Extrapolating incomplete animal population and surveillance data for use in national disease control: Examples from Myanmar and New Zealand

A thesis presented in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

in

Veterinary Epidemiology

School of Veterinary Science

at Massey University, Manawatu, New Zealand.

Mary van Beest van Andel

2019

Abstract

National level databases of animal numbers, locations, and movements provide the essential foundations for disease outbreak investigations, disease control, and disease preparedness activities. These activities are particularly important for managing and mitigating the risks of high impact exotic disease outbreaks like foot-and-mouth disease (FMD) as well as other economically important endemic diseases, which can significantly impact international trade and food security. However, many countries worldwide either lack national animal databases entirely or have multiple, fragmented databases that provide an incomplete picture of animal demographics. Consequently, there has been growing interest in developing novel methods to infer missing information on animal populations from other data sources, to quantify the extent of missing information, and to understand the impacts of missing information on the predictions made from national disease simulation models. This thesis explores these issues in the context of an FMD free country (New Zealand) as well as a country with endemic FMD (Myanmar).

In Chapter 3, regression models were used to predict farm-level animal populations in New Zealand based on available data on farm type and location. When the results were compared against a subset of validated animal population data, the predictions at the farm level were found to be inaccurate especially for small-scale farms that keep animals for personal consumption or as a hobby. These properties are of particular interest to animal health authorities as they have been identified as at risk for exotic disease outbreaks. In Chapter 4, the impacts of having inaccurate herd size data on the predictions made by an FMD disease spread simulation model were explored. The results were analysed using cox proportional hazard models and logistic regression models, which showed that simulations run using actual animal population data indicated different optimal control strategies for FMD than models run with imperfect data and these effects differed by the region in New Zealand where the hypothetical disease outbreak was seeded.

In Chapter 5, high-resolution local survey data and low-resolution national remote sensor data were used alone and in combination to predict the location of FMD positive villages in Myanmar, which were identified by serological sampling conducted as part of a large OIE funded research project in 2016. The performance of both random forest models and logistic regression models were explored using training and testing data sets. Bovine populations and proximity to cattle markets were found to be significant risk factors for FMD seropositivity and the logistic regression models performed as well as machine learning techniques. Chapter 6 compared verbal reports of FMD outbreaks from

village headman and householders against the serological test results from their villages to determine whether using public reports is a viable alternative to conducting resource intensive serological surveys for estimating FMD prevalence in Myanmar. Although village headmen proved to be a better source of FMD reports compared to householders, the verbal reports were still not as accurate as serological tests in an endemic situation where both sensitivity and specificity of observing clinical signs can be complicated by endemic stability and concurrent outbreaks of other diseases.

The work in both chapters 5 and 6 was carried out using data from activities of the Livestock Breeding and Veterinary Department and the OIE and as such separate human ethics approval was not required for the surveys described.

Chapter 7 addressed the issue of estimating the scale of missing data in a national database by comparing intensively collected interview information with recorded movements at the farm level for farms involved in New Zealand's *Mycoplasma bovis* eradication programme. The results showed that dairy farmers often failed to record almost half of high risk movements including leased bulls, calves sent offsite for rearing, and adult cattle sent away for winter grazing. It was also estimated that approximately 60% of animals arriving at abattoirs in New Zealand have multiple movements missing from their life history in the National Animal Identification and Tracing system (NAIT) database. This missing information had a significant impact on the ability of government and industry to effectively respond to the outbreak. However, a positive finding from this study was that the rates of missing data are decreasing over time.

Overall, this thesis demonstrated the importance of enhancing efforts to collect accurate and up-to-date national animal population and movement data. For New Zealand, the changes required to improve the national farm animal data landscape include improving compliance with the legislated requirements to record animal movements and modifying the existing databases to record information on the health status of animals against a unique animal identifier. A unique farm identifier is required at the national level and should be agreed upon by industry representatives, government and researchers. The combination of animal health data associated with the unique animal identifier and a single current farm identifier for all farms will result in a robust animal health and biosecurity system.

Acknowledgements

This work would not have been possible without the support of the Ministry for Primary Industries. In particular, Dr Paul Bingham and Veronica Herrera who gave me the opportunity to pursue my goal of working toward a PhD. Importantly, CEBRA (the Centre of Excellence for Biosecurity and Risk Analysis) started me on my journey and I am particularly indebted to Andrew Robinson, Mark Burgman, Richard Bradhurst and Tracey Hollings from that team. Christine Reed at MPI supported my project and I'm forever grateful for her trust right at the beginning of the journey. I had many formal supervisors on the journey and many mentors who helped me along different parts of the road. Tim Carpenter, Chris Jewell, Daan Vink, Robert Sanson, Geoff Jones, Karyn Froud, Roger Morris, Cord Heuer and Graeme Garner all pitched in along the way in different capacities. Jo McKenzie was a rock and provided moral support when all seemed to be too much. And, finally but by no measure least, Carolyn Gates was a mentor, collaborator, partner in frustration, and effective and efficient supervisor.

I was lucky enough to collaborate with the team from Myanmar Livestock Breeding and Veterinary Department (LBVD) in the course of my work and with their regional OIE supporting team. Dr Ronello Abilla, Dr Scott Zaari and Dr Ian Dacre all should be acknowledged as should each person in LBVD who worked on data amalgamation and generously providing essential local knowledge.

A lot changed in the time it took me to complete this work. I lost my father and had a baby boy. Some things did not change; my wonderful husband remained wonderful. He bore my preoccupations well and was generous enough to express an interest in my doings that extended beyond asking when I'd be finished. My lovely little son bore my preoccupations less well and contributed to some of the delays to finishing. I adore you both and thank you for your understanding. You are everything to me. My father passed away while I was busy with this thesis. He gifted me a thirst for learning that has been unquenchable – I dedicate these efforts to him.

> *“Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise.”*

John Tukey The future of data analysis. Annals of Mathematical Statistics 33 (1), (1962), page 13.

Table of contents

Abstract 2

Acknowledgements 5

Table of contents 7

Table of Tables 10

Table of Figures 14

Abbreviations 16

Publications 17

1 Introduction 19

2 Review: Using national animal datasets to support transboundary animal disease control in developed and developing countries 24

- 2.1 Abstract 24
- 2.2 Introduction 24
- 2.3 Preparedness activities in disease free countries 26
  - 2.3.1 Surveillance to maintain disease freedom status 26
  - 2.3.2 Identifying risk pathways for disease entry 26
  - 2.3.3 Predicting spread post-introduction 27
  - 2.3.4 Impacts of incomplete data on policy decisions based on spread modelling 29
- 2.4 Response activities during an acute response in a disease free country 30
  - 2.4.1 Contact tracing 30
  - 2.4.2 Epidemiological investigations 31
  - 2.4.3 Evaluating control strategies 32
- 2.5 Eradication activities in FMD endemic countries 34
  - 2.5.1 Progressive control pathway framework 34
- 2.6 Data requirements for FMD preparedness 35
- 2.7 Discussion 38

2.8 Conclusions 39

3 Predicting farm-level animal populations using environmental and socioeconomic variables ... 41

3.1 Abstract 41

3.2 Introduction 42

3.3 Materials and Methods 44

3.4 Results 53

3.5 Discussion 63

3.6 Conclusions 67

3.7 Appendix to Materials and Methods 68

3.8 Acknowledgements 75

4 Does size matter to models? Exploring the effect of herd size on outputs of a herd-level disease spread simulator 76

4.1 Abstract 76

4.2 Introduction 77

4.3 Materials and Methods 78

4.4 Results 88

4.5 Discussion 92

4.6 Appendix 97

4.7 Acknowledgements 99

5 Evaluating the utility of national-scale data to estimate the local risk of foot-and-mouth disease in endemic regions 100

5.1 Abstract 100

5.2 Introduction 101

5.3 Materials and Methods 103

5.4 Results 112

5.5 Discussion 120

5.6 Acknowledgements 122

6 Estimating foot-and-mouth disease (FMD) prevalence in central Myanmar: comparison of village headman and farmer disease reports with serological findings 123
  6.1 Abstract 123
  6.2 Introduction 124
  6.3 Materials and Methods 126
    6.3.1 Survey design, respondents and serology: 126
    6.3.2 Data analysis 128
  6.4 Results: 130
  6.5 Discussion: 142
  6.6 Acknowledgements 145
7 Integrating multiple data sources to evaluate the quality and completeness of national animal identification and traceability systems 146
  7.1 Abstract: 146
  7.2 Introduction: 147
  7.3 Materials and Methods: 149
  7.4 Results: 153
  7.5 Discussion: 160
  7.6 Acknowledgements 163
8 Conclusion 164
9 Appendix to Chapter 6: Latent Class Model Winbugs Code 172
10 Bibliography 177

Table of Tables

Table 1: *Species equivalence values used to calculate livestock units (LSUs) on rural properties in New Zealand following the European Union definition of individual species feed requirements (Anonymous, 2013b).* 45

Table 2: *Predictor variables included in a Bayesian Zero Inflated Poisson model predicting counts of cattle and counts of livestock units on New Zealand farms.* 46

Table 3: *Mean and standard deviation of rank probability scores (RPS) for models predicting livestock units (LSUs) and cattle on farms within 3 km zones and administrative regions in New Zealand.* 52

Table 4: *Summary of the mean and standard deviation of the joint posterior estimates of covariates for species distribution models for New Zealand farms using counts of livestock units (LSU) as the outcome variable in the first instance and counts of cattle in the second instance. Risk ratios are reported for the Poisson component and odds ratios for the Bernoulli component for each model along with 95% credible intervals.* 55

Table 5: *Counts of secondary herd types and primary farm types used to parameterise the Australian Animal Disease model (AADIS). Movement patterns and management activities vary by herd type and susceptibility to FMD varies by species.* 80

Table 6: *Descriptive data for each of the herd types used to parameterise AADIS. The minimum, mode and maximum were used to create the beta-pert dataset, the median for the median dataset and remaining descriptors serve to describe the distribution of the gold standard data* 81

Table 7: *Description of the six New Zealand territorial local authorities (TLAs) used as disease index herd for hypothetical foot-and-mouth (FMD) outbreaks in the study.* 84

Table 8: *Tabular representation of a study designed to test the null hypothesis that uncertainty around herd size is not important when interpreting the results of a within-herd spread FMD model. Three control strategies and three herd population datasets were used to create nine model scenarios each of which were run for 1000 iterations in each of six New Zealand territorial local authorities – each cell (1A – 3C) represents 6000 model iterations each using the same order of seed farms.* 86

Table 9: *Number of simulations that generate outbreaks that are not detected (Not detected), outbreaks that last 365 days without being eradicated (Right censored), and number of simulations where the outbreak is detected and eradicated within 365 days (Detected and eradicated), by region and by dataset.* 90

Table 10: *Analysis of variance (ANOVA) table for the Cox Proportional Hazards (CPH) models with infected premises (IPs) and duration as outcome variables. Explanatory variables were identical for both models. Given the large size of the dataset analyzed, the small p-values might be expected, however the large relative deviance for datatype provides an indication of the importance of this variable in the models.* 91

Table 11: *Descriptive five number summaries (minimum, 25th percentile, median, mean, 75th percentile and maximum) for each of 2 outcome variables (count of infected premises and duration in days) and 3 explanatory variables for the models described in table 8. These results are graphically represented in Figures 2 and 3.* 97

Table 12: *Survey questions presented to the village headmen of 168 villages in the project area* 106

Table 13: *Data sources and spatial data layers used to build a data frame of explanatory variables used to explore the usefulness of national level data for modelling FMD prevalence at the village level in 25 townships in central Myanmar. Sources include the Livestock Breeding and Veterinary Department in Myanmar (LBVD) and the Myanmar Information Management Unit (MIMU)* 109

Table 14: *Six number summaries of explanatory variables derived from national level data in a project region of central Myanmar.* 112

Table 15*: Six number summaries of continuous explanatory variables derived from the survey questionnaire data in a project region of central Myanmar.* 113

Table 16: *Counts of categorical explanatory variables derived from the survey questionnaire data in a project region of central Myanmar. Aside from these variables with binary outcomes, when village headmen were asked if any treatment was undertaken for FMD, 3 answered that nothing was done, 3 claimed that FMD affected animals were slaughtered, 43 said they vaccinated in the face of an outbreak and the remaining 119 applied other treatments. Results have been stratified by the village level status showing if the village is regarded as a case of FMD (serological evidence of exposure in young stock).* 114

Table 17: *Regression table of coefficients and p values for three logistic regression models where the outcome was the presence of FMD in young cattle (6-18 months old) in central Myanmar. The three models use survey data, national data and a combination of survey and national data to model the serological presence of FMD in young stock at the village level.* 116

Table 18: *Variable importance in three random forest models predicting FMD at the village level in central Myanmar. One model uses only questionnaire data, one uses only nationally available data and the third uses data combined from both of these sources.* 117

Table 19: *Counts of false negative, false positive and correctly identified villages from the test dataset identified by 2 model types (random forest models and logistic regression models) over 3 datasets (questionnaire data, nationally available data and combined data).* 119

Table 20: *Distribution of the counts of households and FMD susceptible animals present in 160 villages included in a serological survey performed in the Sagaing and Mandalay regions of central Myanmar.* 132

Table 21: *Comparison of how village headman and householder interviews reported outbreaks in 2015 and 2016 (2A) and for 2012 to 2014 (2B) and for all records (2C) displayed as raw counts.* 137

Table 22: *Comparison of the agreement between village headman reports, householder reports and ELISAs both as simple agreement and using Cohen's Kappa in three different time periods. K=0 is no agreement, 0–0.2 is slight agreement, 0.21–0.40 is fair agreement, 0.41–0.60 is moderate agreement, 0.61–0.80 is substantial agreement and 0.81–1 is almost perfect agreement* 138

Table 23: *Model results of 3 latent class models representing 3 ways to identify FMD at the village level (ELISA testing of a sample of animals, survey questions to village headmen and survey questions to householders) and the estimates of incidence, prevalence, sensitivity and specificity derived from each model* 141

Table 24: *Counts and percent of the total of animal movement records that had data source classifications in the Mycoplasma bovis database and where available the type of animals involved in the movement event. Percentages in brackets show the proportion of animal type reflected in each collection method.* 154

Table 25: *For those farms investigated as a part of the Mycoplasma bovis eradication campaign, the proportion of all farm movements which were present in NAIT was calculated by treating the total movements from all sources as the denominator and the movements recovered from NAIT as the numerator.* 154

Table 26: *Description of the available data from 3 abattoirs in 3 time periods in New Zealand. The counts and proportions of instances of missing information and complete information (Tag purchaser and sending NAIT number are the same) are shown along with the counts of suppliers and consignments.* 158

Table 27: *Logistic regression model table showing the role of time period and region against movements with no missing data at three New Zealand abattoirs. The time period between 15th of January and the 15th of March 2017 (A17) is the referent category for time and the abattoir designated as MC is the reference category for the abattoir fixed effect. The outcome variable in our model was a 1 if the animal had no missing information identified in the step from supplier to abattoir.* 160

Table of Figures

Figure 1: *Map showing the 16 regions and the two main islands of New Zealand. Grey shaded areas represent the areas used to compute variograms in the North Island and South Island to evaluate spatial autocorrelation of model fit shown in this case by the rank probability score.* 50

Figure 2: *Variograms for the LSU model (A) and the cattle model (B) showing the spatial covariance in the farm-level Rank Probability Score (RPS) in areas presented in Figure 1 in the North Island (NI) and the South Island (SI).* 51

Figure 3: *Boxplot of RPS for each of the 7 farm types in AgriBase for models predicting livestock units (light grey) and 5 farm classes in AgriBase for models predicting cattle (dark grey) on rural properties in New Zealand.* 60

Figure 4: *Box plots of the RPS from 50 resampled fits of the model where region was removed from A, the LSU model and B the cattle model.* 61

Figure 5: *Receiver operating characteristic (ROC) curve for a model predicting the absence of cattle (dotted line) and the absence of LSUs (shown with a solid line) on rural properties in New Zealand..* 62

Figure 6: *Six New Zealand territorial local authorities (TLAs) selected for locations of simulation models to examine the effects of variation in herd size in the within-herd spread disease model of foot-and-mouth disease (FMD), AADIS.* 83

Figure 7: *Distribution of duration of outbreak for models using each of three herd size data sets across six New Zealand territorial local authorities (TLAs) for three control strategies.* 89

Figure 8: *Distribution of the log10 of final count of infected premises (IPs) for models using each of three herd size data sets across six New Zealand territorial local authorities (TLAs) for three control strategies.* 89

Figure 9: *Ratio between the operational costs generated by a comparison between the strategy indicated as being the lowest by real herd data-set and the costs of the strategy that would have been chosen should an alternate herd data set have been used. The x axis is displayed at the log scale.* 92

Figure 10: *Twenty four project townships in Myanmar (inset top right). Striped townships are in Mandalay Region and spotted townships in Sagaing Region. Both the Mandalay and Sagaing regions are in central Myanmar.* 105

Figure 11: *Distribution of project regions, townships and villages. Twenty four project townships are part of our project area in central Myanmar (inset top right). Striped townships are in Sagaing Region*

and speckled townships in Mandalay Region. Both the Mandalay and Sagaing regions are in central Myanmar. 131

Figure 12: *Variation in the percentage of village householders able to correctly identify a picture shown to them depicting the clinical signs of FMD in a lineup of 4 endemic diseases. Householder responses have been aggregated to the township level (x axis) and Regions are indicated by colour (Mandalay in pink and Sagaing in blue).* 133

Figure 13: *Geographic distribution of total prevalence and prevalence in young stock at the township level. Village level seroprevalence in the sampled population aggregated to the township level (panel A) compared with the same measure for animals 6-18 months of age (panel B). Light grey townships have less than 15% seropositivity, mid grey 15% - 25% and dark grey greater than 25% seropositivity.* 134

