to a poultry strain, based on covariates such as rural living, age, or contact with animals.

The running time of the two models differs significantly, with the Hald model taking approximately 10 times as long as the Island model on the 4 year dataset. Furthermore, the Island model, being essentially just multiple runs of the same model on a changing dataset, is readily amenable to parallelisation which would bring further speed ups, with the time-consuming portion of the model scaling linearly with the number of processors available. This speed advantage must be considered in the light of how regularly the model is

to be run, in comparison to how long it takes. On a standard 2GHz desktop computer, the Hald model takes around 4.5 hours to complete, so this is of no disadvantage even if running the model daily.

The advantages of the Hald model lie in its ease of extension and that estimates always come with confidence intervals. Further analysis, such as incorporating specific temporal models, and testing hypotheses are relatively simple within the Bayesian framework. The clear seasonality in the poultry attributed cases, and their contribution to the summer peak in *Campylobacter* notifications as seen in Figure 7, for instance, would suggest that a model where the poultry temporal factors had a seasonal term might be investigated. Such a model may allow the fitting of simpler temporal factors to the non-poultry sources, such as linear trends, or factors with a lower temporal resolution, allowing simple hypotheses to be tested, such as whether or not cases attributed to a particular source are increasing or decreasing over time. The inclusion of confidence intervals is essential to any further analysis – particularly when the intervals are reasonably large, as is indicated by Figure 5.

6 Recommendation

We recommend that the Modified Hald model is the most appropriate for dynamic source attribution of human cases to sources. The model allows factors associated with each source to change through time, such as adjusting for changes in food preparation practises or industry processing. The Island model should not be disregarded, however, as it is very useful at assigning particular isolates to a source, allowing categorisation of human cases for further modelling purposes. This, however, can be done without the temporal component in place.

Finally, we note that the models may be applicable to other diseases. Mullner et. al. [8] showed an application of the Hald model to *Salmonella* in

New Zealand, and the original Hald model was developed for *Salmonella* in Denmark. The key assumptions that need to hold for the model to be applicable to other diseases is that the distribution of source isolates is reasonably stable through time, and that a sufficient number of human isolates are available in order to balance the requirements of temporal resolution and model accuracy. A temporal resolution allowing an average of at least 20 isolates per time period would be an appropriate starting point.

7 Acknowledgements

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

References

- [1] G. K. Adak, J. M. Cowden, S. Nicholas, and H.S. Evan. The public-health laboratory service national case-control study of primary indigenous sporadic cases of *Campylobacter* infection. *Epidemiology and Infection*, 119:15–22, 1995.
- [2] Wilson D.J, Gabriel P, Leatherbarrow A.J, Cheesbrough J, Gee S, Bolton E, Fox A.J, Fearnhead P, Hart CA, and P. Diggle. Tracing the source of campylobacteriosis. *PLoS Genetics*, in press, 2008.
- [3] J. Eberhart-Phillips, N. Walker, N. Garrett, D. Bell, D. Sinclair, W. Rainger, and M. Bates. Campylobacteriosis in new zealand: results of a case-control study. *Journal of Epidemiology and Community Health*, 51:686–691, 1997.
- [4] I. A. Gillespie. Milkborne general outbreaks of infectious intestinal disease, england and wales, 1992-2000. *Epidemiology and Infection*, 130:461–68, 2003.
- [5] I. A. Gillespie, S. J. O’Brien, J. A. Frost, G.K. Adak, P. Horby, and A.V. Svan. A case-case comparison of *campylobacter coli* and *campylobacter jejuni* infection: A tool for generating hypotheses. *Epidemiology and Infection*, 128:111–18, 2002.
- [6] S. Greenland and K. J. Rothman. *Measure of Effect and Measures of Association*, pages 54–58. Lippincott Williams and Wilkins, 2nd edition, 1998.
- [7] T. Hald, D. Vose, H. C. Wegener, and T. Koupeev. A bayesian approach to quantify the contribution of animal-food sources to human salmonellosis. *Risk Anal*, 24(1):255–69, Feb 2004.
- [8] P. Mullner, G. Jones, A. Noble, S. Spencer, S. Hathaway, and N.P. French. Source attribution of food borne zoonosis in new zealand: A modified hald model. *Risk Anal*, page in press, 2009.

- [9] N. P. French and the Molecular Epidemiology and Veterinary Public Health Group. Enhancing surveillance of potentially foodborne enteric diseases in New Zealand: Human Campylobacteriosis in the Manawatu. http://www.nzfsa.govt.nz/science/research-projects/Campy_Attribution_Manawatu.pdf, 2008.
- [10] L. W. Riley. *Molecular Epidemiology of Infectious Diseases - Principles and Practices*. ASM Press, 2004.
- [11] O. Rosef, G. Kapperud, S. Lauwers, and B. Gondrosen. Serotyping of *Campylobacter jejuni*, *Campylobacter coli*, and *Campylobacter lariidis* from domestic and wild animals. *Appl Environ Microbiol*, 49(6):1507–10, Jun 1985.