Ministry for Primary Industries
Manatū Ahu Matua

Trialling and evaluation of epiclustR to detect potential Campylobacter outbreaks

MPI Technical Paper No: 2012/03

Prepared for the Ministry for Primary Industries
by Ruth Pirie and Kerry Sexton

ISBN No: 978-0-478-38859-6 (online)
ISSN No: 2253-3923 (online)

May 2012

Disclaimer

This report or document ("the Report") is given by the Institute of Environmental Science and Research Limited ("ESR") solely for the benefit of the Ministry of Primary Industries ("MPI"), Public Health Services Providers and other Third Party Beneficiaries as defined in the Contract between ESR and the MPI, and is strictly subject to the conditions laid out in that Contract.

Neither ESR nor any of its employees makes any warranty, express or implied, or assumes any legal liability or responsibility for use of the Report or its contents by any other person or organisation.

Requests for further copies should be directed to:

Publications Logistics Officer
Ministry for Primary Industries
PO Box 2526
WELLINGTON 6140

Email: brand@mpi.govt.nz
Telephone: 0800 00 83 33
Facsimile: 04-894 0300

This publication is also available on the Ministry for Primary Industries website at
http://www.mpi.govt.nz/news-resources/publications.aspx

Client Report

FW12027

Trialling and evaluation of epiclustR

to detect potential *Campylobacter* outbreaks

Dr Stephen On

Food Safety Programme Leader

Ruth Pirie

Project Leader

Graham Mackereth

Peer Reviewer

Disclaimer

This report or document ("the Report") is given by the Institute of Environmental Science and Research Limited ("ESR") solely for the benefit of the Ministry of Primary Industries ("MPI"), Public Health Services Providers and other Third Party Beneficiaries as defined in the Contract between ESR and the MPI, and is strictly subject to the conditions laid out in that Contract.

Neither ESR nor any of its employees makes any warranty, express or implied, or assumes any legal liability or responsibility for use of the Report or its contents by any other person or organisation.

Acknowledgements

The project would not have been possible without the contribution of Massey University staff. In particular, Jonathan Marshall prepared a version of the model for use at ESR that aligned with existing aberration reporting processes. Jonathan also developed a reporting template to allow easy access and viewing of the output and gave an excellent explanation of the model to staff at ESR. Nigel French from Massey University also provided valuable insights into the potential use of the model with *Campylobacter* outbreaks.

Esther Lim at ESR ran the model several times during the trial to ensure the model could be set up and run on another PC. Helpful comments were made by Bronwyn Morris and Shevaun Paine at ESR who reviewed the epiclustR output each week during the trial.

Table of Contents

Background

Between 2007 and 2009, the New Zealand Food Safety Authority funded Massey University to develop a statistical model to detect spatio-temporal clusters in campylobacteriosis notification data. This model (epiclustR) was applied to retrospective New Zealand campylobacteriosis notification data from 2001 to 2007. Details of the model and its ability to detect spatio-temporal clusters have recently been published (Spencer, Marshall, Pirie, Campbell, Baker, et al., 2011; Spencer, Marshall, Pirie, Campbell, & French, 2011)

One of the longer term goals for the development was that the model could be used by public health agencies to aid disease outbreak investigation.

In 2005, ESR implemented EARS (Early Aberration Reporting System), a tool developed by Centers for Disease Control (CDC), to detect increased disease incidence in notification data. This tool is used weekly to monitor notification data at District Health Board (DHB) level for ESR, Ministry of Health (MoH) and Public Health Service (PHS) staff.

Introduction

The goal of this project was to trial and evaluate the use of the Bayesian heirarchical model (epiclustR) developed by Simon Spencer and other researchers at Massey University using weekly New Zealand campylobacteriosis notification data in real time.

The objectives of the project were:

- to set up epiclustR to run at ESR on campylobacteriosis notification data
- to trial the use of the model at ESR on current notification data in real time
- to evaluate the usefulness of the model as a tool in outbreak detection at ESR and PHSs

In addition, the objectives of the trial of epiclustR were to:

1. determine in what circumstances analysis at a finer scale may be helpful,
2. assess the value of epiclustR as a complementary or replacement tool to EARS, and
3. identify any other additional benefits for outbreak detection and response activities.

Methods

The first part of the project was the packaging of epiclustR to run at ESR and development of a reporting template to display the results. This was completed by Massey University in January 2012.

Evaluation criteria were developed and agreed between MAF (now MPI) and ESR in February 2012 (see Appendix 1). This report presents the results of the evaluation according to those criteria. A copy of the installation and weekly running instructions for epiclustR are in Appendix 2.

ESR staff attended a presentation on the epiclustR model given by Jonathan Marshall from Massey University on 13 January 2012. A session was also held at the Ministry of Health in April to explain the model and the output. The explanations of the model for PHU staff were given as output was distributed to them.

The trial was run for 8 weeks covering the surveillance period from 28 January 2012 to 24 March 2012.

1. Weekly data extraction from EpiSurv (EpiSurv number, date reported, census area unit (CAU), surveillance week, address match accuracy, territorial authority (TA), district health board (DHB))
2. Processing the data though epiclustR.
3. Creation of a pdf file that summarised the expected number of notifications, the actual number of notifications and an outbreak probability for each TA for the last four surveillance weeks. This file also showed graphs for the number of notifications and the outbreak probability for each TA for the last 3 years.
4. Production of a kml file (that could be viewed in Google Earth) that displayed the number of cases for each CAU for each week over the 3 year period with the TA outbreak probability.
5. Comparing epiclustR outputs with EARS outputs.
6. Forwarding of model output to PHU and MoH staff, as required, to discuss the investigation of identified outbreaks.