Figure 14: *Scatter plot of the relationship between the percentages of 6-18 month old animals with positive serological results and the percentage older animals at the village level in 2 regions of central Myanmar. The reference lines indicate the cut points for considering a village to be positive (greater than 10% of animals positive on serology).* 135

Figure 15: *ROC curves for the 2 latent class models which represent the ELISA testing data (including sampling uncertainty) as continuous variables.* 140

Figure 16: *Boxplot of the distribution of consignment sizes amongst three abattoirs in 6 time periods in New Zealand. Animals arriving at a particular abattoir on a particular date from a particular supplier were allocated to a consignment.* 156

Figure 17: *Diagram showing the selection process of animal movement records used to estimate the extent of missing data from NAIT over six time periods in 2 years at 3 abattoirs.* 157

Figure 18: *distribution of the number of recorded movements that a single animal had recorded in its lifetime in different time periods and at different abattoirs in New Zealand* 159

Abbreviations

NAIT	The National Animal Identification and Tracing System. The animal movements database in New Zealand. It is managed by OSPRI.
FOL	Farms On Line. The MPI managed database of farm properties
MPI	The Ministry for Primary Industries
CEBRA	The Centre for Excellence in Biosecurity and Risk Analysis at University Melbourne
FMD	Foot and Mouth disease
OSPRI	The New Zealand company that runs NAIT and the NZ bovine tuberculosis eradication programme (TbFree). OSPRI stands for Operational Solutions for Primary Industry
ISP	Interspread Plus. A disease spread simulator used in New Zealand by MPI to model FMD outbreaks (amongst many international users).
AADIS	The Australian Animal Disease Spread Model. A within and between herd spread model designed and built in Australia
AQ	AssureQuality. A New Zealand state owned enterprise that provides operational capability to MPI
TAD	Transboundary Animal Disease
FAO	Food and Agriculture Association of the United Nations
OIE	World Organisation for Animal Health
LBVD	Livestock Breeding and Veterinary Department (Myanmar)

Publications

A. Journal articles produced with the doctoral candidate as the primary author and the publication is included in the thesis

1. Predicting farm-level animal populations using environmental and socioeconomic variables. Preventive Veterinary Medicine 2017, volume 145, page 121-132. Mary van Andel, Christopher Jewell, Joanna McKenzie, Tracey Hollings, Andrew Robinson, Mark Burgman, Paul Bingham, Tim Carpenter

2. Does Size Matter to Models? Exploring the Effect of Herd Size on Outputs of a Herd-Level Disease Spread Simulator. Frontiers in Veterinary Science 2018, volume 5 (May). Mary van Andel, Tracey Hollings, Richard Bradhurst, Andrew Robinson, Mark Burgman, M Carolyn Gates, Paul Bingham, Tim Carpenter

3. Evaluating the utility of national-scale data to estimate the local risk of foot-and-mouth disease in endemic regions. Transboundary and Emerging Diseases - submitted. Mary van Andel, Scott Zaari, Andrew McFadden, Paul Bingham, Barbara Binney, Kelly Buckle, Ben Phiri, Cord Heuer, Ian Dacre, Ronel Abila, Khin Ohnmar Lwin, Htun Htun Win, M. Carolyn Gates

4. Estimating foot-and-mouth disease (FMD) prevalence in central Myanmar: comparison of village headman and farmer disease reports with serological findings. Transboundary and Emerging Diseases - submitted. Mary van Andel, Geoff Jones, Barbara Binney, Kelly Buckle, Ben Phiri, Andrew McFadden, Ian Dacre, Paul Bingham, Cord Heuer, Ronel Abila, Htun Htun Win, Khin Ohnmar Lwin, Scott Zaari, M. Carolyn Gates

5. Integrating multiple data sources to evaluate the quality and completeness of national animal identification and traceability systems. Not yet submitted. Mary van Andel, Karyn Froud, Kara Dawson, Linda Collard, Fanny LeDuc, Paul Bingham, Dale Kaio, Kristy Powell, Chris Morley, M. Carolyn Gates

B. Journal articles produced with the doctoral candidate the primary authors (joint) but not included in the thesis due to delays in doctoral registration process

1. Compatibility between livestock databases used for quantitative biosecurity response in New Zealand. New Zealand Veterinary Journal 2016. 64:3 page 158-164. Christopher Jewell, Mary van Andel, Daan Vink, Andrew McFadden

C. Journal articles produced where the doctoral candidate is one of the authors and the work relates directly to the projects included in the thesis

1. How do you find the green sheep? A critical review of the use of remotely sensed imagery to detect and count animals. Methods in Ecology and Evolution 2018. Volume 9 issue 4 page 881-892. Tracey Hollings, Mark Burgman, Mary van Andel, Marius Gilbert, Timothy Robinson, Andrew Robinson

2. Species distribution models: A comparison of statistical approaches for livestock and disease epidemics. PLoS ONE 2017, 12(8). Tracey Hollings, Andrew Robinson, Mary van Andel, Christopher Jewell, Mark Burgman

D. Report produced relating directly to the projects included in the thesis

1. Estimation of national-level farm demographic data for preparedness of highly infectious livestock disease epidemics.

https://cebra.unimelb.edu.au/__data/assets/pdf_file/0020/2615141/1402_1502C_FINALREPORT-2.pdf

CHAPTER 1

1 Introduction

1.1 Background

One of the important roles of government agencies is planning for the future and preparing for emergencies that may disrupt the country's economy and society. These emergencies include floods, fires, droughts, earthquakes, and disease outbreaks that impact not only human populations, but also domestic and wild animal populations that share our environments. To better support preparedness and response activities, it is essential for government agencies to have detailed information on the exact location and demographic characteristics of the population. However, the availability and accuracy of this type of information as well as the implicit biases in most national databases have the potential to dramatically impact how planning activities are done and how responses are executed, which in turn drives the success or failure of national control efforts.

New Zealand has a maritime climate and is geographically isolated making it an ideal location for primary agricultural activities. The agricultural industry is important in the New Zealand economy where it contributes more than half of export revenue and almost a fifth of national GDP (Statistics NZ, 2014). However, New Zealand's production systems are in some cases based on farming practices which would not be feasible in a situation where additional infectious and vector borne diseases were present. For example, the New Zealand specific dairy practices including moving day[1], calf clubs[2], calf rearing[3], lease bulls[4] and winter grazing[5] would not be profitable in a farming

---

[1] Practice where share milkers (farmers with their own cows but no land) move to another farm with their cattle and equipment to farm for the next season.

[2] Rural school annual events where hand reared livestock are brought together by the school pupils to participate in competitions and mingle at a fair before returning to the farms of origin.

[3] Male dairy calves are purchased in the first week of life and raised on milk along with calves sourced from multiple farms. Calves may be fed waste milk although this is no advised. At 3 or 4 months of age these animals are sold on to be grazed elsewhere before being slaughtered for meat.

[4] After artificial insemination, bulls are introduced to the heifer herd as 'mop-up' bulls. These animals are often leased and returned to their origin farm to be leased out again in the following season.

[5] Animals that are not lactating are sent away to grazing properties a distance from the dairy platform where they may comingle with stock from other farms. They are returned to the platform to calve.

environment which included endemic diseases like Mycoplasma bovis[6], Brucellosis[7] or bovine tuberculosis[8]. Preventing the arrival of agricultural diseases is crucial so that access to international markets is maintained and that reputation, the economy and industry are protected. Biosecurity investment by New Zealand is substantial but the economically damaging incursions of varroa mite *(Varroa jacobsoni*), fire blight (*Erwinia amylovora*), bacterial kiwifruit vines disease (*Pseudomonas syringae pv. actinidiae*) and *Mycoplasma bovis* into New Zealand show that significant biosecurity pests can still pass borders undetected.

The costs of such incursions to disease free countries like New Zealand are primarily measured in terms of the cost of eradication and compensation and how that would compare with the production and market losses that would be incurred if the disease were to become established. According to an economic assessment performed by New Zealand's central animal health authority, the Ministry for Primary Industries (MPI), an outbreak of foot-and-mouth disease (FMD) would result in a total loss in real GDP of up to $16.2bn over 8 years (Forbes and van Halderen, 2014b). A high priority for New Zealand is therefore to reduce the risk of an outbreak of FMD and to prepare as best it can for such an outbreak should it occur. To this end, MPI has responsibility for investigation, response and preparedness against exotic and emerging disease outbreaks.

The situation is often different for developing countries with endemic FMD where the main costs of the disease are associated with decreased animal production at the individual household level as well as loss of draft power, which both have a profound impact on household income and food security (Nampanya et al., 2016, 2015; Perry et al., 1999; Young et al., 2013). In 2010, it was estimated that the benefit to cost ratio of eradication of FMD from South East Asia was 3:1 (Mcleod, 2010). In light of these benefits, control of foot-and-mouth disease (FMD) has been identified as a priority for the livestock production sector in Myanmar. Myanmar is taking part in the Food and Agriculture Organisation of the United Nations (FAO) and World Organisation for Animal Health (OIE) Progressive Control Programme (PCP) for FMD*;* a programme which aims to provide epidemiological support and benchmarking to those countries that wish to progress toward FMD freedom (Jamal and Belsham, 2013).

Despite having different socioeconomic and FMD statuses, both New Zealand and Myanmar require information on their animal populations to undertake disease preparedness and response activities.

---

[6] First identified in New Zealand in 2017 and currently the subject of an ongoing eradication programme

[7] New Zealand is free from *Brucella abortus*

[8] Present at very low prevalence and in the end stages of an eradication programme

A possible solution to unknown animal population information is to model the size and location of farms based on satellite imagery or known land use characteristics. This has proved to be a useful strategy in the United State of America(Burdett et al., 2015) and Australia and is widely used in wildlife ecology(Hollings et al., 2018, 2017). The Food and Agriculture Organisation of the United Nations (FAO) has released a world atlas of modelled domestic livestock which is a valuable resource(Robinson et al., 2014, 2006).

Once the animal population at risk is known, identifying the highest risk geographic regions for disease outbreaks and prioritising those for surveillance becomes possible. In the case of FMD in New Zealand, the most likely pathway of introduction is believed to be an illegal pathway and it is believed that backyard pigs pose the greatest threat of being the primary case of disease(H J Pharo, 2002). Although FMD is endemic across Myanmar, cattle markets and areas with large numbers of cattle and migratory populations are thought to have the highest disease prevalence. A flow of animals has been documented in the Greater Mekong Subregion (GMS) which has created a textbook example of transboundary animal disease spread. The GMS comprises China, Vietnam, Laos PDR, Myanmar and Thailand. Cambodia, Lao PDR, Thailand and Myanmar can be considered to be net suppliers of cattle while China is a net importer. Vietnam is regarded as being an exporter as well as an importer. This is because Vietnam not only exports cattle to China, but also imports cattle from within the GMS for domestic use. Cattle move through and intricate network of official and unofficial pathways throughout the GMS (Cocks et al., 2009; Smith et al., 2012). These relationships result in a flow of animals from south to north (toward China) and from west to east (toward Vietnam).

In an endemic FMD situation where resource is limited, disease prevalence and incidence estimates must often be based on report cases. Official national reports to the World Organisation for Animal Health (OIE) are affected by a multitude of factors including the costs and logistical difficulties associated with laboratory testing. In these cases, the use of disease reports as a way to estimate prevalence using local knowledge is well established (Bellet et al., 2012a; Calba et al., 2015; Catley et al., 2012; Kenton L. Morgan et al., 2014; Robertson et al., 2010; Sawford, 2011). Prevalence estimates derived from disease reports and population estimates derived from simulation models are essential data sources. But, how these estimates compare with sero-surveys or animal censuses is almost never known. In some cases, it might be possible to compare estimated information with something which is thought to approximate ‘the truth’. An example of this is a comparison of farmer reports of FMD with serological status of the animals in question (Bellet et al., 2012a). When it comes to validating the completeness of national animal movement databases, the opportunity to estimate the extent of the missing data seldom presents itself. The impact of missing animal movement data in the face of an outbreak of disease would lead to a delay in restricting the spread of disease from infected farms and result in an outbreak with longer duration and more farms and animals affected. In the case of an exotic disease in a disease free country, this impact could be economically crippling (Carpenter et al., 2011; Forbes and van Halderen, 2014; Mardones et al., 2013).

Spatially explicit stochastic disease simulators are widely used by FMD free countries to assess the impact of hypothetical FMD outbreaks and different control options in naïve animal populations. (Michael G. Garner et al., 2016; Kitching et al., 2006; Roche et al., 2015). Some of these models have been refined over many years and require specialised expertise as well as many parameter values to run (Owen et al., 2011). In countries where disease is endemic, the competent authority must have existing knowledge of where disease is located before any model can be parameterised. In the absence of such a model (or while such a model is being developed) decisions must be made about deployment of control efforts. In endemic disease situations where some historical data is available, maximum likelihood models and machine learning models can be used to predict outbreaks based on the presence or absence of risk factors. Logistic regression is considered to be a standard approach to predicting a binary outcome (for example outbreak in the period of interest or no outbreak in the period of interest) especially when modelling is undertaken not only for prediction but also to describe and increase understanding of risk factors that could cause the outcome of interest (Langford et al., 2009). By contrast, the random forest prediction algorithm (a machine learning technique) is focused on prediction rather than explanation of risk factors. Furthermore, RF models have been found to out-perform logistic regression models in cases where prediction is the primary goal of the modelling activity (Breiman, 2001; Brownstein et al., 2005; Cutler, 2007; Nicolas et al., 2016; Prasad et al., 2006).

1.2 Thesis Overview

This thesis was started as a part of a project for the Centre of Excellence for Biosecurity Risk Assessment (CEBRA) at University Melbourne in which the overlaps and gaps between national datasets in New Zealand were described (Jewell et al., 2015). The CEBRA project went on to assess the use of satellite imaging to count farm animals (Hollings et al., 2018) and then moved on to consideration of different methods that could be used to model animal counts (Hollings et al., 2017; van Andel et al., 2017). The CEBRA project ended with some exploration of the use of modelled farm animal populations in a spatially explicit FMD spread simulator built specifically for the Australian context and adjusted for New Zealand conditions (Bradhurst et al., 2015). This work was extended in this thesis by building FMD spread models in New Zealand in the AADIS platform using modelled and real animal population estimates and comparing the impacts of the herd data on the inferences that could be drawn from the model outputs (van Andel et al., 2018).

The second section of this thesis was completed as a part of the activities of the OIE's South East Asia and China Foot and Mouth Disease Campaign (SEACFMD) which seeks to coordinate animal disease control activities at the regional level and to provide technical advice. One of the required activities in Myanmar was the description of geographical risk factors of reported outbreaks of FMD in a specific project area and provision of advice about extrapolation of these risk factors to the rest of the country. The purpose of the risk factor identification was to provide some information that could be used to prioritise scarce vaccination resource. Subsequently a strategy was examined for assessing the efficacy of clinical signs reporting when compared with sero-surveillance using latent class analysis in an endemic FMD situation. The final chapter uses two analyses to examine the completeness of data accessed from New Zealand's national animal movement database.

CHAPTER 2

2 Review: Using national animal datasets to support transboundary animal disease control in developed and developing countries

2.1 Abstract

National level databases of animal numbers, locations, and movements provide the essential foundations for disease outbreak investigations, disease control, and disease preparedness activities. These activities are particularly important for managing and mitigating the risks of high impact exotic disease outbreaks like foot-and-mouth disease (FMD), which can significantly impact international trade and food security. In countries where livestock production systems are highly subsidised by the government, producers are often required to provide animal movement and demographic data as a condition of business. In the remaining countries, it can be difficult to maintain these types of databases and impossible to estimate the extent of missing or inaccurate information due to the absence of gold standard datasets for comparison. Given the global importance of this issue, the objective of this paper was to review the activities for which a national biosecurity authority would require animal population, location, and movement data both in the case of an FMD free and an FMD endemic situation. In this review, we first discuss the different activities that competent authorities use farm-level animal population data for to support (1) preparedness activities in disease free countries, (2) response activities during an acute outbreak in a disease free country, and (3) eradication and control activities in an endemic country. We then discuss (4) data requirements needed to support epidemiological investigations, surveillance, and disease spread modelling both in disease free and endemic countries.

2.2 Introduction

According to the Food and Agriculture Organisation of the United Nations (FAO), a transboundary animal disease (TAD) is defined as an infectious disease of significant economic, trade, and/or food security importance for a considerable number of countries that can easily reach epidemic proportions when it spreads into disease-free countries and for which disease control or management requires cooperation between several countries (Otte et al., 2004). An outbreak of such a disease can prove devastating to countries whose economies rely on the export of live animals and/or animal products due to the trade restrictions that are applied until proof of freedom can be established (Anonymous, 2013a). As such, it is the responsibility of the competent animal health authority to prevent and prepare for outbreaks in disease-free countries and to eradicate, manage, and report outbreaks in affected countries.

A relevant example of a TAD is foot-and-mouth disease (FMD), which is a rapidly spreading viral disease caused by a picornavirus that causes significant production losses due to high morbidity, but has very limited mortality (Alexandersen et al., 2003b). The presence of FMD in a country limits the options for primary sector exports and so countries that are regarded as FMD-free have strong economic incentives to mount national eradication campaigns in the event of an FMD incursion. As highlighted by the 2001 FMD outbreak in the United Kingdom, the impact of the disease and the associated eradication efforts are profound at every level including effects on gross domestic profit, international trade, the primary sectors, human and animal welfare, tourism, and the environment (Blake et al., 2003; Hayama et al., 2017, 2016; Hibi et al., 2015; Leslie and Black, 2006; O'Toole et al., 2002; Tildesley et al., 2012; Williams and Ferguson, 2005; Wilson and Kinsella, 2004).