The evaluation of the usefulness of model outputs included evaluation criteria divided into two groups. One group of measures covered the installation and running of the model. The other group assessed the usefulness of the model output, especially for outbreak detection, and Included an assessment of the benefits and resources needed to maintain an epiclustR implementation at ESR.

Results

Installation and running of epiclustR model

a) Installation of model

A significant amount of effort was made by Massey University staff to package the model so it could be installed and run on a regular basis within ESR's corporate information technology environment. The model ran successfully for the first time at ESR on 11 January 2012. Some development decisions were made and constraints imposed by the time available and need to provide model output that was readily available to ESR, MoH and PHU staff during the trial. It was decided to use a pdf format for the main output (instead of html pages) which in addition to the installation of R and specific libraries also required the installation of MikTex and some libraries on the PC running the model. The installation of MikTex and the additional libraries proved to be problematic on both occasions when an installation was attempted. There were difficulties with both the installation of the basic MikTex installation and the proxy settings required to install the additional libraries.

b) Data preparation

It was intended that the model would produce output at the level of territorial authority (TA) and for drinking water zones. However due to the large number of drinking water zones in New Zealand (over 700 community supplies/drinking water distribution zones), resulting in an impracticably large number of reporting areas to monitor each week, only TA level analysis was implemented during the trial.

c) Running of the model

The default and recommended parameters (see model interface in Appendix 2) appeared to be sufficient to run the model over the evaluation period.

Significant effort was made to align the output from epiclustR with EARS in terms of surveillance weeks (the surveillance week in EARS runs from Saturday to Friday). However it was still possible for someone running the model to enter the wrong date (either in the EpiSurv extract or in the interface) and end up running a model that produced output for future partial weeks.

A facility to continue if the model failed part way through the iterations process or during the report production process was provided. This was used successfully several times (during the setup phase and during the evaulation trial) to complete the model iterations and produce the reports.

Data was extracted from EpiSurv on a Monday evening at approximately 5pm and set to run overnight on a desktop PC at ESR to match output from EARS which runs on Tuesday morning at 7am. The goal was to produce output by 9am on Tuesday morning. The process ran successfully to completion on 6 out of the 8 occasions. One failure was due to operator error i.e. misspecification of dates in the EpiSurv extract file. The reason for the other failure was unknown but some unidentified process prevented epiclustR from writing to the drive. Features such as power management were disabled and virus scans and file archiving were carried out at times that did not overlap with when epiclustR was run.

A number of trial runs had been completed prior to the February/March trial and from these the main issues identified in running the model successfully were incorrect dates being included in the data file extract and the model being interrupted by another process on the PC. Some data checking was already included in the epiclustR but an automated extraction and running process would overcome most of the more minor issues.

For the evaluation epiclustR was run using CAU level EpiSurv data and TAs as the output. The standard parameters for running epiclustR were 12000 iterations with 2000 iterations burn in. The best running speed for epiclustR with these parameters on a networked PC at ESR was 14.5 iterations per minutes (the process took approximately 12 hours). This could be compared to 24.5 iterations per minute (approximately 8 hours) on a four year old (non-networked) PC and 64 per iterations (3 hours 15 minutes) on a 3 month old (non-networked) PC with a 2.70GHz Intel Core i7-2620M processor.

A significant component of running the model is multiple writes to the hard drive. Running a RAM drive utility did make some improvement to the time taken to run the model (approximately 10% improvement in speed). Writing to a network drive considerably slowed the progress of the model and writing to a local drive appeared to make a significant improvement to the running speed of the model.

For comparison the complete EARS process (running statistical analysis and outputting results) runs in less than 20 minutes. However EARS uses less geographic areas (there are only 20 DHBs) and information from the 1900 CAUs is not used. During the trial some comparisons were made between EARS and epiclustR. However as EARS runs at DHB level and epiclustR at TA level these outputs were not directly comparable.

Use of epiclustR in disease outbreak monitoring

a) EpiclustR output format

The surveillance week and run day (Tuesday) in the EARS and epiclustR outputs were aligned which made comparison between the two tools (and EpiSurv) straightforward. TA names were ordered logically from north to south in the output making it easy to track particular areas each week. Changing the parameters of the output was not attempted in the evaluation period as the focus was on understanding how epiclustR worked over a number of weeks with the same parameters.

Only the clusters with the highest outbreak probability (highlighted red in the TA summary table) aligned with the evaluator's subjective "investigation threshold".

There was some incorrect or missing explanatory information with the tables in the pdf file but this was not a major barrier to interpretation of the model output. The level of detail on the tables and figures was sufficient, although it would have been useful to document that only notifications with an exact or nearest address accuracy were included in the analysis and output as this provided some initial confusion.

Navigation from the TA summary table, down to an associated figure or list of associated cases was simple (one mouse click). As the output was in pdf format (as opposed to the web format of EARS),

returning to the initial TA summary table or elsewhere in the output required scrolling or use of page up/page down buttons which made it more cumbersome than navigating through EARS.

The entire output file was forwarded to Community and Public Health without any difficulty. TA graphs were extracted from the output and emailed successfully to both Public Health South and Community and Public Health.