In countries that are currently free of FMD, preparedness and response efforts are focussed on preventing the introduction of FMD and minimising the time from introduction to declaration of freedom should an outbreak occur. Countries where FMD has become endemic also generally wish to gain disease freedom since the impacts of subclinical production losses on food security and clinical production losses associated with the circulation of new strains are deemed to be substantial (Knight-jones and Rushton, 2013). In 2010, for example, it was estimated that the benefit to cost ratio of eradication of FMD from South East Asia was 3:1 (Mcleod, 2010). The losses attributed to FMD in South East Asia include lost work from draft animals and lost production at the household level, which both negatively affect food security and household income (Nampanya et al., 2016, 2015; Perry et al., 1999; Young et al., 2013). The control of FMD in endemic countries can also have positive externalities for disease-free countries by reducing the risk of FMD entry since border control and import regulations cannot completely eliminate the movements of high risk products and fomites (Rweyemamu et al., 2008).

Irrespective of the current disease status of a country, it is important to have accurate baseline data on the numbers and locations of animal populations that are potentially susceptible to the TAD to adequately plan, resource, and prioritise operations in the field as well as to quantify the amount of disease in a population (Buhnerkempe et al., 2014; T Porphyre et al., 2013; Thrusfield, 2013;

Woolhouse, 2003). In the unfortunate event of an incursion of a fast moving disease like FMD in a disease free country, it is essential to have these data available and in place prior to the introduction of these disease. This will give the best chance that the eradication programme does not transition to a control programme. In those countries with endemic FMD, the availability of these population data will provide the basis for planning and progressing a control programme. Although these data are acknowledged to be important for disease control, maintenance of national databases is problematic. Even in countries where national databases exist, the quality and timeliness of the data is questionable (Honhold and Taylor, 2006; Jewell et al., 2015). In many developed countries, data are collected for purposes other than biosecurity and protected by privacy legislation and so are not fit for purpose or not available for use. When considering a TAD, the ideal data situation would include linking of movement and demographic datasets in a region to enable a coordinated approach to disease control. There are significant legal and political challenges to such data linking activities, but without them, epidemiologists must rely on the results of population models or extrapolated surveys to furnish the essential population information.

In this review, we first discuss the different activities that competent authorities use farm-level animal population data for to support (1) preparedness activities in disease free countries, (2) response activities during an acute outbreak in a disease free country, and (3) eradication and control activities in an endemic country. We then discuss (4) data requirements needed to support epidemiological investigations, surveillance, and disease spread modelling both in disease free and endemic countries.

2.3 Preparedness activities in disease free countries

2.3.1 Surveillance to maintain disease freedom status

FMD can broadly be diagnosed in three ways: PCR (or Antigen ELISA) to detect the presence of the virus, Antibody ELISA to detect a susceptible host's immune response at some time in the past to the virus, and the observation of clinical signs (Alexandersen et al., 2003b; Longjam et al., 2011). Diagnosis by PCR and Ab ELISA requires samples either submitted to laboratories or analysed using pen-side tests (Longjam et al., 2011; Reid et al., 2001). In most disease free countries, diagnosis based on clinical signs alone is not sufficient as endemic differentials exist and must be ruled out (T.W. Bates et al., 2003; Holliman, 2005; Watson, 2004). These tests are usually part of a larger active or passive surveillance system which provides the necessary level of assurance required for the country to fulfil its obligations to the OIE and trading partners (Tana, 2014).

2.3.2 Identifying risk pathways for disease entry

Given the severe impacts of an outbreak of FMD and the fast spread of the disease, it is essential that the competent authority of a disease free country have a policy framework in place in preparation for an FMD outbreak. The length of time between introduction and detection of a disease outbreak has a profound impact on the extent of the spread of any infectious disease outbreak. Especially in the case of a highly infectious disease like FMD, any delay between introduction and detection may result in a large number of additional disease cases (Bouma et al., 2003; Davies, 2002; Dubé et al., 2008; Ferguson et al., 2001; Muroga et al., 2012; Park et al., 2013; Yang and W. and Sung, 1999) and substantial economic impacts (Hayama et al., 2013; Ward et al., 2009). The objective of surveillance for exotic disease is then to direct the available surveillance resources toward the sector or geographic region where disease is most likely to be introduced and to perform surveillance in the manner which optimises detection sensitivity (M. G. Garner et al., 2016; Kompas et al., 2017). For example, in New Zealand, the risk of introduction of FMD is deemed highest by the pathway of illegal feeding of imported waste food to backyard pigs (H. J. Pharo, 2002) and being able to develop accurate risk models therefore requires data on importation of animal products as well as knowledge of the location and management practices of backyard pig operations.

2.3.3 Predicting spread post-introduction

It is difficult to completely eliminate all risk of introductions and so the competent authorities are also interested in understanding how disease may spread after introduction to predict the potential size and scale of outbreaks. Spatially explicit stochastic disease simulators are widely used by FMD free countries to assess the impact of hypothetical FMD outbreaks and different control options in naïve animal populations (Michael G. Garner et al., 2016; Kitching et al., 2006; Roche et al., 2015). Spatially stochastic spread modelling of FMD has historically been centred in FMD free countries and used to compare the costs and impacts of different control options in the event of an outbreak of FMD (Kompas et al., 2017; Roche et al., 2015; Sanson et al., 2017). The investigation of the early characteristics of large outbreaks is also of great interest (Sarandopoulos, 2015; Webb et al., 2017) . These models are typically visual representations of the primary production sector with the location of farms, movement patterns and counts of susceptible animal populations used as starting parameters. This allows disease investigators to create different disease introduction, spread and control scenarios which can then be compared when making policy decisions.

These simulation models rely on accurately replicating disease transmission pathways, which can be difficult to predict without having information from past outbreaks in the country. FMD is able to be spread both by susceptible live animal movements and via mechanical contact with contaminated personnel, vehicles, fomites, and wildlife (e.g., rats, mice and birds) (Alexandersen et al., 2003b; Sutmoller et al., 2003). Some parameters can be extrapolated from previous outbreaks in other historically free countries and some from endemically affected countries. Much remains unknown about the between and within herd transmission rates due to varying animal population structures, movement patterns and country specific farming practices. Disease spread is affected by the number of animals present on farms, as well as animal and vector movements that link farms to each other spatially. Modelled predictions of spread are affected by the scale of the model, for example the use of polygon or farm-level area information as opposed to point features that represent the weighted centroids of farms will affect model predictions of spread (Flood et al., 2013). In the event of prioritisation of farms to receive surveillance visits, those with a boundary that is shared with an infected place (IP) can be expected to have higher priority than those in close proximity, but not sharing a fence line.

Although some of these models have been refined over many years, most still require specialised expertise as well as many input parameter values to run (Owen et al., 2011). Animal population estimates will have variable effects on the model outputs depending on how susceptible populations are used in the spread simulation part of the model and in the economic part of the model (Keeling et al., 2001; Thibaud Porphyre et al., 2013a; Shea et al., 2014). For example, in the Australian Animal Disease spread model (AADIS) (Bradhurst et al., 2015) the population of animals on each farm is used to generate infectivity for each farm according to latent periods, within herd contact rates and incubation periods specific to the species and numbers of each species present; this creates a unique probability of transmission curve for each farm based on its population. In the disease spread simulator InterSpread Plus (ISP) (Stevenson et al., 2013) a probability of transmission curve is applied to each farm type based on the average size (in chosen number of geographic strata) of a herd of a particular type.

Advances in modelling techniques and increased computational ability have resulted in the current trend in veterinary epidemiology of using individual animals to generate infectivity at the farm-level by the inclusion of intra-herd spread within disease spread simulators. Due to this high level of detail at the farm-level, these models are being used to evaluate the impacts of management decisions at a microeconomic level on individual farms and to address the effects of vaccine strategies, false positive results and test and removal control strategies for some diseases. Livestock density strongly influences the farm-level reproductive number used in these models (Thibaud Porphyre et al., 2013b) and can only be estimated for a particular farm by knowing the count of animals present on the property. Porphyre et al (2013) further show that it is the spatial variation of these farm-level reproductive numbers that best inform a decision to deploy vaccination for FMD in the United Kingdom situation, along with knowledge of the circulating strain of virus.

In situations where it is not possible to survey every farm in a population, risk based surveillance is an option. The rationale is that failure to detect disease in a high risk subgroup at a specified design prevalence provides greater confidence of freedom from disease than a random sample of the whole population at the same design prevalence (Astudillo et al., 2016; Caporale et al., 2016b). However, to use this technique, the population demography is a prerequisite and the integration of animal movement patterns is common (Frössling et al., 2014, 2012, M. C. Gates et al., 2013a, 2013c; Gates and Woolhouse, 2015).

In countries where disease is endemic, the competent authority must have existing knowledge of where disease is located before any model can be parameterised. In the absence of such a model (or while such a model is being developed), decisions must be made about deployment of control efforts. In endemic disease situations where some historical data is available, maximum likelihood models and machine learning models can be used to predict outbreaks based on the presence or absence of risk factors. Logistic regression is considered to be a standard approach to predicting a binary outcome (for example outbreak in the period of interest or no outbreak in the period of interest) especially when modelling is undertaken not only for prediction but also to describe and increase understanding of risk factors that could cause the outcome of interest (Langford et al., 2009). By contrast, the random forest prediction algorithm (a machine learning technique) is focused on prediction rather than explanation of risk factors. Furthermore, RF models have been found to out-perform logistic regression models in cases where prediction is the primary goal of the modelling activity (Breiman, 2001; Brownstein et al., 2005; Cutler, 2007; Nicolas et al., 2016; Prasad et al., 2006).

2.3.4 Impacts of incomplete data on policy decisions based on spread modelling

There is an increasing body of literature related to the impacts of incompleteness on simulation modelling of disease spread. This work focusses for the most part on comparing ways to predict missing data that ensures maximal model performance. Both animal movement networks and farm locations have been studied in this way. It has been found that a realistic entire animal movement network can be predicted using targeted node surveillance or snow ball sampling techniques. The authors point out that the method has some limitations in application to entirely unknown networks as it is not possible to know when sufficient nodes have been sampled to accurately predict the entire extent (Dawson et al., 2015). When exact animal locations are not available, land cover maps can be used to infer the location of the susceptible population (Burdett et al., 2015; Louz et al., 2013; Tildesley and Ryan, 2012). The performance of this method varies based on the scale of aggregation with larger aggregations performing less well than small scale ones (Tildesley and Ryan, 2012). When these modelled populations are included in disease spread simulators and policy options compared, it is possible to make policy recommendations that would be the same as those made with disaggregated data in some cases (Tildesley et al., 2010), however this ability is specific to the question at hand and the parameterisation of the model being used (which includes country specific factors that influence disease spread).

Other studies focus on the value that perfect information adds to a response to an infectious disease by identifying where more information should be collected as a part of preparing for an outbreak (Bradbury et al., 2017; Shea et al., 2014).

All of these studies provide essential insight into the functioning of our model frameworks, our available data sets and the biases that could be introduced into preparedness planning as a result of incomplete and flawed data. What these studies cannot do is provide data that is sufficiently nuanced for operational purposes in the field when a disease outbreak occurs.

2.4 Response activities during an acute response in a disease free country

2.4.1 Contact tracing

Disease control policies must be developed prior to the identification of an outbreak as the field activities will have to be implemented urgently. In particular, the immediate establishment of movement controls and controlled areas can greatly limit the spread of an FMD disease outbreak (Carpenter et al., 2011). Although all possible efforts should be made to develop a robust policy framework prior to the discovery of an outbreak, some strategic changes in direction will be required as the outbreak progresses and data from field operations become available. The ability to process and analyse the outbreak data in real time allows decision makers to strategically evaluate control policies and make appropriate changes (Mansley, 2004). Epidemiologists employed by the competent authority for biosecurity provide information and reports to decision makers on progress of disease control activities and advise on strategic and operational changes that could be implemented in the field to improve outcomes. These reports will include advice of whether an outbreak is under control (Ferguson et al., 2001; Paine et al., 2010) and inform policy adjustments in the face of an outbreak. Ongoing real-time analysis of testing data is also required.

It is clear that not all farms will present the same risk of onward spread of disease. In particular, the number and species of animals present on a farm as well as the characteristics of the animal populations in the surrounding 10km as important for estimating the risk of infection with FMD. Similarly with live animal contact tracing and fomite tracing, those farms with the presence of large numbers of virus excretors or those with many contacts should be prioritised for immediate visits (Eames and Keeling, 2003). Farms in the 3km zone surrounding an infected place are known to be at higher risk of infection and all farms in this zone are visited by surveillance teams during a response. However, those farms which share contiguous borders are at even greater risk than those which are nearest neighbours but one or those separated by a road or other geographic feature and should be visited first (Flood et al., 2013; Gibbens et al., 2001; Thrusfield et al., 2005). In addition to geographic proximity, those farms with large numbers of links to and from other farms act as hubs in the network of movements formed by farming practices. These farms with high in and out degree are important sentinel sites and provide information about unrecognised disease circulating in the population. These farms can act as a proxy for large numbers of farms which supply animals and are of particular value following eradication of all known infectious premises when scanning of the population is being undertaken to gather evidence of disease freedom in the population (Caporale et al., 2016b; M.C. Gates et al., 2013d; Gates and Woolhouse, 2015; Schärrer et al., 2015).

2.4.2 Epidemiological investigations

The situation and best course of action can change very rapidly during a fast moving disease response. It is essential that field priorities are continuously reviewed and the updated priorities communicated to the field control centres for implementation (Paul R Bessell et al., 2010; Mansley, 2004). As an example, those properties with intensively housed pig populations within a 3km surveillance zone around a property that has been infected with FMD may be visited first due to the high risk of aerosolised virus attributed to pigs. These results should be examined in combination with climatological modelling of viral plume spread to gain further information on properties at highest risk. Populations of animals present on the source farm, populations of animals present on farms under the modelled plume and the prevailing weather conditions are required to complete this analysis (Alexandersen et al., 2002; Alexandersen and Donaldson, 2002; Paul R. Bessell et al., 2010; Donaldson et al., 2001; Hess et al., 2008; Sanson, 1994; Sorensen et al., 2000).

Each of these trace events must be documented, assessed for the risk they present and followed up with field visits. A further source of surveillance data during an outbreak response to FMD will be the investigation of disease reports by the public. During these investigations the contact details of farm managers and information on other holdings which present the same level of epidemiological risk as the farm under investigation because of farming activities on non-contiguous land, and knowledge of farms location in relation to other landmarks, which assist in assessment of spatial risk factors (Brangenberg and van Andel, 2011). The disease investigator on the farm must collect the movement and demographic information, and clinically examine and sample an appropriate subset of animals on the farm to adequately determine the true disease status of the farming operation (Alexandersen et al., 2003a; P. C. Cleland et al., 1995; Thrusfield, 2013).

2.4.3 Evaluating control strategies

2.4.3.1 Test and cull

Once FMD is identified in a historically free country, depopulation with accompanying cleaning and disinfection of all fomites in contact with the farm are required immediately to limit the potential for onward disease spread. This urgency relates to the infection pressure that a single farm infected with FMD presents and the fact that the spread of FMD may occur aside from recorded animal or vector movements to nearby farms. This unclassified spread over short distances has at times been labelled under the heading 'local spread' due to lack or resource to investigate every infectious event. Failure to have sufficient sites available to bury infected animals resulted in disease spread in Japan which delayed eradication efforts (Flory et al., 2017; Hayama et al., 2012; Muroga et al., 2013). Aside from immediate impact on disease spread, insufficient planning for associated disposal activities has numerous impacts that may extend many years after the end of an outbreak and can include economic losses, groundwater contamination, and air pollution from burial sites (Gwyther et al., 2011; Hseu and Chen, 2017; Joung et al., 2013; Kim and Kim, 2012). These major challenges were encountered during the 2001 FMD epidemic in the UK and the Korean FMD outbreaks in recent years, with drinking water contamination problems being encountered in both countries as a result of inappropriately positioned burial pits. Alternative strategies to burial include composting, high-temperature gasification or the negotiation with government bodies regarding the transportation of carcasses across multiple regions to reach suitable landfill sites. These negotiations must be informed by geospatial analyses to compare livestock densities with local capacities to bury or burn carcasses to identify areas that require special measures.

2.4.3.2 Vaccination strategies

Various FMD vaccination policies have been explored by disease free countries using spread modelling including single species vaccination, zonal vaccination, vaccinate to live (VTL) and vaccinate to die (VTD) policies. Models of hypothetical FMD outbreaks have returned conflicting results regarding the benefit of vaccinating cattle only (Roche et al., 2015; Sanson et al., 2017). Laboratory transmission studies have shown that cattle and sheep are equally susceptible to FMD virus infection, but cattle are more infectious than sheep (de Rueda et al., 2014). From an epidemiological perspective, there might be merit in vaccinating cattle in preference to sheep in order to reduce transmission rates in the population when vaccination resources are limited. In New Zealand livestock industries are primarily focussed on sheep and cattle farming. As cattle and pigs are better indicator species for FMD than sheep and goats, cattle-only vaccination may mask infection in an area where diagnosis of disease was dependent on clinical signs. This could delay eradication or interfere with proof of freedom surveillance testing after an outbreak. If vaccinating cattle only were to be shown to be equally effective as vaccinating all species, this could result in substantial savings in the number of vaccine doses required to confer benefits, and decrease the number of animals to be destroyed in "vaccinate-to-die" policies.