The kml files were not sent to the PHUs during the evaluation period. The kml output was somewhat confusing as the outbreak probability calculated at TA level was shown at CAU level with a count of all notifications for that week. Counts for CAUs and weeks were only shown on the map where the TA had an outbreak probability of 0.75 or higher.

b) Contribution to outbreak investigation

The epiclustR evaluation period commenced after significant aberrations in campylobacteriosis notifications had already been detected through EARS (particularly for Canterbury and Southern DHBs). No potential outbreaks were identified by epiclustR that had not already been "flagged" in EARS. However, it was useful for the ESR outbreak team that epiclustR further refined the DHB-level EARS aberrations to specific TAs. We would have been able to identify the TA clustering through further exploration of the notification data but it was helpful that this was already available in epiclustR output. In practice, when contact was made with PHU staff, both Public Health South and Community and Public Health were already aware of the TA-level clusters.

The kml file map output was not accessed by ESR during the evaluation period. Community and Public Health produced and distributed their own map of the campylobacteriosis cases from Christchurch City TA. This showed that the cases were widely distributed. Further exploration by viewing the kml output in Google Earth from epiclustR was considered unlikely to provide any additional intelligence. The Clutha TA cluster was not viewed via Google Earth either.

Maps incorporating drinking water zones were discussed during the set up phase of the evaluation but were not implemented for the evaluation phase. If drinking water zone maps had been available, it is likely that these would have been viewed as part of the evaluation of the detected clusters. The potential utility of a map output incorporating drinking water zones remains unknown.

Discussion

Installation and running of the model

A significant portion of the project time was spent setting up the model to run successfully in ESR's IT environment.

The model was set up during the trial to run and produce output for TA areas. Further thought is needed as to how drinking water distribution zones can be used as an epiclustR output area. The large number of zones could be reduced by combining zones to indicate water treatment instead of water distribution or only including water supplies that serve a large population.

Running the model each week led to a few problems with specifying parameters correctly. Ideally a stored procedure would run the model each week (including an ODBC connection to extract the data from EpiSurv) which allows less room for user error.

Time taken to run the model at ESR remains an area for improvement. The running of the model at ESR could be significantly improved by running the model on a local drive instead of across a network drive and the use of a PC with a fast processor.

The model could be run for less than 12 000 iterations. The probability estimates will be less stable but the trade-off of time may be worthwhile. A comparison between running 12 000 and 1200 iterations was made and while there were differences in the probabilities produced these differences were unlikely to result in markedly different public health action.

One of the other options discussed to improve the time taken to run the model was setting up the model so it ran just the last few weeks instead of the whole time period every week. However, increasing complexity would provide more chance for error and present greater difficulty in resolving errors when running the model. So this would not be a high priority development.

Use of epiclustR in disease outbreak monitoring

The output format worked well for the epiclustR trial. The trial primarily involved ESR's review of the model output and further work would be required to assess whether the display of the model output is appropriate for PHS and Ministry of Health users.

From ESR's perspective epiclustR added little value to the disease outbreak investigation process for campylobacteriosis during the evaluation period. This may have been due to aberrations in DHBs having already been detected through the EARS system when the evaluation period started. Clusters were not identified in any additional DHBs.

There would be benefit in evaluating epiclustR over a different time period to see whether clusters were identified in addition to those in EARS, or at an earlier stage. In addition it would be useful to run EARS at TA level to assess whether similar clusters were identified and which method is most appropriate. The only benefit from running epiclustR during the evaluation period was that TA-level clusters were easier for ESR staff to identify.

Extracts from the pdf file or the entire pdf file from the epiclustR weekly report were forwarded to PHUs during the evaluation. The PHUs were already aware of the TA-level increases. There was a misunderstanding of the epiclustR output by Community and Public Health. Cases listed in two consecutive weeks were interpreted as relating to separate outbreaks and all cases listed in each cluster were assumed to be part of an outbreak cohort. The use of the word cluster with epiclustR is somewhat misleading and confusing if PHU staff are familiar with other cluster detection tools such as SatScan where space time clustering is identified and reported in a similar manner.

The evaluation is primarily focussed on ESR's evaluation of epiclustR. The two month period was too short for PHU and the Ministry of Health staff to appreciate the methods that epiclustR was using identifying outbreaks. A much longer period (e.g. six months) would be required to really understand what the model does and how it might be useful.

Even with both the EARS and epiclustR tools showing potential outbreaks of high probability, it was difficult to inspire action within PHUs for investigation beyond routine practice. The exception being that the epiclustR output did motivate Community and Public Health to phone interview every case flagged in epiclustR for two consecutive weeks.

The visualisation and interpretation of the probabilities, especially for smaller TAs, and prioritising of outbreaks for investigation needs some further work. During the trial the focus for further investigation was on TAs and weeks with very high probability (0.95 or higher) that were coloured red or orange. For this reason Christchurch and Dunedin cities attracted the most attention. However there was an elevated probability (0.75) in Hastings District and Marlborough District over several weeks that may have also warranted attention. It was likely overlooked as it not the highest probability and was coloured with a light cream colour. As the numbers reported each week were small (1-2 expected notifications and 4-5 cases notified) there was likely to be insufficient power to produce a higher probability.

The value in identifying outbreaks is difficult to realise if the PHUs assign a low priority or have little interest in investigation of the identified clusters. Staff at CPH were interested in isolate typing to assist with investigation. PHUs other than CPH were generally disinterested in carrying out any further investigation. Protocols for forwarding of *Camplyobacter* isolates from clinical laboratories to ESR and typing of these isolates do not currently exist making it difficult to initiate typing of isolates during an outbreak.

The process of reviewing epiclustR data on a weekly basis is likely to take very little time (less than half an hour) for PHU staff once they are familiar with the epiclustR output but the potential impact of identifying additional clusters, and the time required to investigate these clusters, will not be known until a longer trial is undertaken.

A similar amount of time for the review process would be required at ESR. Again, until a longer review period for epiclustR is undertaken, it will not be known whether epiclustR will identify clusters additional to those identified by EARS and the impact of this on staff time at ESR. Some resources will be also required to maintain epiclustR at ESR e.g. implement changes to census data, troubleshoot model interruptions, reinstallation of model following PC upgrades/server changes etc. A rough estimate of this would be 40 hours a year.

From the trial it appears epiclustR running on a weekly cycle is most suited to an endemic disease such as campylobacteriosis with a regular significant number of sporadic cases reported as these historical cases determine the probability of an outbreak. EpiclustR may add some value to identifying increased incidence with other non-enteric diseases. However for most non-enteric notifiable diseases the numbers are much smaller and will likely run into the lack of statistical power issues mentioned above. Some changes could be made to the epiclustR model parameters to better suit the low number of cases e.g. changes to output areas such as using DHB for output and use of an appropriate time period such as months. Other non-enteric notifiable diseases with a large number of notifications e.g. pertussis have a more cyclical pattern of epidemics and are therefore unsuitable for epiclustR approach which relies on a more consistent historical sporadic pattern of notification.

Conclusions

Installation and running of the model

During the trial most of the effort went into getting the model up and running at ESR. The model is now available at ESR and can be run if required for further evaluation. A number of areas were identified for further improvement of the model. Further evaluation work is required to optimise the parameters (e.g. number of iterations) for running epiclustR and a trial of EARS at TA level alongside epiclustR is needed for a direct comparison. More work is also required to improve the time taken to run the model and produce reports.

Use of epiclustR in disease outbreak monitoring

The model trial period was short but sufficient to identify a number of areas that impact on the potential use of epiclustR. It took most of the trial period and examination of the model output over multiple weeks before ESR staff had a reasonable understanding of the output.

A much longer period of time, as long as six months, is needed for public health staff in ESR, PHUs and MoH to really understand how to correctly interpret the output from the model and evaluate the real value of epiclustR in identifying *Campylobacter* outbreaks. The development of a training module would help public health staff interpret the model output from epiclustR.

The policy for investigation of potential campylobacteriosis outbreaks (including when an investigation should be triggered within a TA and the use of typing to identify common sources) needs development. Once this is in place a further trial over a longer period would be worthwhile carrying out to assess the value of epiclustR for the identification and investigation of *Campylobacter* outbreaks. The estimated resourcing for this trial would be one day a week for one staff member at ESR along with some support from Massey University (estimated 2-3 days work) to make some small changes to the epiclustR output format.

Appendix 1. Evaluation Criteria

MAF Project Title: Trialling and evaluation of epiclustR to detect potential campylobacter outbreaks (MRP11/06)

Final evaluation criteria for epiclustR

The contract stated that *"Evaluation criteria will be developed to assess the success of the implementation and use of the model at ESR and how the model may be used with, or if it could replace, existing aberration detection tools such as EARS (Early Aberration Reporting System)."*

The first phase of the project was the installation of epiclustR at ESR to run weekly with campylobacteriosis notification data and the development of the reporting template.
A working version of epiclustR is now running and the next phase will focus on

a) Evaluation of functionality currently present in epiclustR, the ideal future state and how much further investment (by ESR/Ministry of Health) might be required and the priorities for further development.
b) High level assessment of potential value of running epiclustR for campylobacteriosis outbreak monitoring. This will be assessed within ESR, with Ministry of Health staff and with PHU staff.

Tables 1 and 2 summarise the aspects of epiclustR that will be evaluated following the trial period.

Table 1. Evaluation criteria for epiclustR installation and running of model

Item	Ideal state	Evaluation Measure
Installation	Model can be installed and run in a corporate business environment e.g. at ESR	Description of installation requirements and success.
Data preparation	The steps for preparation of the data for running the model are documented and have been carried out at least twice on two different datasets.	Description of steps involved in converting cases, area and population data into required format for use in model Comment on level of skill required to prepare data.
Running model	Interface allows setting parameters for time period, input file locations, output file locations, monitoring of model progression.	Description of any further development required for parameter setting
	Parameters set by user are checked for misspecified parameters e.g. incorrect dates, error diagnostics	Description of any further development required for detection of misspecification errors
	Facility to continue model running if interruption occurs and process stops.	Description of any further development required for continuing model
	Model will process the required iterations and complete processing the data successfully each time it is run.	Number of model runs that successfully complete. Description of the reason(s) for any model run interruption or failure where known.
	The model will run, without interruption, and produce reports in less than 4 hours.	Describe computation time for model with different parameters and hardware configuration. Identify potential improvements and quantify resources required.
	Model can be run overnight on Monday evening so that results are available at 9am on Tuesday morning	Number of weeks that results are available by 9am on Tuesday. Description of any issues.