The options for vaccination under the OIE code for historically FMD free countries with animal and animal product export markets are limited to either a "vaccinate-to-live" (VTL) or a "vaccinate-to-die" (VTD) policy (Anonymous, 2017). Under the former policy, vaccinated animals continue to be farmed normally for the duration of their productive lives, and the country seeks to prove to the Office International des Épizooties (OIE) that is it "free from FMD with vaccination". The latter means that the vaccinated animals are slaughtered as soon as there is culling capacity available, and the country seeks to then prove that it is "free from FMD without vaccination". Under a VTD policy, vaccinating animals places a near immediate death sentence on them. The associated costs, including loss of genetic material, compensation to farmers, time to depopulate and dispose the animals following the epidemic are expected to be very high. In these cases, vaccination would likely be deployed around infected farms to stop local spread of the disease and would be deployed from the outside of an agreed size circular zone inwards. The OIE allows countries to regain disease-freedom 3 months after culling all vaccinated animals (although importing countries may still take longer to restore trade) if a country adopts a VTD policy (Anonymous, 2017). Under a VTL policy a country can only regain FMD-freedom 6 months after the last case is identified and surveillance on all vaccinated animals is completed. The objective of this surveillance is confirmation that no infection is present in the national herd. A blanket vaccination strategy that supplied complete coverage with a trusted vaccine was successful at ending the 2001 outbreak in Argentina (Knight-Jones et al., 2016; Perez et al., 2004) and similar approaches have been successful at ending outbreaks in large parts of Brazil, Chile and Uruguay (Barteling and Sutmoller, 2002). Under the current OIE regulations, the fast and aggressive use of vaccination, which maximises the usefulness of its deployment, is unattractive to decision makers who must minimise the length of the outbreak to rapidly return to trade (Roche et al., 2015). This must be weighed against the impacts of a VTD policy.

Under specific circumstances, it is possible to practice zonal vaccination for FMD and for remainder of the country to be declared free of disease without vaccination by the OIE, thus allowing international trade in animal products. This is the case in South Africa where the South African Territory (SAT) strains of FMD are endemic in the Kruger National Park and surrounding reserves. These strains are maintained in Africa Buffalo *(Syncerus caffer)* and the disease is successfully contained by the maintenance of game proof fences, vaccination of susceptible cattle in the directly proximal area (known as the buffer zone) and maintenance of a surveillance zone between the buffer and the rest of the country (Bruckner et al., 2004, 2002).

2.5 Eradication activities in FMD endemic countries

2.5.1 Progressive control pathway framework

To address the situation of FMD endemic countries, the OIE has created the progressive control pathway (PCP), which allows the countries to step toward freedom (Rweyemamu et al., 2008; Sumption et al., 2012) by participating in a 7 stage programme which can run over 30 years (Rweyemamu et al., 2008). The first stage requires the disease distribution and epidemiology in the country be described and assessed at the country level. Once this is adequately understood, movement control and vaccination can be sensibly instituted (stage 2). The purpose of stage 2 is to establish disease free zones and to develop strategies in both the free and infected areas. Stage 3 has the objective of supressing virus transmission by preventing clinical disease. During stage 4, the country applies to the OIE to be recognised as free from FMD with vaccination in designated zones. In stage 5 the objective is to transition the designated zone from free with vaccination to free without vaccination. This is followed by expansion of the free zones (stage 6) and prevention of reintroduction of FMD (stage 7). The South-East Asia and China FMD (SEACFMD) Campaign implements the Food and Agriculture Organisation of the United Nations (FAO) and World Organisation for Animal Health (OIE) Global FMD Control Strategy based on the Progressive Control Pathway (PCP) for Foot and Mouth Disease (Jamal and Belsham, 2013; OIE and FAO, 2012; OIE World Organization for Animal Health, 2016).

For those countries where FMD is endemic the PCP offers a framework to move toward disease freedom. Stage 3 of the PCP suggests the use of vaccination to limit clinical disease expression and production losses in particular sectors or regions. Current examples are the pig industry in the Philippines and Thailand (Gleeson, 2002a). The objective of the PCP is for small regions to be expanded over time so that the countries are in time, able to become FMD free (Abila and Foreman, 2006). Some countries vaccinate in endemic zones to control spread to free areas. This is the case in South Africa where SAT strains are endemic in wildlife (Bruckner et al., 2004). In these cases, the prioritisation of resources is essential if zones are to be established and successfully maintained. In those countries with endemic disease, the value of vaccination should be examined by collection of information on household level economic factors especially where vaccination is being done outside of commercial enterprises. The value of vaccination in dairy cattle has been documented in Vietnam at the householder level at a benefit to cost ratio of 3 (CI 0.76-7.19) and even less for beef cattle in the same country (Truong et al., 2018). The clinical signs of FMD in endemic regions where large numbers of carriers are present and endemic strains are circulating are not particularly well described in the literature (Bertram et al., 2018b) but are expected to be less dramatic than those seen in naïve populations (Kitching, 2002). In addition, there is widespread recognition that disease reports made by human observers are by no means perfect indicators of disease presence or absence – they vary based on the disease under investigation, the individual surveyed (e.g. knowledge and experience etc.) and the production system (Bellet et al., 2012a; Kenton L Morgan et al., 2014; Vergne et al., 2012). In addition to this, even in disease free countries, the presentation of FMD may be indistinguishable from endemic differentials (T.W. Bates et al., 2003).

For developing countries (many of which have endemic FMD), resource to perform disease testing on large scale and the logistical difficulty of obtaining reagents, and getting samples to laboratories can prove to be prohibitive. In these limited conditions, verbal disease reports may be used as a source of surveillance information and may provide additional insights about the extent of clinical disease that cannot be gained from serology alone (Goutard et al., 2015; Muellner et al., 2016; Robertson et al., 2010; Sawford, 2011). When the disease distribution in a country is uncertain, the direction of resource to the sector which will benefit most becomes challenging and the external validity of surveys is difficult to establish.

The testing strategy deployed during the delimiting, eradication and proof of freedom phases of an FMD outbreak will in part be informed by the sample numbers required and the capacity of the national laboratory to process large volumes of samples. A strategic decision must be made (preferably prior to any outbreak) relating to diagnoses based on clinical signs vs the presence of DNA or serological disease markers. Return to trade relies on the testing of large numbers of serum samples as specified by the OIE. In the case of FMD, while “vaccinate to live” policies are attractive from a compensation and disposal point of view the delay in return to trade and the extensive DIVA (differentiating infected from vaccinated animals) testing has to be considered (Caporale et al., 2016b; Longjam et al., 2011; Paton et al., 2014, 2009).

2.6 Data requirements for FMD preparedness

In the context of livestock diseases, the basic building blocks field operations and preparedness for disease response include knowing the locations, counts and movement patterns of all individual animals in the national herd (Forman et al., 2012; Riley, 2010). The availability of these data at the country level varies greatly from countries which have mandatory centrally held agricultural property and animal movement databases to countries where no information on animal movements or farm locations exist. Every permutation and combination of voluntary, mandatory, regional and national system exists between these two extremes (Cheneau et al., 2004). Even amongst those countries with central property databases there are those that have property boundaries recorded (as spatial polygons) and those which record the property centroids.

Countries with a single mandatory national farm and livestock population database available for TAD preparedness and response are limited to a relatively small number of developed countries. Most competent authorities must amalgamate data from various sources to populate disease spread models or augment their existing preparedness data with data collected during response operations for the purposes of disease control activities in the field. In some countries industry sectors would share information related to farm populations and locations at the time of an outbreak.

In multi-jurisdiction countries, field operations are generally managed at the regional/state/ provincial /territory level, while disease spread modelling for preparedness purposes typically happens at the federal level. This difference means that there may be disparities at different levels in access to datasets and that preparedness and operational activities may be done using different data. When national datasets are available, the data contained within the databases may not be up to date or the data may have been collected for purposes other than disease control, limiting its use in supporting epidemiological investigations (Honhold and Taylor, 2006; Jewell et al., 2015). As an alternative to these national data sets, cross-sectional surveys can be performed, but these are constrained by resources to perform them at a sufficiently large geographic scale and it may be difficult to extrapolate the findings from one or more regions or time periods to other situations. Expert opinion has frequently been used when data are not available and strategies for making the best use of expert opinion and eliciting unbiased estimates have been well described (Burgman, 2015). Techniques for combining survey data, national level information and local knowledge exist so that all sources may be included (Sumption et al., 2008).

In a country where primary sector exports make up a significant part of GDP, preparing for a catastrophic situation that an outbreak of FMD would present is essential. However reaffirmation of the value of investment in these activities is needed in a resource scarce economy. The United Kingdom Rapid Analysis and Detection of Animal-related Risks (RADAR) was established in the wake of the 2001 FMD outbreak when it became clear that population information available to the competent authority was not fit for purpose (Paiba et al., 2007). RADAR created a way to rationalise animal location, population, movement and disease data in a central data warehouse. Information on the internet suggests that this project is no longer fully supported and personal communications with the RADAR team by email confirm that funding has been scaled back[9]. A way to quantify the value of accurate and timely animal population and movement data is urgently required. While economic modelling that includes both farm and country level detail provides an indication of the cost of an incursion of disease, information on the additional cost that would be incurred if data are missing or inaccurate and the magnitude of the harm caused by this data is difficult to quantify (Forbes and van Halderen, 2014; Michael G. Garner et al., 2016; Kompas et al., 2017; Mitchell, 2003; Van Andel et al., 2018).

Animal movement databases enable the rapid tracing of animals to and from infected premises and are essential to identify population level movement patterns. These patterns include regional differences in movement volumes and high risk locations which could be targeted for surveillance activities. The alternate to having an national animal movement recording system is for interviewers to collect movement information at each infected premises for the purposes of disease outbreak response and for national level surveys to be completed to identify movement patterns (Sanson, 2005). These time consuming and manual nature of these activities had led to uptake of centralised animal movement databases. However, the benefits of national level movement databases can only be realised if the data are complete and well understood. Incomplete data can be used appropriately only if the limitations are known. In this case, additional time and resource will be required to amalgamate additional sources of information with the existing national records to effectively trace and contain disease spread. The success of animal movement databases require the full cooperation of the farming sector along with intelligent integration of the technological solutions at critical control points in the production system to streamline collection of information. Establishing such a system is resource intensive and demands that both the drivers for reporting of animal movements and the requirements of the farming sector and the data users are fully understood. Restrictive legislation, lack of consultation with system users and ignorance of farming practices will result in the failure of such a system or at the very least in complicated and expensive retro-fitting of solutions after implementation has already occurred. It is important to note that all ungulates are affected by FMD

---

[9] Email communication with Helen Gartner, Helen.Gartner@apha.gsi.gov.uk on 14 July 2015

virus and are able to infect individuals of other species. This means that maintaining an integrated single system for cattle, deer, sheep, pigs and goats will be mandatory for a successful FMD response.

2.7 Discussion

The daunting extent and complexity of preparedness and eradication activities for FMD are widely documented and the high costs and farm reaching consequences of these activities can be described in detail (Forbes and van Halderen, 2014; Honhold et al., 2004; Mansley et al., 2016; Wada et al., 2016b). Despite this, the suitability and accuracy of the national data sources used during such activities are seldom assessed or discussed (Honhold and Taylor, 2006; Jewell et al., 2015). Given that the data are time consuming and resource intensive to collect and maintain, the question arises of how much resource should be allocated to the activities surrounding their upkeep. When deciding how much to invest in their upkeep, it is important to understand how large the impact of inaccuracies on decision making could be (van Andel et al., 2018; Wada et al., 2016a). Inaccurate data causes delays to field operations which can extend the duration of an outbreak, cause confusion and distress to farmers and embarrassment to the competent authority. Undesirable as these are, the effect is acknowledged and potentially quantifiable (Carpenter et al., 2011). The effect of inaccurate data on preparedness activities is more difficult to quantify (Bradbury et al., 2017; Jewell et al., 2015; Probert et al., 2015; van Andel et al., 2018). If the impacts on both operations and preparedness could be adequately quantified, an estimate of the value of national farm movement, population and location databases could be obtained. This estimate could then be compared with the cost of the maintenance of such databases in a traditional cost-benefit analysis.

Such an analysis would need to take into account that ongoing upkeep of the national data resource is essential as it is required for preparedness activities. Although disease control field activities will be performed on an ongoing basis, however it is essential (in the case of an FMD free country) that those who perform these activities are familiar with them so that they can be performed quickly under stressful conditions. The best way to ensure this is for the participants to regularly practice these activities and for preparation to focus on the availability of the necessary data and development of appropriate systems, processes and capacity before an incursion. All of these tasks depend on a knowledge of the size and location of the animal populations so that adequate human resources can be deployed to efficiently perform the tasks at hand. Accurate information on how many livestock and what species are present is essential information at disease control centres and the information must be updated in real time to ensure that the right people and supplies are deployed on visits and so that targeting and prioritising of surveillance, vaccination, culling and cleaning and disinfection activities is optimised Control activities must be prioritised to fit within limited resources.

One way to improve data quality in a fragmented data landscape is to attempt the linking of different levels of nationally held data (Jewell et al., 2015; Paiba et al., 2007). While technically attractive as a solution, the costs, complexities of government levels and legal restrictions on the use of different data sources make the maintenance and linking of these data sets in real time from multiple sources unfeasible for most countries. The technical difficulties inherent in linking datasets which are collected for varying purposes and have inherent biases are complex and may in some cases not be possible to resolve (Paiba et al., 2007). An alternate to data linking is the use of statistical and geospatial techniques to simulate farm sites and animal counts from census and survey information for national preparedness (Burdett et al., 2015; Hollings et al., 2018, 2017; van Andel et al., 2017). This is a logical and responsible choice for the competent authority in the case of animal population information when it is not available elsewhere. In the case of animal movement information (where databases are not available to analyse), surveys and expert opinion must be used to fill the information gaps that exist.

Even with perfectly up to date and accurate national data available, any strategic policy choice in a country that does not have FMD can only rely on models which provide an estimate of variability and indicate the comparative outcomes of different control options (Kitching et al., 2006). The extent of the mismatches and inaccuracies in national level databases are most often only realised during a disease outbreak of national concern. The 2001 FMD epidemic in the UK came under intense scrutiny by the public and academics. Some authors have suggested that model based decisions may have been made differently if different animal demographic data were used to build the models (Savill et al., 2007; Tildesley et al., 2010; Tildesley and Ryan, 2012). Aside from these policy decisions, the impacts on field operations are well documented even if the size of the impact is not quantified (Honhold and Taylor, 2006).

2.8 Conclusions

This review has collected a cross section of epidemiological uses for nationally held animal demographic and animal movement information in FMD eradication programmes. Sufficient evidence exists to accept that animal demography and up to date animal movements are essential for the veterinary government authority to adequately perform its function both in an endemic and disease free situation. Urgent improvements to data systems are required to create an environment that encourages the contribution of essential data by producers and this must be supported by adequate investment by regulators to curate, store and streamline this resource. Only when there is a single coherent source of unbiased animal demographic and animal movement data can the biosecurity authority hope to fulfil its responsibilities.

Research chapters

CHAPTER 3

3 Predicting farm-level animal populations using environmental and socioeconomic variables

3.1 Abstract

Accurate information on the geographic distribution of domestic animal populations helps biosecurity authorities to efficiently prepare for and rapidly eradicate exotic diseases, such as Foot and Mouth Disease (FMD). Developing and maintaining sufficiently high-quality data resources is expensive and time consuming. Statistical modelling of population density and distribution has only begun to be applied to farm animal populations, although it is commonly used in wildlife ecology. We developed zero-inflated Poisson regression models in a Bayesian framework using environmental and socioeconomic variables to predict the counts of livestock units (LSUs) and of cattle on spatially referenced farm polygons in a commercially available New Zealand farm database, Agribase.

Farm-level counts of cattle and of LSUs varied considerably by region, because of the heterogeneous farming landscape in New Zealand. The amount of high quality pasture per farm was significantly associated with the presence of both cattle and LSUs. Internal model validation (predictive performance) showed that the models were able to predict the count of the animal population on groups of farms that were located in randomly selected 3km zones with a high level of accuracy. Predicting cattle or LSU counts on individual farms was less accurate. Predicted counts were statistically significantly more variable for farms that were contract grazing dry stock, such as replacement dairy heifers and dairy cattle not currently producing milk, compared with other farm types.

This analysis presents a way to predict numbers of LSUs and cattle for farms using environmental and socio-economic data. The technique has the potential to be extrapolated to predicting other pastoral based livestock species

3.2 Introduction

An exotic disease outbreak can be economically devastating for the primary production industries and related sectors in nations that have substantial primary sector export economies (Blake et al., 2003; Carpenter et al., 2011). The preparation for and rapid control of outbreaks of exotic animal disease are high priorities for the Competent Authorities of these countries.

Information on the size and distribution (i.e., the demography) of the farm animal populations at risk, and their proximity to sources of infection, are required to adequately prepare for and respond to these outbreaks (Woolhouse, 2003). Important aspects of planning for exotic disease outbreak management include: defining a strategy to reduce disease spread, vaccine bank resourcing, and identification of the locations of burial or pyre burning sites for disposal of animal carcases. Each of these activities relies on knowledge about the farm animal demography of the area in question (Tomassen, de Koeijer et al. 2002). Impacts of insufficient planning for disease outbreak management are numerous and may extend many years after the end of an outbreak as economic losses, ground water contamination and air pollution from carcase disposal sites (Gwyther et al., 2011; Joung et al., 2013).

Demography fundamentally influences how disease will spread (Buhnerkempe et al., 2014; Thibaud Porphyre et al., 2013b). Estimating the risk of spread of infection of contagious disease from known infected sites is important for prioritising interventions during an outbreak with finite disease control resources (Paul R. Bessell et al., 2010; Carpenter et al., 2004; Haydon et al., 2003; Jewell and Brown, 2014; Keeling et al., 2001; Thrusfield et al., 2005). Farm animal demographic information facilitates a more timely identification of sites that are at high risk of becoming infected and/or at risk of spreading infection and guides decisions on control strategies early in an outbreak when epidemic information is scarce (Paul R. Bessell et al., 2010; Hugh-Jones, 1972; Mansley, 2004; Sarandopoulos, 2015; Tomassen et al., 2002) . This prioritisation is critical because delays in identification of infected premises (IP) can have a major effect on epidemic impacts (Carpenter et al., 2011).