Table 2. Evaluation criteria for use of epiclustR in disease outbreak monitoring

Item	Ideal state	Evaluation Measure
Output format	Model results are output in a format that a) can be matched to the surveillance week used in other processes e.g. EARS b) output regions e.g. TAs can be ordered in tables, maps etc c) users can select dates, probabilities, cluster sizes to control model results output d) individual TA maps/graphs/tables can be extracted and exported to PHUs/MoH etc as required e) graphs/maps/tables have appropriate level of detail to navigate information quickly and easily f) size of output files is reasonable for distribution via email. g) PHU staff can view map output through a readily available mapping interface.	Qualitative assessment of how well functionality meets needs of users.
Usefulness	Model results contribute to identification of potential outbreaks	Description of how epiclustR results contributed to the identification of potential outbreaks.
	epiclustR identifies outbreaks that are also identified through EARS or identifies additional outbreaks.	What proportion of "potential" outbreaks detected through EARS were also detected through EpiclustR? How many "potential" outbreaks were detected through EpiclustR and not EARS? How many "potential" DHB-level outbreaks identified through EARS were refined as TA-level outbreaks through EpiclustR?
	Map output contributes to interpretation of potential outbreaks	How often was the Google Map presentation of the epiclustR data accessed by ESR surveillance data staff and what value did it add to the investigation process?

Appendix 2. Installation instructions for EpiclustR at ESR

1. Install R (at least version 2.12.2) with maptools, spatstat, RColorBrewer, xtables, classInt and spdep libraries from one of the CRAN mirrors linked from here http://www.r-project.org/

2. Install .NET framework version 4 from http://www.microsoft.com/download/en/details.aspx?id=17851

3. Install Microsoft visual studio files from http://www.microsoft.com/download/en/details.aspx?id=5555.

4. Install MikTex from here http://miktex.org/2.9/setup. Choose the basic setup and then you can install packages required later.

5. Optionally install RAMdisk software (e.g. http://memory.dataram.com/products-and-services/software/ramdisk) to allow writing to a RAM partition instead of the PC hard drive. Allocate 500 MB RAM for this.

6. Copy the latest release of EpiclustR (epiclustR_20120201) from the **\epiclustR\Releases** folder on Dropbox.

Running the model each week

1. Run EpiSurv report "epiclustR extract" with dates as below (choose appropriate week end date)

Start	End
week beginning *Saturday*	week ending *Friday*
31/01/2009	3/02/2012
7/02/2009	10/02/2012
14/02/2009	17/02/2012
21/02/2009	24/02/2012
28/02/2009	2/03/2012
7/03/2009	9/03/2012
14/03/2009	16/03/2012
21/03/2009	23/03/2012

3. Export csv file and name as yyyymmdd with week ending date

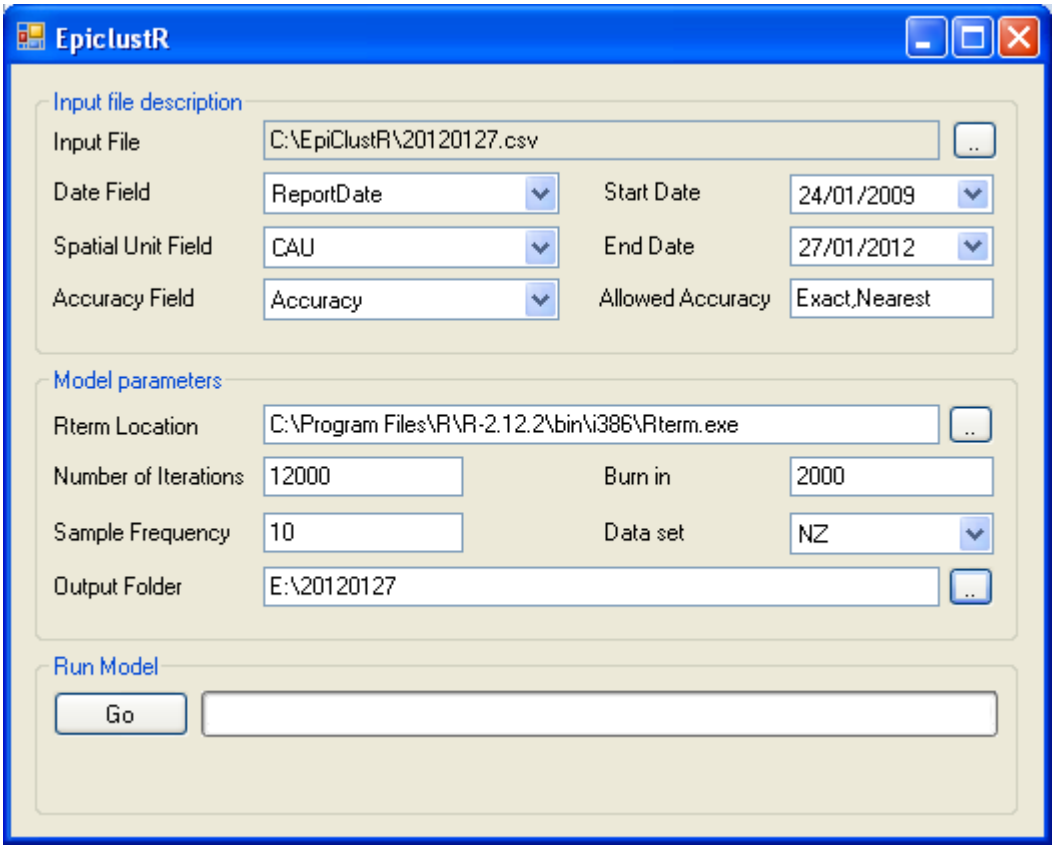
4. Open csv file in Excel and separate data into columns based on commas (Data, Text to columns, Delimited, Commas)

5. Save as csv (comma delimited) file

6. Create a folder for the output using week ending date as in step 3.

7. Open EpiclustR.exe

8. Enter values as per example screenshot below.



9. Click go. Leave to run for 12 hours.

Note: When the model runs it makes a copy of all the R scripts into your output folder (i.e. the one you named yyyymmdd). This is where all the parameters for your model run are kept.

10. If the model ran successfully you will see the files **outbreaks.pdf** and **outbreaks.kml** in your output folder (i.e. yyyymmdd\NZ\RUX2_region).

11. If the model did not produce these files check the r_output.txt file in your yyyymmdd output folder and see instructions for troubleshooting below.

Troubleshooting

If the model did not complete successfully try the following (requires some R knowledge)

1. Read the **r_output.txt** log file to identify where the model failed.

2. Load up R and load in the **epiclustR.R** file (File->Source R Code... menu) from the yyyymmdd folder. You will see a mini menu as below

```
> source("C:\\EpiClustR\\20120127\\epiclustR.R")
Welcome to EpiclustR
====================

Model run from 2xxx to 2xxx

This is the simple R interface for re-running portions of the model.

Type continue() to continue a model that didn't finish all iterations
Type analysis() to run the analysis stage of the model
Type full_model() to run the full model from scratch (overriding any progress)
>
```

3. Type **analysis()** for R to run the analysis part again or **continue()** to continue the model iterations if it did not reach 12000 iterations. The analysis alone takes about 10-15 minutes to run (depending on processor speed and file writing to disk speed).

Appendix 3. Sample pages from epiclustR pdf output

EpiclustR run on cases from 2009-03-14 to 2012-03-16

12000 iterations run with a burn in of 2000 and sample frequency of 10

TA	DHB	8, 2012	9, 2012	10, 2012	11, 2012
Far North District	Northland	2 (0.9)	0 (0.9)	1 (0.8)	1 (0.7)
Whangarei District	Northland	4 (1.7)	1 (1.6)	3 (1.5)	2 (1.3)
Kaipara District	Northland	0 (0.4)	0 (0.4)	1 (0.4)	1 (0.3)
Rodney District	Waitemata	5 (3.1)	3 (2.9)	1 (2.8)	1 (2.5)
North Shore City	Waitemata	9 (6.9)	4 (6.6)	4 (6.2)	4 (5.5)
Waitakere City	Waitemata	7 (5.1)	1 (4.8)	4 (4.5)	1 (4.0)
Auckland City	Auckland	9 (11.3)	6 (10.7)	8 (10.1)	9 (9.0)
Manukau City	Counties Manukau	7 (6.9)	6 (6.5)	7 (6.1)	8 (5.4)
Papakura District	Counties Manukau	1 (1.2)	3 (1.2)	0 (1.1)	1 (1.0)
Franklin District	Counties Manukau	1 (2.1)	2 (2.0)	1 (1.8)	1 (1.6)
Thames-Coromandel District	Waikato	1 (0.7)	0 (0.7)	1 (0.6)	1 (0.6)
Hauraki District	Waikato	2 (0.5)	1 (0.5)	0 (0.4)	1 (0.4)
Waikato District	Waikato	3 (1.6)	3 (1.5)	2 (1.4)	1 (1.2)
Matamata-Piako District	Waikato	0 (1.0)	1 (1.0)	1 (0.9)	0 (0.8)
Hamilton City	Waikato	3 (4.2)	2 (3.9)	3 (3.7)	2 (3.3)
Waipa District	Waikato	3 (1.6)	1 (1.5)	1 (1.4)	2 (1.3)
Otorohanga District	Waikato	1 (0.3)	1 (0.3)	0 (0.3)	0 (0.3)
South Waikato District	Waikato	0 (0.7)	0 (0.7)	0 (0.6)	0 (0.5)
Waitomo District	Waikato	0 (0.3)	0 (0.3)	0 (0.2)	0 (0.2)
Ruapehu District	Waikato	0 (0.3)	0 (0.3)	1 (0.3)	0 (0.2)
Taupo District	Lakes	0 (0.9)	0 (0.9)	2 (0.8)	1 (0.7)
Rotorua District	Lakes	0 (2.0)	1 (1.9)	2 (1.7)	0 (1.5)
Western Bay of Plenty District	Bay of Plenty	1 (1.2)	0 (1.2)	4 (1.1)	2 (1.0)
Tauranga City	Bay of Plenty	3 (2.8)	6 (2.6)	3 (2.5)	1 (2.2)
Whakatane District	Bay of Plenty	0 (1.0)	1 (0.9)	0 (0.9)	0 (0.8)
Kawerau District	Bay of Plenty	0 (0.1)	0 (0.1)	0 (0.1)	0 (0.1)
Opotiki District	Bay of Plenty	0 (0.2)	1 (0.1)	0 (0.1)	0 (0.1)
Gisborne District	Tairawhiti	0 (0.8)	0 (0.8)	1 (0.7)	0 (0.6)
New Plymouth District	Taranaki	5 (2.2)	3 (2.1)	1 (1.9)	0 (1.7)
Stratford District	Taranaki	0 (0.3)	0 (0.3)	1 (0.3)	0 (0.3)
South Taranaki District	Taranaki	0 (0.9)	0 (0.8)	1 (0.8)	2 (0.7)
Wairoa District	Hawkes Bay	0 (0.2)	0 (0.2)	0 (0.2)	2 (0.2)
Hastings District	Hawkes Bay	5 (2.4)	5 (2.3)	5 (2.2)	2 (1.9)
Napier City	Hawkes Bay	0 (2.1)	2 (2.0)	3 (1.8)	2 (1.6)
Central Hawkes Bay District	Hawkes Bay	1 (0.6)	1 (0.6)	0 (0.5)	1 (0.5)
Wanganui District	Whanganui	2 (1.1)	2 (1.0)	0 (0.9)	0 (0.8)
Rangitikei District	Whanganui	1 (0.4)	1 (0.4)	0 (0.4)	0 (0.4)
Manawatu District	Midcentral	1 (0.8)	0 (0.7)	0 (0.7)	0 (0.6)
Palmerston North City	Midcentral	3 (1.7)	3 (1.7)	2 (1.5)	0 (1.4)
Tararua District	Midcentral	2 (0.5)	0 (0.5)	0 (0.4)	0 (0.4)
Horowhenua District	Midcentral	2 (0.9)	0 (0.9)	0 (0.8)	0 (0.7)
Kapiti Coast District	Capital and Coast	0 (1.7)	1 (1.6)	2 (1.5)	0 (1.3)
Upper Hutt City	Hutt	1 (1.6)	1 (1.5)	1 (1.4)	0 (1.3)
Lower Hutt City	Hutt	3 (3.7)	1 (3.6)	2 (3.3)	2 (3.0)
Porirua City	Capital and Coast	0 (1.6)	2 (1.6)	1 (1.5)	1 (1.3)
Wellington City	Capital and Coast	13 (7.6)	6 (7.2)	2 (6.7)	1 (6.0)
Masterton District	Wairarapa	2 (0.6)	0 (0.6)	1 (0.6)	0 (0.5)
Carterton District	Wairarapa	0 (0.3)	0 (0.2)	1 (0.2)	1 (0.2)
South Wairarapa District	Wairarapa	0 (0.3)	1 (0.3)	0 (0.3)	1 (0.3)
Tasman District	Nelson Marlborough	1 (0.9)	1 (0.8)	1 (0.8)	0 (0.7)
Nelson City	Nelson Marlborough	0 (0.9)	4 (0.8)	2 (0.8)	0 (0.7)
Marlborough District	Nelson Marlborough	2 (1.1)	5 (1.1)	5 (1.0)	4 (0.9)
Buller District	West Coast	0 (0.3)	0 (0.2)	0 (0.2)	0 (0.2)
Grey District	West Coast	0 (0.3)	1 (0.3)	0 (0.3)	0 (0.3)
Westland District	West Coast	0 (0.3)	0 (0.3)	0 (0.3)	0 (0.2)
Kaikoura District	Canterbury	0 (0.1)	1 (0.1)	0 (0.1)	0 (0.1)