Disease spread simulation models, which use animal demographic information, are useful for planning and preparing for exotic animal disease outbreaks. Models of disease transmission in farm animal populations typically assume that infection rate is a function of farm-level characteristics and the distance between pairs and contact structure of farms. Thus the distribution of farms and populations on the farms at risk are critical components of the effective reproduction number, R_e (Dohoo et al., 2009, Porphyre et al., 2013). In the model, estimates of R_e indicate whether an outbreak is under control and can inform policy adjustments which are implemented in the field (Ferguson et al., 2001; Paine et al., 2010). If the population data are inaccurate the model predictions of disease spread and impact will be inaccurate in ways which are not readily quantifiable without some model based structure to capture their uncertainty.

Competent Authorities may maintain national animal registries of farm and livestock population distribution (Paiba et al., 2007). The maintenance of accurate data within central animal registries is an ongoing challenge; particular areas of difficulty include: recording of births and deaths, real time animal movements, and mismatches and inconsistencies between different sources of data in the repository (Jewell et al., 2015; Savill et al., 2007). Other Competent Authorities have addressed a lack of animal population data by simulating livestock populations within a country (Burdett et al., 2015) or globally (Robinson et al., 2014). Although simulated livestock population data are available on a global scale at 1 km^2 resolution from the Food and Agriculture Organization's Gridded Livestock of the World Project (Robinson et al., 2014), this is not sufficiently accurate at the farm level to plan disease control operations, which require knowledge of the numbers and species present on each IP (Mansley, 2004). An alternative approach, which is required in some countries for which details of farm locations and their animal populations are not available, is to model the location of farm sites and farm types based on environmental variables (Burdett et al., 2015).

New Zealand livestock farming is predominantly pasture based and the main stock categories farmed are dairy cattle, beef cattle, sheep and deer. These populations are represented in multiple repositories of animal demographic information (Jewell et al., 2015), which are maintained by the government and private industry bodies. One of these, AgriBase, is owned by AsureQuality Limited, a New Zealand state-owned enterprise that provides operational services to the Ministry for Primary Industries (MPI). MPI maintains a second properties database FarmsOnLine (FOL) for biosecurity purposes but livestock counts are not available in this data source. Both AgriBase and FOL hold spatially referenced information on properties used for all farming activities including those used for recreational farming or "lifestyle" farming as it is known in New Zealand. However, national databases are expensive to maintain, and require full engagement of the landowners to keep current both in terms of land parcels and livestock kept. The result is that there may be a high level of unquantified uncertainty around what we would otherwise like to regard as census data (Jewell et al., 2015). In this study, we investigate an approach to making use of sample-based data with known accuracy to predict counts of counts of livestock units (LSU's) and cattle present in production systems where animals live mainly outdoors and eat a predominantly pasture based diet on unsampled farms in New Zealand using environmental and socio-economic data. An LSU is a reference unit with allows various species of livestock to be aggregated based on nutritional requirement of each animal type; we have used the European Union definition of an LSU for this study (Anonymous, 2013b).

The balance of the paper is constructed as follows. In the next section we describe the data sets, data handling, and modelling strategy, and the means by which we assessed and compared the models. We then provide a detailed description of livestock prediction in New Zealand, with Discussion, and our Conclusion. Supplementary material to the materials and methods section is provided.

3.3 Materials and Methods

Databases and study variables:

The predictive model was built using farm boundary and animal population data from AgriBase (Sanson and Pearson, 1997) which contained 135 002 farms. AgriBase contains spatially referenced farm boundary data and species-specific counts of animal populations for each farm. Some farm boundaries in the dataset overlapped one another as they are farmed under cross-lease agreements. These 6586 (4.5% of the total available farm records) overlapping farms were deleted from the dataset to avoid bias in predicting animal counts. In this case both of the overlapping farms were deleted. We deleted a further 36815 properties (27.02% of the total available farm records) that had no information on production type of any kind (agriculture or horticulture based) because we assumed that these represent missing data rather than the true absence of the animal species of interest. The remaining 91601 farms (68.50%) were used for model building and validation.

The two outcome variables of interest were farm level counts of LSUs and of cattle; both were drawn from the AgriBase dataset (Sanson and Pearson, 1997). LSUs were calculated for sheep, goats, cattle, pigs, deer and horses following the European Union definition of individual species requirements in relation to a dairy cow (Anonymous, 2013b) shown in Table 1. Poultry data were not included in the model.

Table 1: *Species equivalence values used to calculate livestock units (LSUs) on rural properties in New Zealand following the European Union definition of individual species feed requirements (Anonymous, 2013b).*

	Livestock Units (LSUs) relative to a grazing dairy cow
Dairy cattle	1
Beef cattle	0.8
Deer	0.8
Sheep & Goats	0.1
Pigs	0.4
Horses	0.8

Publicly available land cover, environmental and socio-economic data held by Land Information New Zealand (LINZ), Statistics New Zealand, Land Environments of New Zealand (LENZ) and Land Resources Information New Zealand (LRINZ) were used to generate candidate predictor variables (Table 2). The New Zealand Transverse Mercator (NZTM) projection was used for geographic information projection in this analysis. Predictor variables were loaded into a geographic information system (GIS) relational database (PostgreSQL 8.4 with PostGIS 2.2.0 extensions). This provided a convenient platform for interfacing with statistical software (Rv3.0.1, Foundation for Statistical Computing, Vienna, Austria), enabling large amounts of geospatial data to be processed and extracted for each farm. Predictor variables included in the models, where they were sourced and how they were derived are presented in Table 2.

Table 2: *Predictor variables included in a Bayesian Zero Inflated Poisson model predicting counts of cattle and counts of livestock units on New Zealand farms.* Predictor variables included in the final models. Δ Predictor variables significant in the LSU model. ¥ Predictor variables significant in the cattle model*

Predictor Variable	**Description (units)**	**Reference**
Distance to sealed roads *Δ¥	Distance from the centroid of the farm to the nearest major sealed road (km)	Land and Information NZ (LINZ) https://data.linz.govt.nz/
Distance to urban centres *Δ¥	Distance from the centroid of the farm to the nearest urban centre (km)	Land and Information NZ (LINZ) https://data.linz.govt.nz/
Distance to rivers *Δ¥	Distance from the centroid of the farm to the nearest river (km)	Land and Information NZ (LINZ). https://data.linz.govt.nz/
Social deprivation index *Δ¥	Value at the farm centroid (categorical variable between 1 and 10 with 1 being the greatest degree of deprivation)	Statistics NZ census data. Decile measure of socio-economic status
Interaction term*	Interaction term between deprivation index and distance to urban centres to account for the effect of lifestyle farming	
Area of high quality pasture (log of) *Δ¥	Area of grassland per farm classified as high producing by the national land cover database (ha)	Land Cover Database (LCDB) Version 4. A satellite derived GIS vector polygon layer of land use. https://lris.scinfo.org.nz
Area of low quality pasture (log of)*Δ¥	Area of grassland per farm classified as poorly producing grassland by the national land cover database (ha)	Land Cover Database (LCDB) Version 4. A satellite derived GIS vector polygon layer of land use. https://lris.scinfo.org.nz
Mean annual temperature	The average annual temperature across the farm calculated by averaging each raster cell of a modelled temperature surface contained within the farm polygon	Land and Environments NZ (LENZ) layer (Leathwick et al. 2002). www.koordinates.com
Mean minimum temperature	The average minimum temperature across the farm calculated by averaging each raster cell of a modelled temperature surface contained within the farm polygon	Land and Environments NZ (LENZ) layer (Leathwick et al. 2002). www.koordinates.com
Temperature difference*	Mean annual temperature minus the mean minimum temperature for each farm	
Mean solar radiation	The annual average solar radiation across the farm calculated by averaging each raster cell of a modelled solar radiation surface contained within the farm polygon	Land and Environments NZ (LENZ) layer (Leathwick et al. 2002). www.koordinates.com
Mean winter solar radiation	The average minimum daily solar radiation across the farm calculated by averaging each raster cell of a modelled solar radiation surface contained within the farm polygon	Land and Environments NZ (LENZ) layer (Leathwick et al. 2002). www.koordinates.com
Solar difference*Δ¥	The difference between the mean solar radiation and the mean winter solar radiation for each farm	

Mean water deficit*¥	The average of the annual water deficit across each farm calculated by averaging each raster cell of a modelled water deficit surface contained within the farm polygon	Land and Environments NZ (LENZ) layer (Leathwick et al. 2002). www.koordinates.com
Mean aspect	Mean of farm aspect or the direction in degrees which a slope faces (degrees)	
Mean slope	Mean of farm slope (slope is a unitless ratio)	
Slope x Aspect*	Adjustment to slope and aspect to decrease collinearity	(Stage, Salas 2007)
Sine of Slope*Δ¥	Adjustment to slope and aspect to decrease collinearity	(Stage, Salas 2007)
Cosine of Slope*Δ¥	Adjustment to slope and aspect to decrease collinearity	(Stage, Salas 2007)
Standard deviation of slope*Δ¥	Standard deviation of slope for the farm	
Mean and standard deviation of elevation*Δ¥	Mean and standard deviation of elevation for the farm calculated by averaging each raster cell of a surface generated by a digital elevation model contained within the farm polygon (m)	Digital elevation model from Land and Information NZ (LINZ). https://data.linz.govt.nz/
Regional Council*Δ¥	One of the 17 New Zealand Regional councils that the farm centroid falls within (categorical variable)	https://koordinates.com/layer/4240-nz-regional-councils-2012-yearly-pattern/
Size of farm	Farm area (ha)	Agribase™ for the farm boundaries and post gis
Forest cover (log of)*Δ¥	The sum of endemic and exotic forest cover for each farm (ha)	https://koordinates.com/layer/168-nz-native-polygons-topo-150k/ & https://koordinates.com/layer/131-nz-exotic-polygons-topo-150k/
Remaining land cover (log of)* Δ¥	Remaining area of the farm once high quality pasture, low quality pasture and forest cover had been taken from the total (ha)	

Spearman correlation matrix was used to identify significant pairwise collinearity amongst the predictor variables, defined arbitrarily as a correlation of $|r| > 0.7$. Our choice to use the Spearman's correlation coefficient rather than a variable selection process within the modelling framework allowed the analysts to make variable selections based on knowledge of the biological system under examination. Collinearity was identified between the following pairs of variables: mean *annual temperature* and mean *minimum temperature*, mean *solar exposure* and minimum *solar exposure,* and *slope* and *aspect*. Collinearity was addressed by creating bivariate transformations to include in the model and by retaining one of each of the collinear pairs (see supplementary material equations 1 and 2 for further detail of the bivariate transformations).

An interaction term was included to represent the relationship between *deprivation index* and *distance from urban centres* to capture the population of lifestyle farms. Statistics New Zealand notes that this population of lifestyle farms in rural areas with high urban influence has the highest median and mean income (and household expenditure) of any demographic in New Zealand (Pink, 2011). Land held by this population demographic, which commutes to city centres for work, is thus unlikely to be used in the same way as land further from cities, where agriculture is the principal source of income. The log of the area in hectares of each farm covered by *high quality pasture*, *low quality pasture*, *native forest* and *endemic forest* and the complement of the farm area were used in the final model. Each of these was offset by 0.01 to make the log transformation possible for farms with no coverage of some of the categories. The log transformation of these variables was used in the modelling to promote for numerical stability.

Statistical model framework

A zero-inflated Poisson (ZIP) regression model was used to generate the predicted counts of LSUs and of cattle for farm properties. A ZIP model was chosen because many rural properties are not engaged in livestock production; AgriBase records 37% of properties as having no animals present (Jewell et al., 2015). Consequently, the ZIP model fit the data significantly better than did a comparable Poisson model.

The model was trained to predict farm level populations using a Bayesian approach that calculated a distribution of the predicted number of cattle or LSUs within a farm polygon on the land surface in New Zealand. The outcome variable for an individual farm (either number of cattle or LSUs) was modelled as a ZIP random variable, conditional on the values of the predictor variables for that farm (Mullahy, 1986). A two-stage process was thus used, in which (i) a Bernoulli random variable first represented the probability of animals being present on a farm; (ii) then, assuming that animals are present, the number of animals was assumed to follow a Poisson distribution with mean related via a log link function to the predictor variables. Both the Bernoulli and the Poisson stages involved modelling using the suite of predictor variables. Further detail of the model framework, model equations and model code are available in the supplementary materials.

The vectors of coefficients are calculated for the same set of predictor variables for the both the Poisson and Bernoulli components. We have allowed the prediction of the excess zeros and the prediction of the counts on each farm to be conditional on the same predictor variables, however as Table 4 shows, different predictors are identified as important in the Poisson regression step when compared with the Bernoulli step, thus $\boldsymbol{x}_i$ (equation 5 and 6 in the supplementary material) appears in the linear predictors for both the Poisson and the Bernoulli steps of the model. The importance of each covariate in the final model was explored using reversible jump MCMC which is fully described in the supplementary material (Carlin and Chib, 1995; Green, 1995).

The MCMC was coded in R using a single-site adaptive Metropolis-Hastings step for each coefficient (Chib and Greenberg, 1995; Jewell, 2016a). A total of 50 000 iterations of each model was performed and the first 20 000 runs were discarded as burn-in leaving 30 000 iterations which were thinned systematically every 30 iterations to 1000 observations to be used in all further analyses.

Predictive performance:

To assess how well the model was able to predict the data which it was built from, a frequentist model residual analogue was calculated (rank probability score, referred to hereafter and in the supplementary material as RPS) and five “residual”, i.e., RPS checking analyses were completed. The statistical detail of the calculation of the RPS and the analyses performed on this residual analogue are explained in detail in the supplementary material.

The first two methods of residual (RPS) checking were: i) an examination of spatial autocorrelation of RPS in two large regions (Figure 1 shows the regions and Figure 2 the variograms), and ii) the differences in model performance amongst farm types (Figure 3). In these analyses the farm is the unit of interest. The third and fourth methods were performed at iii) small (3km radial zone) and iv) larger (New Zealand administrative regions) spatial scales and calculated area level RPS values (Table 3 and Figure 4). In these analyses the area (either the region or the 3km zone) is the primary unit of interest. The final method v) used Receiver Operator Characteristic (ROC) curves (Figure 5) to examine the performance of the Bernoulli component of the ZIP model (which predicts the presence or absence of livestock or cattle on the farm of interest).

Figure 1: *Map showing the 16 regions and the two main islands of New Zealand. Grey shaded areas represent the areas used to compute variograms in the North Island and South Island to evaluate spatial autocorrelation of model fit shown in this case by the rank probability score.*

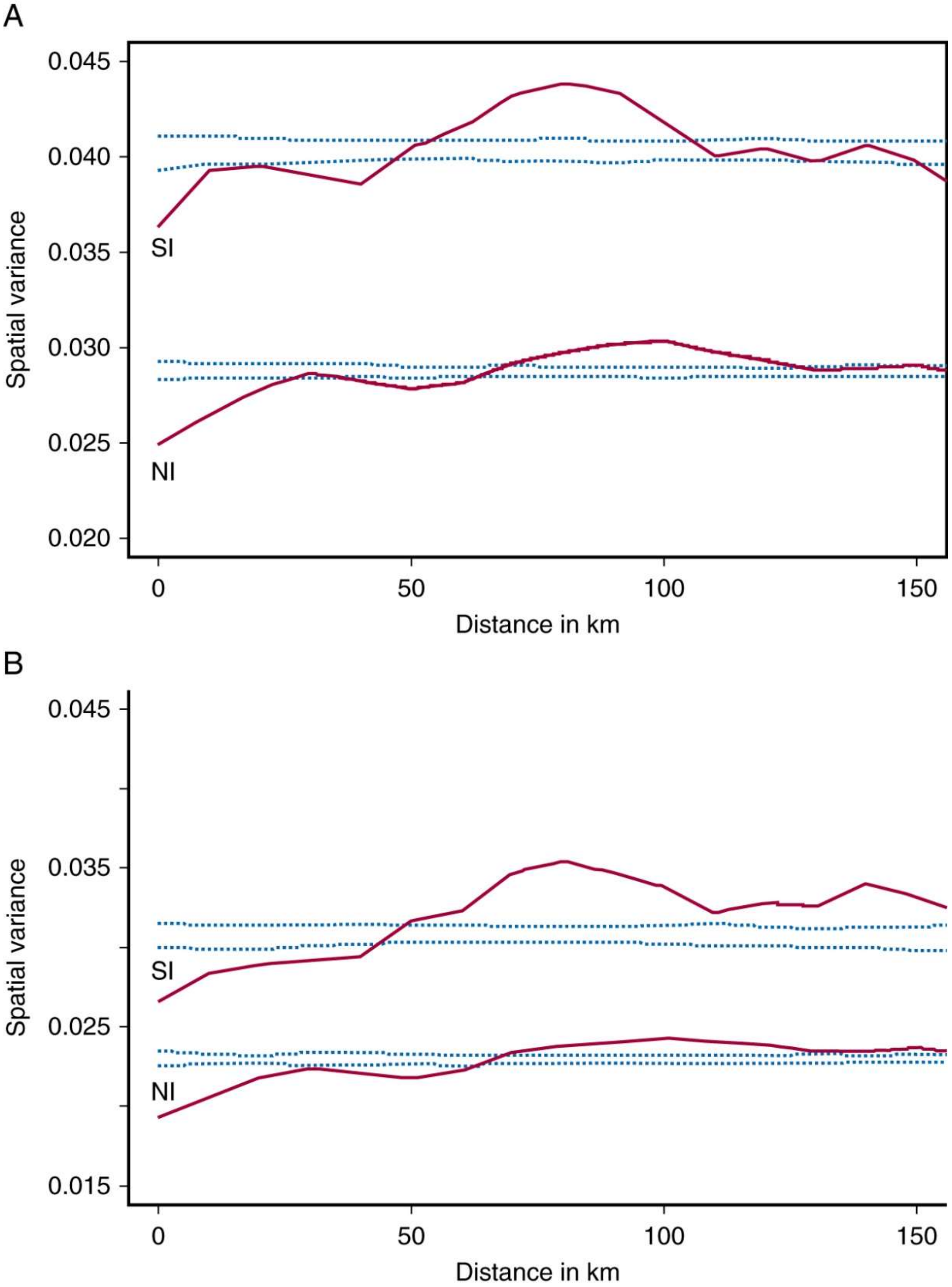

Figure 2: *Variograms for the LSU model (A) and the cattle model (B) showing the spatial covariance in the farm-level Rank Probability Score (RPS) in areas presented in Figure 1 in the North Island (NI) and the South Island (SI). The dotted lines represent the 5% and 95% of the prediction interval estimated by calculating a variogram for 100 random permutations of the real RPS values for each region. The distance at which the errors stop being correlated is read off where the solid lines cross into the 95% prediction intervals (dotted lines) for each island and for each model.*

Table 3: *Mean and standard deviation of rank probability scores (RPS) for models predicting livestock units (LSUs) and cattle on farms within 3 km zones and administrative regions in New Zealand.*

Target data predicted	**Number of fits**	**Level of RPS calculation**	**LSU model**		**Cattle model**	
			mean RPS	**Standard deviation**	**mean RPS**	**Standard deviation**
200 3km zones around randomly selected farms	10	farm	0.221	0.006	0.214	0.014
	10	area (3km zone)	0.093	0.009	0.162	0.0122
3 North Island and 2 South Island regions	50	farm	0.317	0.062	0.247	0.033
	50	area (administrative region)	0.261	0.227	0.265	0.219

We chose 3km radial zones as a unit of interest to examine model performance because the outcomes (farm level counts of cattle or LSUs) of these models are primarily required for biosecurity purposes. In the event of a suspected or actual outbreak of FMD, the population in the 3km radial zone around an infected farm is at high risk of local spread (Thrusfield et al., 2005). This means that accurate data on susceptible animal populations on farms within 3km of a farm with a suspected outbreak of FMD should be available as early as possible in the investigation and risk assessment process. In the New Zealand context, a 3km zone is an appropriate choice based on local observations and farm density.