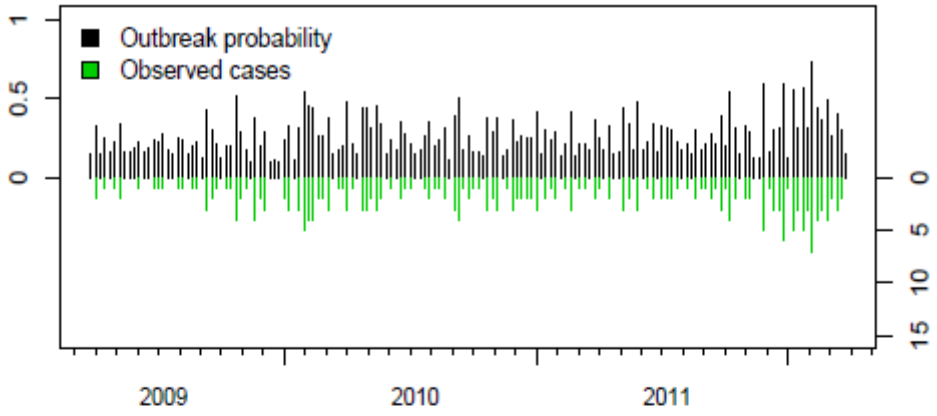
1

Hurunui District	Canterbury	0 (0.4)	1 (0.4)	1 (0.4)	2 (0.3)
Waimakariri District	Canterbury	2 (1.4)	3 (1.3)	2 (1.3)	0 (1.1)
Christchurch City	Canterbury	19 (8.3)	15 (7.9)	17 (7.4)	13 (6.6)
Selwyn District	Canterbury	3 (1.3)	3 (1.3)	1 (1.2)	0 (1.1)
Ashburton District	Canterbury	3 (1.1)	2 (1.1)	1 (1.0)	2 (0.9)
Timaru District	South Canterbury	5 (1.6)	1 (1.5)	5 (1.4)	3 (1.2)
Mackenzie District	South Canterbury	1 (0.2)	0 (0.1)	0 (0.1)	0 (0.1)
Waimate District	South Canterbury	1 (0.3)	0 (0.3)	2 (0.3)	1 (0.3)
Waitaki District	Southern	1 (0.7)	1 (0.6)	3 (0.6)	0 (0.5)
Central Otago District	Southern	1 (0.6)	0 (0.5)	0 (0.5)	0 (0.5)
Queenstown-Lakes District	Southern	1 (0.8)	1 (0.8)	4 (0.7)	1 (0.7)
Dunedin City	Southern	2 (3.1)	4 (3.0)	12 (2.8)	6 (2.5)
Clutha District	Southern	0 (0.8)	0 (0.8)	4 (0.7)	3 (0.6)
Southland District	Southern	1 (1.1)	2 (1.0)	4 (1.0)	0 (0.9)
Gore District	Southern	0 (0.5)	1 (0.5)	0 (0.4)	1 (0.4)
Invercargill City	Southern	1 (1.1)	1 (1.0)	0 (1.0)	0 (0.9)

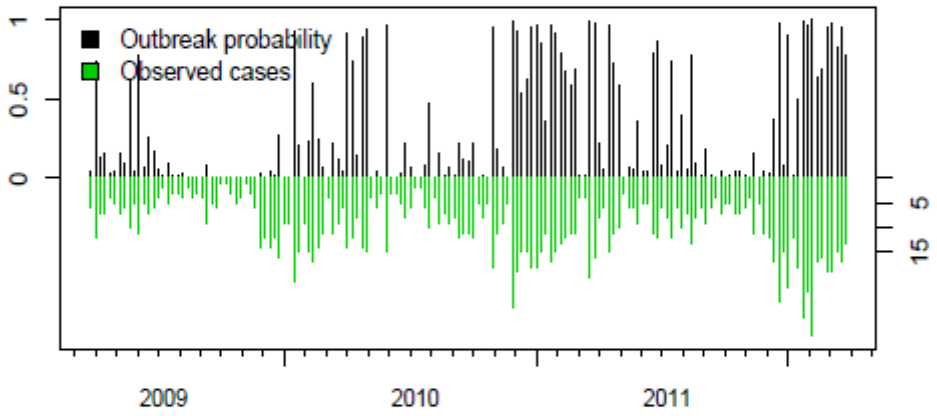
Table 1: Number of observed (and expected) cases for the 4 week period ending 2012-03-16 (week 11, 2012) by Territorial Authority and District Health Board. Outbreak probabilities are shaded red (87th percentile), dark orange (98th), light orange (95th) and cream (75th).

Region	Week	Probability	Cases
Christchurch City, Canterbury	8, 2012	0.975	12-[redacted]-CH
			12-[redacted]-CH
			12-[redacted]-CH
			12-[redacted]-CH
			12-[redacted]-CH
			12-[redacted]-CH
			12-[redacted]-CH
			12-[redacted]-CH
			12-[redacted]-CH
			12-[redacted]-CH
			12-[redacted]-CH
			12-[redacted]-CH
			12-[redacted]-CH
			12-[redacted]-CH
			12-[redacted]-CH
			12-[redacted]-CH
			12-[redacted]-CH
			12-[redacted]-CH
Dunedin City, Southern	10, 2012	0.971	12-[redacted]-DN
			12-[redacted]-DN
			12-[redacted]-DN
			12-[redacted]-DN
			12-[redacted]-DN
			12-[redacted]-DN
			12-[redacted]-DN
			12-[redacted]-DN
			12-[redacted]-DN
			12-[redacted]-DN
			12-[redacted]-DN
			12-[redacted]-DN
Christchurch City, Canterbury	10, 2012	0.961	12-[redacted]-CH
			12-[redacted]-CH
			12-[redacted]-CH
			12-[redacted]-CH
			12-[redacted]-CH
			12-[redacted]-CH
			12-[redacted]-CH

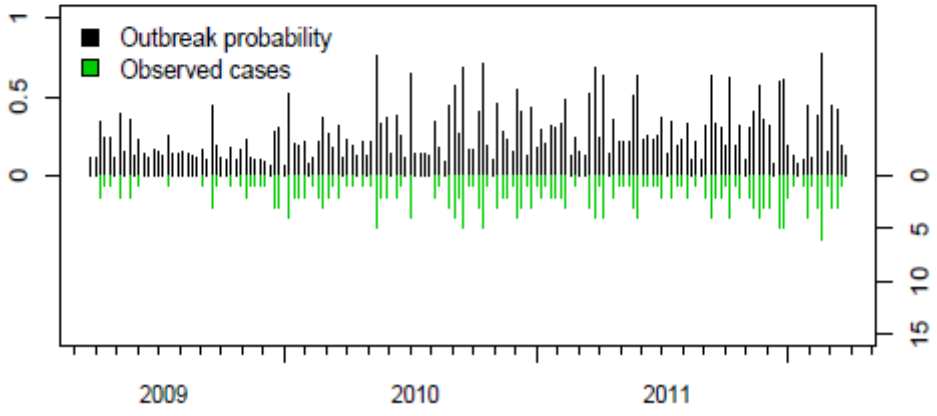
Waimakariri District, Canterbury



Christchurch City, Canterbury



Selwyn District, Canterbury



28

References

Spencer, S. E. F., Marshall, J., Pirie, R., Campbell, D., Baker, M. G., & French, N. P. (2011). The spatial and temporal determinants of campylobacteriosis notifications in New Zealand, 2001-2007. *Epidemiology and Infection*. doi: 10.1017/S0950268811002159

Spencer, S. E. F., Marshall, J., Pirie, R., Campbell, D., & French, N. P. (2011). The detection of spatially localised outbreaks in campylobacteriosis notification data. *Spatial and Spatio-temporal Epidemiology, 2*(3), 173-183. doi: 10.1016/j.sste.2011.07.008