3.4 Results

Predictor variables

Predicted coefficients of covariates for the Bernoulli and Poisson components of the ZIP models and their associated odds ratios and risk ratios are presented in Table 4. The most important covariate for predicting the presence or absence (Bernoulli component of the ZIP model) of cattle and of LSUs was the log of the area (hectares) of high quality pasture followed by region in which the farm was located. Region was the most important predictor variable for counts of LSUs and cattle (the Poisson part of the models). The use of the zero-inflated distribution allows the excess appearance of zeros to be linked to the covariates. As an example, consider the Bernoulli component of the cattle model. As shown in the model covariate table presented in Table 4, the size of the area of high quality pasture present on the farm of interest is predictive of the presence of cattle (OR 1.91, CI between 1.91 and 1.92), however some regions have additional increases in the likelihood of cattle being present (multiplicative at the log scale with the covariate representing the size of high quality pasture). The difference between the covariates for the regional fixed effects for Northland and Nelson illustrate this point well. If the Manawatu- Wanganui region (a region with a large amount of farm land and a sizeable cattle population) is taken to be the referent category, then the Nelson region is half as likely (OR of 0.52 with CI between 0.5 and 0.54 to have cattle present on a particular farm of interest, and farms in the Northland region are 1.77 times more likely with CI between 1.75 and 1.83 to have cattle present on the property. No evidence to remove any of the covariates from either the Poisson or Bernouilli parts of the model was found when the covariates were explored using reversible jump MCMC. We have included the OR for the interaction terms in Table 4 even though the OR for an interaction term cannot be interpreted without taking into consideration the OR for both of the terms involved in the interaction. Table 4 is a summary of the joint posterior distribution over the coefficients. It displays marginal information for each coefficient and ignores correlations between the respective marginal posterior distributions, and the fact that for some realisations of the posterior the value of certain coefficients may be 0.

Table 4: *Summary of the mean and standard deviation of the joint posterior estimates of covariates for species distribution models for New Zealand farms using counts of livestock units (LSU) as the outcome variable in the first instance and counts of cattle in the second instance. Risk ratios are reported for the Poisson component and odds ratios for the Bernoulli component for each model along with 95% credible intervals.*

	LSU model			**Cattle model**		
Bernoulli component						
Predictor variable	**mean**	**std dev**	**OR (5%-95%CI)**	**mean**	**std dev**	**OR (5%-95%CI)**
intercept	3.353	0.007	28.58 (28.48 - 28.68)	3.074	0.012	21.62 (21.53 - 21.79)
Other land cover in hectares (log)	0.013	0.000	1.01 (1.01 - 1.01)	0.008	0.000	1.01 (1.01 - 1.01)
High quality pasture in ha (log)	0.628	0.002	1.87 (1.87 - 1.88)	0.649	0.003	1.91 (1.91 - 1.92)
Low quality pasture in ha (log)	0.009	0.000	1.01 (1.01 - 1.01)	0.006	0.000	1.01 (1.01 - 1.01)
Forest cover in ha (log)	0.014	0.000	1.01 (1.01 - 1.01)	0.009	0.000	1.01 (1.01 - 1.01)
Distance to a road from farm centroid	0.050	0.000	1.05 (1.05 - 1.05)	0.072	0.001	1.07 (1.07 - 1.08)
Distance to an urban centre from farm centroid	0.006	0.000	1.01 (1.01 - 1.01)	0.003	0.000	1 (1 - 1)
Temperature difference	0.006	0.000	1.01 (1.01 - 1.01)	0.004	0.000	1 (1 - 1)
Solar difference	0.011	0.000	1.01 (1.01 - 1.01)	0.000	0.001	1 (1 - 1)
Distance to a river from farm centroid	0.007	0.000	1.01 (1.01 - 1.01)	0.007	0.000	1.01 (1.01 - 1.01)
Mean slope	-0.029	0.000	0.97 (0.97 - 0.97)	-0.047	0.001	0.95 (0.95 - 0.96)
Standard deviation of slope	-0.015	0.000	0.99 (0.98 - 0.99)	-0.019	0.000	0.98 (0.98 - 0.98)
Mean digital elevation model	-0.001	0.000	1 (1 - 1)	-0.002	0.000	1 (1 - 1)
Standard deviation of digital elevation model	0.002	0.000	1 (1 - 1)	0.002	0.000	1 (1 - 1)

Cosine of slope	0.051	0.000	1.05 (1.05 - 1.05)	0.049	0.001	1.05 (1.05 - 1.05)
Sine of slope	0.008	0.001	1.01 (1.01 - 1.01)	0.006	0.001	1.01 (1 - 1.01)
Product of slope and aspect	0.000	0.000	1 (1 - 1)	0.000	0.000	1 (1 - 1)
Deprivation index	0.022	0.000	1.02 (1.02 - 1.02)	0.027	0.001	1.03 (1.03 - 1.03)
Deprivation index * Distance to an urban centre	-0.001	0.000	1 (1 - 1)	-0.002	0.000	1 (1 - 1)
Mean water deficit	-0.002	0.000	1 (1 - 1)	0.019	0.002	1.02 (1.02 - 1.02)
Mean solar hours	-0.009	0.000	0.99 (0.99 - 0.99)	-0.024	0.002	0.98 (0.97 - 0.98)
Mean temperature	-0.001	0.000	1 (1 - 1)	-0.001	0.000	1 (1 - 1)
Fixed effects for region (Manwatu-Wanganui as the referent category)						
Auckland	-0.159	0.004	0.85 (0.85 - 0.86)	0.145	0.012	1.16 (1.14 - 1.18)
Bay of Plenty	0.204	0.004	1.23 (1.22 - 1.23)	0.333	0.008	1.4 (1.38 - 1.41)
Canterbury	-0.047	0.004	0.95 (0.95 - 0.96)	-0.055	0.016	0.95 (0.92 - 0.96)
Gisborne	0.548	0.004	1.73 (1.72 - 1.74)	0.764	0.008	2.15 (2.13 - 2.16)
Hawkes Bay	0.106	0.003	1.11 (1.11 - 1.12)	0.130	0.004	1.14 (1.13 - 1.14)
Marlborough	-0.359	0.005	0.7 (0.69 - 0.7)	-0.361	0.016	0.7 (0.68 - 0.71)
Nelson	-0.718	0.016	0.49 (0.47 - 0.5)	-0.660	0.025	0.52 (0.5 - 0.54)
Northland	0.169	0.004	1.18 (1.18 - 1.19)	0.573	0.016	1.77 (1.75 - 1.83)
Otago	-0.344	0.004	0.71 (0.71 - 0.71)	-0.586	0.021	0.56 (0.53 - 0.57)
Southland	-0.288	0.003	0.75 (0.75 - 0.75)	-0.547	0.023	0.58 (0.55 - 0.59)
Taranaki	0.179	0.002	1.2 (1.19 - 1.2)	0.340	0.005	1.4 (1.4 - 1.41)
Tasman	-0.459	0.005	0.63 (0.63 - 0.64)	-0.310	0.015	0.73 (0.71 - 0.75)
Waikato	0.257	0.003	1.29 (1.29 - 1.3)	0.447	0.007	1.56 (1.56 - 1.58)
Wellington	-0.066	0.003	0.94 (0.93 - 0.94)	-0.157	0.004	0.85 (0.85 - 0.86)
West Coast	-0.341	0.004	0.71 (0.71 - 0.71)	-0.179	0.010	0.84 (0.82 - 0.85)

	LSU model			**Cattle model**		
Poisson component	**mean**	**std dev**	**RR (5%-95%CI)**	**mean**	**std dev**	**RR (5%-95%CI)**
intercept	-2.438	0.046	0.09 (0.08 - 0.09)	-1.394	0.036	0.25 (0.23 - 0.26)
Other land cover in hectares (log)	0.022	0.002	1.02 (1.02 - 1.02)	0.004	0.001	1 (1 - 1.01)
High quality pasture in ha (log)	-0.040	0.003	0.96 (0.96 - 0.97)	-0.086	0.003	0.92 (0.91 - 0.92)
Low quality pasture in ha (log)	-0.015	0.003	0.98 (0.98 - 0.99)	-0.018	0.002	0.98 (0.98 - 0.99)
Forest cover in ha (log)	0.009	0.002	1.01 (1.01 - 1.01)	-0.002	0.001	1 (1 - 1)
Distance to a road from farm centroid	-0.049	0.027	0.95 (0.91 - 0.99)	-0.199	0.027	0.82 (0.78 - 0.86)
Distance to an urban centre from farm centroid	-0.011	0.002	0.99 (0.99 - 0.99)	-0.025	0.002	0.98 (0.97 - 0.98)
Temperature difference	0.001	0.002	1 (1 - 1)	0.000	0.001	1 (1 - 1)
Solar difference	0.020	0.008	1.02 (1.01 - 1.03)	0.038	0.006	1.04 (1.03 - 1.05)
Distance to a river from farm centroid	-0.005	0.003	1 (0.99 - 1)	-0.011	0.002	0.99 (0.99 - 0.99)
Mean slope	0.107	0.006	1.11 (1.1 - 1.12)	0.123	0.006	1.13 (1.12 - 1.14)
Standard deviation of slope	-0.160	0.008	0.85 (0.84 - 0.86)	-0.145	0.006	0.86 (0.86 - 0.87)
Mean digital elevation model	0.002	0.000	1 (1 - 1)	0.003	0.000	1 (1 - 1)
Standard deviation of digital elevation model	-0.008	0.001	0.99 (0.99 - 0.99)	-0.013	0.001	0.99 (0.99 - 0.99)
Cosine of slope	0.032	0.016	1.03 (1.01 - 1.06)	0.007	0.012	1.01 (0.99 - 1.03)
Sine of slope	0.001	0.017	1 (0.97 - 1.03)	-0.036	0.014	0.96 (0.94 - 0.99)
Product of slope and aspect	0.000	0.000	1 (1 - 1)	0.000	0.000	1 (1 - 1)
Deprivation index	-0.006	0.006	0.99 (0.98 - 1)	-0.044	0.005	0.96 (0.95 - 0.96)

Deprivation index * Distance to an urban centre	0.002	0.000	1 (1 - 1)	0.001	0.000	1 (1 - 1)
Mean water deficit	-0.002	0.008	1 (0.99 - 1.01)	-0.011	0.007	0.99 (0.98 - 1)
Mean solar hours	0.045	0.006	1.05 (1.04 - 1.06)	0.037	0.005	1.04 (1.03 - 1.05)
Mean temperature	0.001	0.000	1 (1 - 1)	0.003	0.000	1 (1 - 1)
fixed effects for region (Manwatu-Wanganui as the referent category)						
Auckland	0.057	0.088	1.06 (0.92 - 1.23)	0.213	0.068	1.24 (1.11 - 1.39)
Bay of Plenty	0.954	0.085	2.6 (2.27 - 2.99)	0.453	0.068	1.57 (1.41 - 1.77)
Canterbury	1.085	0.078	2.96 (2.6 - 3.37)	1.469	0.061	4.35 (3.93 - 4.81)
Gisborne	0.300	0.118	1.35 (1.11 - 1.63)	0.386	0.096	1.47 (1.26 - 1.73)
Hawkes Bay	0.683	0.070	1.98 (1.76 - 2.23)	0.576	0.057	1.78 (1.62 - 1.96)
Marlborough	1.156	0.096	3.18 (2.71 - 3.72)	1.283	0.084	3.61 (3.14 - 4.14)
Nelson	0.627	0.215	1.87 (1.32 - 2.66)	0.870	0.182	2.39 (1.77 - 3.22)
Northland	0.026	0.116	1.03 (0.85 - 1.24)	-0.089	0.091	0.92 (0.79 - 1.07)
Otago	1.661	0.108	5.27 (4.41 - 6.29)	2.141	0.084	8.5 (7.41 - 9.81)
Southland	1.409	0.127	4.09 (3.31 - 5.05)	2.240	0.093	9.4 (8.07 - 10.95)
Taranaki	0.214	0.068	1.24 (1.11 - 1.39)	-0.226	0.054	0.8 (0.73 - 0.87)
Tasman	1.373	0.099	3.95 (3.35 - 4.63)	1.224	0.079	3.4 (2.99 - 3.88)
Waikato	0.209	0.073	1.23 (1.09 - 1.39)	-0.047	0.057	0.95 (0.87 - 1.05)
Wellington	0.625	0.065	1.87 (1.68 - 2.08)	0.664	0.049	1.94 (1.79 - 2.1)
West Coast	1.691	0.098	5.42 (4.6 - 6.37)	1.643	0.077	5.17 (4.56 - 5.87)

Predictive performance

The primary measure of the performance of our models is how well they predicted counts of LSUs or cattle at various spatial scales as measured by the RPS. Results for internal validation at each of the five scales are presented below.

i) The spatial variogram of farm-level RPSs showed slight spatial autocorrelation in the LSU model over short distances in both the North Island and South Island test regions (Figure 2). This correlation was apparent up to about 30km in the North Island and about 50km in the South Island for the LSU model. In contrast, in the cattle model in the North Island, correlation of the RPS persisted up to 60km and in the South Island it reached a sill at around 50km. Autocorrelation in RPS was consistently larger in the South Island when compared with the North Island for the LSU model indicating a higher level of spatial correlation of model error.

ii) Significant differences in mean RPS values were found for groups of farms stratified by the farm type recorded in AgriBase (dairy farms, pastoral livestock farms, pig breeder and finisher farms, grazing and dairy heifer rearing properties and other farm types) when tested with the Kruskal-Wallis test ($p < 0.001$ with 6 degrees of freedom) for both the LSU and the cattle models. In both models, the farm type 'dry stock' which included contract grazing and dairy heifer rearing properties were predicted with the least accuracy (Figure 3).

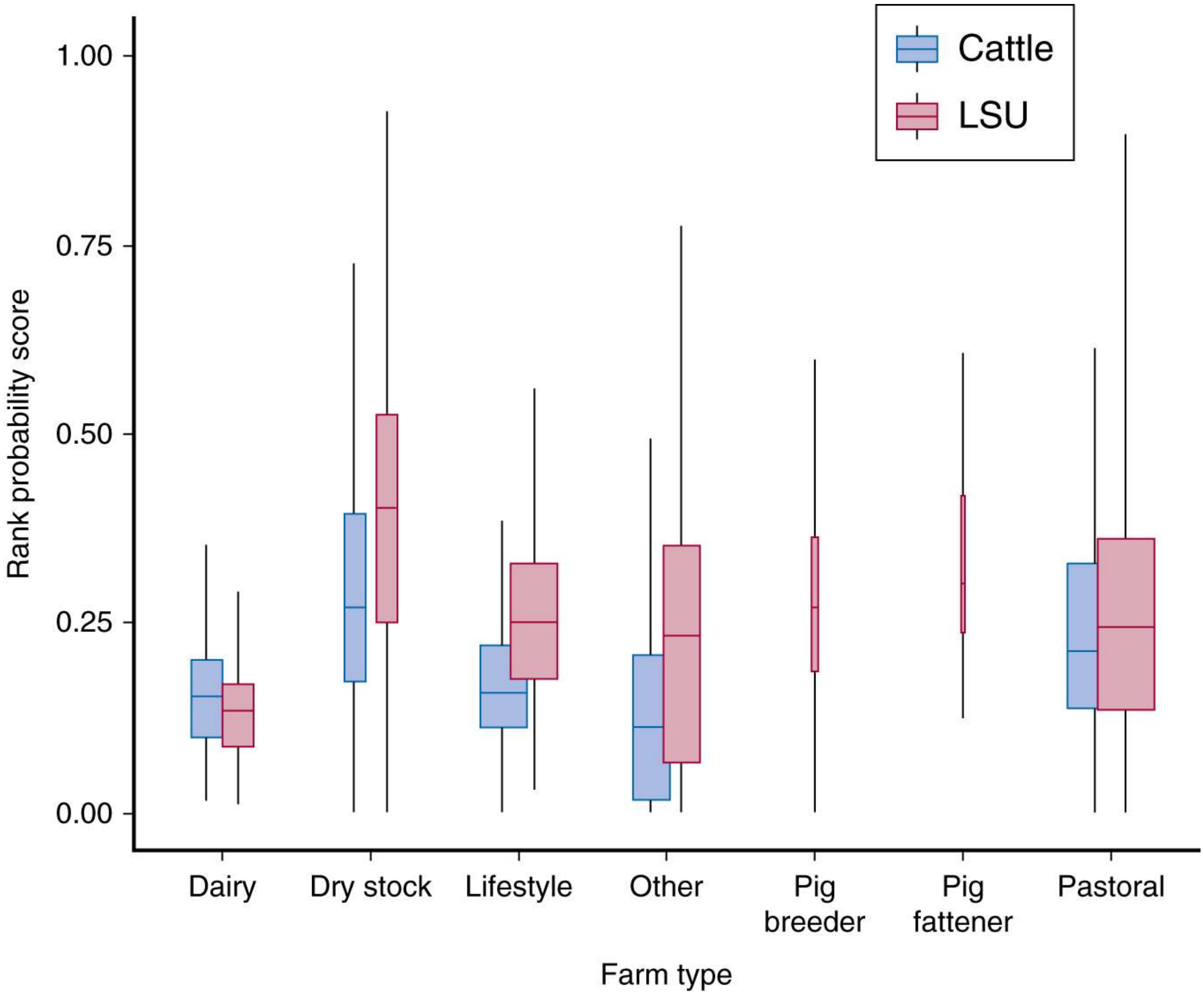

Figure 3: *Boxplot of RPS for each of the 7 farm types in AgriBase for models predicting livestock units (light grey) and 5 farm classes in AgriBase for models predicting cattle (dark grey) on rural properties in New Zealand. This figure shows that in both models predictions for dry stock (contract grazing and dairy heifer rearing properties) were least accurate. Box widths are proportional to the square root of the number of observations present in each group.*

iii) Prediction at the 3km zone level was most accurate with mean RPS of 0.093 for the LSU model and 0.162 for the cattle model summarised over the 10 model fits (Table 3).

iv) When region was removed from the model to allow internal validation between regions, the model that predicted cattle numbers performed better overall than the model that predicted LSUs (mean RPS of 0.317 and standard error of 0.062 for the LSU model compared with a mean of 0.247 and standard error 0.033 for the cattle model) (Table 3). When regional RPSs from each model run were aggregated and plotted a large amount of variability in the RPSs was observed at the region level (Figure 4).

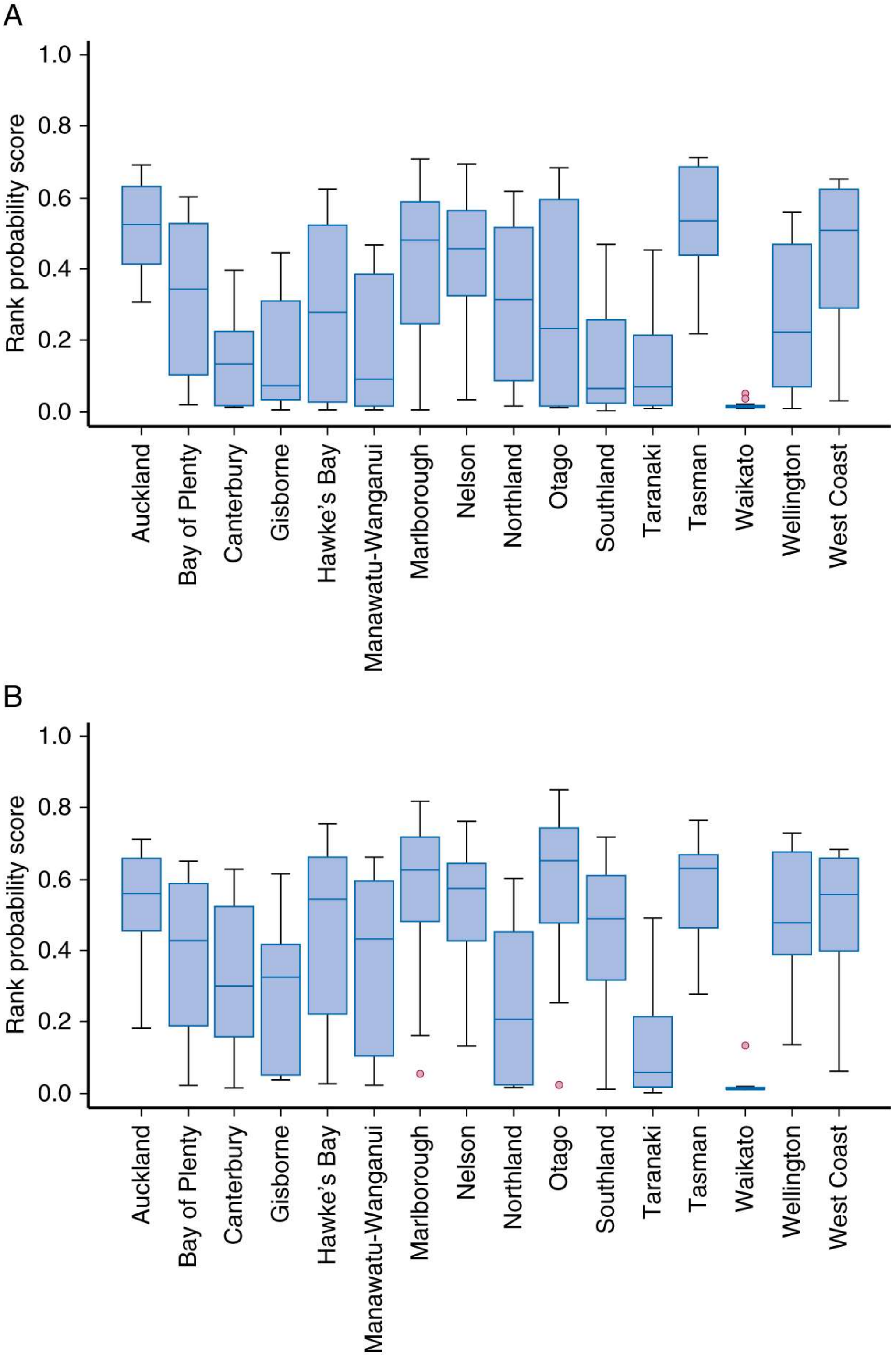

Figure 4: *Box plots of the RPS from 50 resampled fits of the model where region was removed from A, the LSU model and B the cattle model. The upper bound of each box is the 75th percentile, lower bound is the 25th percentile, the median is indicated within the box and the dotted whiskers terminate at the 95th and 5th percentiles. Outliers are shown as hollow circles.*

v) The area under the curve (AUC) was estimated for the models and thresholds that maximised both sensitivity and specificity. As an example, for the model that predicted LSUs, when the cut-off for representing the posterior prediction of having no animals on each property was 0.25, 90% of farms with no animals were correctly identified (specificity). The sensitivity of this estimate was poor (40%) with many false negative results (i.e. farms identified by the model as having no animals, which in fact had animals present). The largest area under the curve (AUC) for the LSU model was found at a point mass of zeros of 0.15, representing correct identification of 68% of the farms without animals. In the cattle model the area under the curve was maximised at a cut-off of 0.35 where 68% of the area was captured.

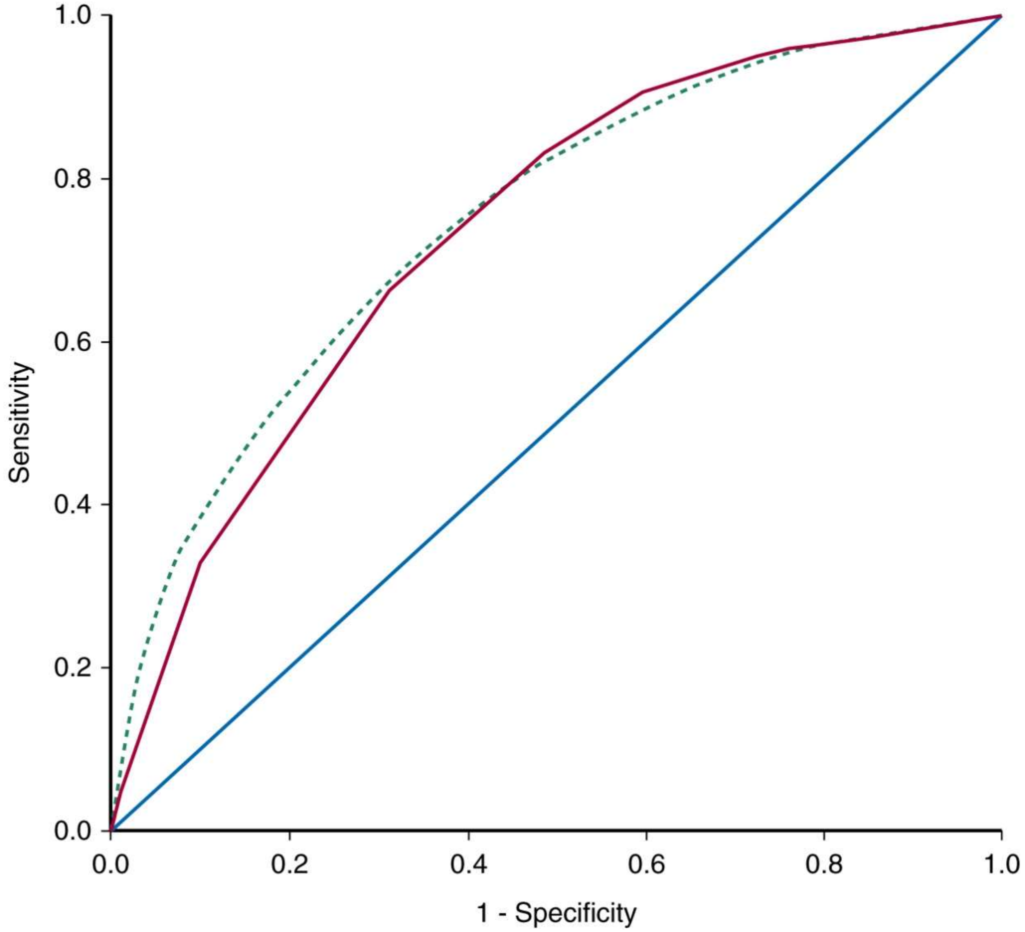

Figure 5: *Receiver operating characteristic (ROC) curve for a model predicting the absence of cattle (dotted line) and the absence of LSUs (shown with a solid line) on rural properties in New Zealand. The x-axis of the ROC curve shows 1-(Pred Animals| Data Animals), traditionally Sensitivity. The y-axis represented (Pred no animals| Data no animals), traditionally 1-Specificity.*

3.5 Discussion

To the best of our knowledge no work like this has been published in the peer-reviewed literature for New Zealand. The small scale of our spatial unit of interest (the farm), our focus on operational uses of the data for controlling disease outbreaks (as opposed to use in disease spread models only) and the large amount of existing farm-level animal data for New Zealand set this study apart from similar studies performed in other countries. Our technique differs from the FLAPs approach (Burdett et al., 2015) as it does not model farm locations but rather models the animal counts for known farm polygons. When compared with the approach taken by the Gridded Livestock of the World (GLW) project (Robinson et al., 2014) the main difference is that our model provides counts at the farm level (rather than a raster cell level count or animal density) , allowing farm level (rather than raster cell level) attributes to play a role in predicting animal counts.

A model that included the log of the area (hectares) of high quality pasture and region in which the farm was located predicted presence or absence of cattle and of livestock in general (represented as LSUs) with reasonable accuracy. Region was the most important predictor variable for counts of LSUs and cattle.

The models performed better when used to predict the aggregated cattle or LSU population of 3km radial zones. Prediction at the level of single farms was poorer (as could be expected). Even though the purpose of the models is to predict presence or absence of animals and the counts of animals present on the farms rather than to examine the effects of different covariates, the relatively uninformative nature of many covariates in the models should be mentioned (that is, the odds ratios for many covariates in the predictive models are close to one, as shown in Table 4). We believe that these covariates still explain some of the variance (and improve prediction) without necessarily explaining an amount of variance beyond the traditional probability threshold value of 0.05. In addition, many of these covariates that do not explain a large amount of variance may have been expected to play a larger role if the modelling had not provided evidence to the contrary. At the farm level we were able to identify that while counts on dairy farms were predicted more accurately (low RPS shown in Figure 3), predicted counts were less accurate for farms classified as contract grazing and dairy heifer rearing. These results could be due to real differences in practices such as constantly changing stock numbers on dairy heifer rearing and contract grazing properties as animals are bought and sold and animals arrive for and leave after grazing contracts. These factors mean that contract grazing properties do not have the same predictable animal counts as properties of a similar size which have more stable population of resident animals. Alternatively, the lower accuracy in prediction could be due to misclassification bias in the dataset used to build the model.

Only an exercise that collects actual farm level data will clarify the reason for low predictability of LSUs and cattle numbers on contract grazing and dairy heifer rearing farms. The high level of regional heterogeneity in the farm-level animal populations in New Zealand makes some regions appropriate training data sets for particular regions but not for others (Figure 4). In particular, the Waikato (an area of high dairy cattle farming density) is well predicted by the models and the Tasman Area is very poorly predicted.

The livestock unit (LSU) model is important to test the hypothesis that in an animal production system where animals live mainly outdoors and eat a predominantly pasture based diet, remotely sensed geospatial layers provide sufficient information to extrapolate counts of LSUs at the farm level. As biosecurity applications include stocking of vaccine banks and predicting the spread of disease based on species composition at the farm level, such a model or the outputs of the LSU model must be adjusted to predict single (we present the cattle model as an example) or specified combinations of species populations on each farm (Burdett et al., 2015; Robinson et al., 2014). A useful characteristic of our model is the ability to predict LSU and cattle counts in the 3km radial zones with a high degree of accuracy reflected by the lower RPS for these zones compared with the predictions for individual farms (Table 3). As accurate data on susceptible animal populations on farms within 3km of a farm with a suspected outbreak of FMD are needed by the responders to a disease event as early as possible in the New Zealand context our model may provide some idea of the magnitude of the risk presented by an IP before farm level information can be collected.

The advantage of using Bayesian inference to estimate the coefficients of the ZIP model is that a predictive distribution of the number of animals on a farm may be estimated explicitly, without resorting either to asymptotic assumptions or complex bootstrapping procedures (West, 2014). This provides a precise measure of uncertainty surrounding predicted animal numbers for each farm. Samples from the predictive distribution may then be used as input for Monte Carlo disease simulation models, such as the Davis Animal Disease Spread model (DADs) and the Australian Animal Disease Spread model (AADIS), so that uncertainty in the population at risk of infection is reflected in the simulated outputs (Bates et al., 2003; Bradhurst et al., 2015). This is an area of future exploration. Having information on farm-level populations is important because in a disease outbreak, not all farms have the same risk of disease (Gates et al., 2014). In particular (in the context of this study), the importance of herd size is well established as a risk factor for disease spread based on detailed data collected from outbreaks in naïve populations (Hugh-Jones, 1972; Keeling et al., 2001; Kitching et al., 2006; Muroga et al., 2013; Tildesley et al., 2008; Tildesley and Ryan, 2012).

Advances in modelling techniques and increased computational ability have resulted in a trend toward using individual animals as the unit of interest in disease spread modelling. The inclusion of within-herd spread in detailed process disease spread models is being employed to represent the effects of farm and herd level heterogeneity (Bates et al., 2003; Bradhurst et al., 2015). These models are being used to evaluate the impacts of management decisions at a microeconomic level on individual farms (Halasa et al., 2013; Porphyre et al., 2013a) and variation in start conditions (farm level populations). The effect of herd size on disease spread has the potential to alter the model outputs that inform policy decisions. The adoption of such modelling techniques requires farm level animal population data and a clear understanding of the inadequacies in the data used and the impacts of any inadequacies on model outcomes. The accuracy of these data are seldom understood or examined (Honhold and Taylor, 2006; Jewell et al., 2015). Recent studies have explored how different models may be used to represent similar outbreak situations and have compared model outputs within and between countries (Dubé et al., 2007; Roche et al., 2015; Sanson et al., 2013, 2011) but little attention has been paid in the published literature to the effect of the population information on the model outcomes and resulting policy advice (Wada et al., 2016). Similarly, although demographic data were identified as lacking during the 2001 FMD outbreak in the UK, the extent of the impact of the inaccuracies on either operations or decisions made based on disease spread modelling is hard to establish (Honhold and Taylor, 2006; Keeling et al., 2001; Kitching et al., 2006; Thrusfield et al., 2005).

Spatial correlation of the RPSs was present at distances less than 18km for the LSU model and distances less than 70km for the cattle model after inclusion of region and other environmental variables. This spatial autocorrelation of RPS could be reduced in future models by identifying missing predictor variables that capture the spatial autocorrelation or by the inclusion of a statistical effect which would extend the ZIP model to a geostatistical model (Diggle and Ribeiro, 1998). Given the large number of farms present in our dataset the addition of this statistical effect would pose a significant computational challenge.

The interpretation of the ROC curves depends on the purpose for which the modelled data are being used. Unless we are able to measure the cost of error and uncertainty around knowing and not knowing if an animal is present on a particular property, it is difficult to choose a preferred cut-off at which to interpret the performance of the Bernoulli component of the model and to appreciate the impact of uncertainty associated with that cut-off. In our case we used the ROC curves as a predictive performance exercise to examine the performance of the Bernoulli component when predicting the absence of animals on a farm. Our interpretation of the results of the ROC curves is that our proposed model does provide valuable information about the demographics of animals present on a farm but it is not able to predict with acceptable accuracy the presence or absence of an animal population. This finding agrees with other studies that used spatial environmental variables to predict the distribution of animal species - distinguishing between areas with low numbers of animals and areas where animals were absent is seldom achievable (Porphyre et al., 2013b).

Aside from the statistical variability that our model is able to estimate around each population estimate (the error), ongoing fluctuations can be expected to occur due to seasonal changes in production and changing long term trends in species composition in different areas. An appreciation of these population trends is important for animal disease outbreak response, as having incorrect data representing a population early in a disease response may delay or influence decision making to the same degree as a lack of information could delay or influence response efforts. Further work is needed to examine how long term and seasonal demographic changes are best represented in the model. An approach to long term changes is to use the presented LSU model and combine this with information present in the New Zealand agricultural census. Strategies that could be explored to address seasonal changes include the incorporation of regional level animal movement data into the model or the use of normalised difference vegetation index (NDVI), which has successfully been used in the past to predict herbivore migration associated with pasture availability in ecological models (Marshal et al., 2006; Mueller et al., 2008; Myneni et al., 1995; Reed et al., 1994; Tveraa et al., 2013). In addition to seasonal and production linked trends, herd size is affected by economic factors such as the milk price. These sources of uncertainty resulting from seasonality and economics suggest that maintenance of up to date national level registries is still an important function for the competent authority for biosecurity.

We are able to identify some sources of bias in our models. Aside from bias present in the data used to build the models, which may be affected by response bias by those populating the data and will be affected by the currency of the data held in the databases, our methods may have introduced bias. We have mentioned that 4.5% of farm records were deleted due to overlapping polygons and that another 27% were deleted due to missing data on production type. If either missing-ness or overlapping-ness are non-randomly distributed amongst the full dataset, bias will result from their removal. When properties were selected to have a 3km zone situated around them, this was done based on the property overlapping the centroid of the grid square. This biases our selection method toward larger farms raising the question of whether the accuracy of fit is affected by the selection of farms in a 3km zone surrounding larger farms when compared with those farms surrounding small farms. Finally when distances to features were measured, the calculation was made from the centroid of the farm to the closest feature rather than from the farm boundary to the closest feature. It is possible that this might bias the estimate of the distance from a farm to a feature for larger farms when compared with smaller farms predisposing our models to type 1 error.

Although our goal was to predict the animal numbers on farms for which data were not available, we have only been able to test the model using withheld samples of farms for which the data were available, that is, we have performed internal validation. Internal validation is inherently unsatisfying; it demonstrates that the sample of data that is available shows reasonable consistency from the point of view of fitting the model, but does not provide solid evidence of the extensibility of the model to novel situations. Therefore for the deployment of the model it is necessary for us to assume that the farm population for which data are missing is represented statistically by the farms for which we do have information. There are numerous reasons that would undermine this assumption, for example, that the missing farms have a greater proportion of peri-urban properties, that non-inclusion is not homogenously distributed across all production types or that overlapping properties are not randomly distributed in space. However, we proceed under the assumption that some information is better than no information. Further work is underway to develop an independent dataset that can be used for a more complete external validation of our models, but that is beyond the scope of this article.

3.6 Conclusions

We have constructed a model that predicts the presence and absence as well as the numbers of livestock units and cattle at the farm level in New Zealand. While wildlife ecology uses similar models for conservation and research purposes, the use of modelled or inaccurate point estimates of animal populations for biosecurity operations (rather than when preparing for a disease outbreak) may introduce systematic error which could increase the number of properties affected by, and the eventual cost of a fast moving animal disease. Predictions were accurate at a wide spatial scale (3km zones or administrative regions) and our model allows us to explore uncertainty around a point estimate in animal numbers at the farm level by incorporating farm population posterior distribution estimates in disease simulators. Future work will further refine the modelling approach to improve its internal validity and provide external validation of the model using data collected from the field to fully assess the precision of the estimates of livestock numbers and of cattle on farms and to test the extensibility of the derived covariates to predict farm level counts on an unrelated spatial farms database (FarmsOnLine). While probabilities of presence and distributions of animal counts are useful in exotic disease preparedness and disease spread simulation modelling, when faced with the already considerable uncertainty inherent early in a disease outbreak response the decision maker requires the most accurate and correct information to make high quality decisions. Every reasonable effort must be made to strengthen data linkages between existing animal population data sources by aligning data collection efforts undertaken by government and industry and by the use of a single, national farm identifier.

3.7 Appendix to Materials and Methods

Addressing collinearity in predictors:

Collinearity between the *solar exposure* and *temperature* variables was addressed by creating new variables using the difference (T) between the mean *annual temperature* (MAT) and mean *minimum temperature* (MMT) for each farm (*i*) so that:

$$(1)\ \ \mathrm{T}_i = \mathrm{MAT}_i - \mathrm{MMT}_i$$

and the difference (S) between mean *annual solar exposure* (MASE) and *minimum solar exposure* (MMSE) for each farm (*i*) so that:

$$(2)\ \ \mathrm{S}_i = \mathrm{MASE}_i - \mathrm{MMSE}_i$$

Mean *annual solar exposure* and mean *annual temperature* were retained in the models along with the bivariate transformations T_i and S_i shown in equations (1) and (2).

Aspect defines the orientation of the land surface and is measured as a compass bearing with 0 degrees representing a surface facing true north. Given this system of measurement, a simple average or standard deviation is not a meaningful summary statistic and it is necessary to further transform *slope* and *aspect.* The method proposed by Stage (1976) was used to integrate the effects of slope and aspect so that for each farm the slope and aspect were summarised across the farm polygon by including the product of slope and the cosine function of aspect and the product of slope and the sine function of aspect (Stage, 1976). Mean and standard deviation of elevation for the farm were calculated by averaging each raster cell of a surface generated by a digital elevation model contained within the farm polygon to take into account the effect of elevation (Stage and Salas, 2007).

Model framework:

In our approach to $\boldsymbol{y}$ denotes the vector of outcomes (cattle or LSU counts) for each of $i = 1, \ldots, n$ farms and $\boldsymbol{X}$ the *n* x 18 design matrix of predictor variables shown in Table 2. Under the ZIP model, it is assumed that

$$(3)\quad y_i = \begin{cases} 0 \text{ with probability } p_i \\ z_i \text{ with probability } 1 - p_i \end{cases}$$

Where

$$(4)\quad z_i \sim \text{Poisson}(\lambda_i)$$

Logit and log-linear models are then specified for p_i and λ_i respectively, such that

$$(5)\quad \text{logit}\left(\frac{p_i}{1-p_i}\right) = \alpha_{\gamma} \boldsymbol{x}_i^{\boldsymbol{T}} \boldsymbol{\gamma}$$

and

$$(6)\quad \log \lambda_i = \alpha_{\beta} \boldsymbol{x}_i^{T} \boldsymbol{\beta}$$

where $\boldsymbol{x}_i$ represents the *i*th row of the design matrix of predictor variables $\boldsymbol{X}$ (Table 2), and $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are column vectors of coefficients. In a ZIP model, a portion of the zero predictions arise from the Bernoulli component with the remainder from the Poisson component (Lambert, 1992). Independent Normal (0, 10) prior distributions were chosen for each component of the coefficient vectors $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$. The model was fitted to the data $(\boldsymbol{y}, \boldsymbol{X})$ using a Markov Chain Monte Carlo (MCMC) algorithm, from which the joint posterior distribution $\pi(\boldsymbol{\gamma}, \boldsymbol{\beta} | \boldsymbol{y}, \boldsymbol{X})$ was estimated numerically. This normal prior was selected as the coefficient values were expected to fall between -20 and 20 – the interval encompassed by 95% of this distribution.

Since probabilistic predictions of $\boldsymbol{y}$ are required, we wish to work in a Bayesian framework. Moreover,

we may wish to perform both model selection (for interpretation of covariates) and model averaging for

the final predictions. To do this, we expand the notation of the design matrix as X_β and X_γ representing

(possibly different) design matrices for the Poisson and Bernoulli components of the model respectively, with corresponding coefficients β_M and γ_M respectively. The number of columns in both matrices together define a particular model which we denote by M with dimension $d(M)$, and with M_{max} denoting the “full” model containing all covariates. Thus, we wish to calculate a posterior distribution $\pi(M, \beta_M, \gamma_M \mid y, X)$ over the set of models $\mathcal{N} = \{M : d(M) \leq d(M_{max})\}$ and coefficients β and γ.

We denote the joint posterior

$$(7)\quad \pi(M, \beta, \gamma \mid y, X) = \frac{f(y \mid \beta, \gamma, M) f(\beta \mid M) f(\gamma \mid M) f(M)}{\int_{\mathcal{N}} \int_{\Gamma} \int_{\theta} f(y \mid \beta, \gamma, M) f(\beta \mid M) f(\gamma \mid M) f(M) \mathrm{d}\theta \mathrm{d}\gamma \mathrm{d}M}$$

Conditional on a particular model M, we draw samples from the conditional distributions $\pi(\beta_M \mid \cdot)$ and $\pi(\gamma_M \mid \cdot)$ using standard adaptive Metropolis Hastings updates within a Markov chain Monte Carlo (MCMC) algorithm (for example Chib and Greenberg, 1995). For our problem, the number of possible models is potentially large (e.g. $d(M) = 10$ gives 2^{10} possible models), so we employ a reversible jump algorithm as in Green [1995], rather than relying on Gibbs sampling techniques such as Carlin and Chib (1995). Our setup incorporates reversible jump proposals that either add a covariate $(d(M) \rightarrow d(M) + 1)$ or delete a covariate $(d(M) \rightarrow d(M)-1)$ within the MCMC scheme as a whole. Thus jumps are proposed that increase or decrease the dimension of the parameter space by 1 both in the Poisson and Bernoulli parts of the model independently. The covariate to add or delete is chosen uniformly at random, and a deterministic mapping function is employed to propose new values of all parameters within the linear predictor associated with each

part of the model (Equations 5 and 6). For each part of the model, we then employ the reversible jump algorithm.

Reversible jump MCMC:

Consider a generalised linear model with noise distribution $G(\mu)$ and link function $\eta = g(\mu)$, with design

matrix X and parameters β. We wish to mix over the set of models $N = \{M : d(M) \leq d(M_{max})\}$ where

$d(M)$ denotes the dimension of model M (i.e. the number of covariates), and M_{max} denotes the full, or

saturated, model. Since the number of models $2^{d(M_{max})}$ may be large, we choose a reversible jump algorithm over other model selection techniques such as the Carlin and Chibb Gibbs sampling approach. Our reversible jump algorithm proposes to either increase or decrease $d(M)$ by 1 at each update, using a deterministic dimension-matching proposal for the parameter values based on pivoting the model around the value of the linear predictor:

Adding a covariate

1. Calculate $\eta = X_M \beta_M$
2. Propose to move $d(M') = d(M) + 1$ by choosing a covariate to add uniformly with probability $\frac{1}{D-d(M)}$
3. Propose $\beta_{M'} = (X_{M'}^T X_{M'})^{-1} X_{M'}^T \eta$;
4. Accept $M = M'$, and $\beta_M = \beta_{M'}$ with probability $1 \wedge \frac{\pi(M', \beta_M' | y, X_M')}{\pi(M, \beta_M | y, X_M)} \cdot \frac{D-d(M)}{d(M')}$.

Deleting a covariate

1. Calculate $\eta = X_M \beta_M$
2. Propose to move $d(M') = d(M) - 1$ by choosing a covariate to add uniformly with probability $\frac{1}{d(M)}$
3. Propose $\beta_{M'} = (X_{M'}^T X_{M'})^{-1} X_{M'}^T \eta$;
4. Accept $M = M'$, and $\beta_M = \beta_{M'}$ with probability $1 \wedge \frac{\pi(M', \beta_M' | \text{ƴ}, y, X_M')}{\pi(M, \beta_M | \cdot)} \cdot \frac{d(M)}{D-d(M')}$.

The Markov chain Monte Carlo algorithm proceeds by first updating each component of β_M and γ_M using

single-site Metropolis Hastings, followed by a reversible jump update according to the following scheme:

1. Initialise the chain at $\beta_M^{(0)}, \gamma_M^{(0)}, M^{(0)}$;
2. Update β_M and γ_M component-wise using Metropolis Hastings
3. With probability 0.5 do :
    a. Add a component to X_β, $d(M) \rightarrow d(M)+1$;
    Or
    b. Delete a component from X_β, $d(M) \rightarrow d(M)+1$;
4. With probability 0.5 do :
    a. Add a component to X_γ, $d(M) \rightarrow d(M)+1$;
    Or
    b. Delete a component from X_γ, $d(M) \rightarrow d(M)+1$;
5. Go to 2

As usual with MCMC, steps 2-4 above are iterated as many times as required for inference, bearing in mind thinning of the Markov chain to reduce autocorrelation between successive samples. Convergence is checked by initialising multiple chains at different points, and monitoring convergence to a common limiting distribution.

Measuring predictive performance:

For prediction purposes, the Bayesian predictive distribution for the outcome variable $\boldsymbol{y}^\star$ of the target data $\boldsymbol{X}^\star$ (i.e. the values of the predictor variables calculated for each farm polygon in the target dataset) is estimated numerically by simulating from the ZIP model specified above, using draws of the model coefficients $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ as represented in the MCMC output.

The rank probability score (RPS), which compares a predictive distribution against an observed value, was used to assess the performance of the Bayesian prediction using a variety of scales of validation data sets (Brier, 1950). The RPS is defined as

$$(8)\quad RPS(\Pi, y) = \frac{1}{K-1}\sum_{k=0}^{K}(\Pi(k) - \mathbf{1}[k \geq y\ \ ])^2$$

where Π is the cumulative mass function for a discrete probability distribution representing a prediction, y is the observed value, and K is the largest value of the outcome variable, predicted or observed, for any farm in the dataset. Small values of the RPS indicate a good prediction, taking the value 0 if the predictive distribution takes point mass on the observed value (i.e. a "perfect" prediction with no bias or uncertainty). In the case where the posterior prediction is a point estimate, the RPS reduces to the mean absolute error.

Residual checking:

The RPS (as an analogue of a frequentist model residual) was analysed in five ways to asses the fit of the model and to identify further areas of study.

i) Spatial autocorrelation of the RPS for each farm in New Zealand was examined using variograms (Diggle and Ribeiro, 1998). Given the size of the dataset a parallel implementation of the variogram function was written as an R package, making use of Graphics Processing Unit-accelerated computation (Jewell, 2016b).The variogram was plotted, and visually examined for evidence of spatial autocorrelation. This was done for an area in the North Island that comprised the Waikato Region, the Bay of Plenty Region and the Auckland Region and for an area in the South Island comprising the Southland Region, the Otago Region and the Canterbury Region (Figure 1). The analysis was limited to 350km, being approximately half the maximum distance between farms in either selected area, with a step size of 10km for both the LSU and the cattle models. A null distribution for the variogram, assuming no spatial correlation, was calculated by permuting the RPS scores among spatial locations (Diggle and Riberio, 1998). The 5th and 95th percentiles of the variogram null distribution are shown by the dotted lines in Figure 2.

ii) To assess how farm type affected model fit, the RPS for farms classified as dairy farms, pastoral livestock farms, pig breeding and pig finishing establishments and properties used for grazing young stock and dairy cattle not currently in production was compared. Results of this comparison for both the cattle and the LSU models were visually examined as boxplots and were statistically tested with the Kruskal-Wallis test for homogeneity amongst independent non-parametric groups (R-Core Team, 2013).

The models for the area level validation were built from subsets of training farms with target or validation farms selected spatially either by 3km radial zones or by New Zealand administrative region. While the primary unit of interest for these analyses is RPS at the level of the specified area, mean farm level RPS values are also presented.

iii) To assess model performance at the scale of 3km radial zones, 10 sets of 200 3km zones were selected by randomly selecting 200 cells from a grid of 9000m x 9000m cells which was overlaid across a map of New Zealand. Each zone was generated by creating a 3 km buffer around the farm polygon that intersected the centroid of each randomly selected grid cell. We chose to use 200 3km radial zones following experimentation which found a good trade-off between having enough 3km radial zones for each model instance to provide sufficient information on the quality of fit, but not having so many that overlap between farms within adjacent zones became a substantial issue. All farms that intersected with this 3km buffer zone were extracted using spatial libraries (Bivand et al., 2014a, 2014b; Pebesma et al., 2014) in R (R-CoreTeam, 2013). The training data comprised all properties not selected by this method. Models were validated by comparing the observed and predicted counts of LSUs or cattle in zones in the validation data set. The posterior prediction for each farm within each zone was aggregated and regarded as the posterior prediction for the zone. The RPS was calculated by comparing the predicted zone-level counts against the observed counts aggregated for the farms in each zone. The mean RPS at the zone level was then the sum of all zone-level RPSs divided by the number of zones (200). This process of model building and validation for the 3km areas was repeated for 10 independent sets of randomly selected training and target zones. The results were summarised across the 10 model fits by calculating the mean and standard deviation for RPSs at the level of the 3km zones (Table 3).

iv) To assess model performance at the scale of regions, training and target datasets were generated by stratifying first by island (North Island and South Island), and then randomly selecting 6 of 9 regions from the North Island and 5 of 7 regions from the South Island to act as training data. The remaining 3 North Island regions and the 2 South Island regions were target data sets for validating each of the LSU and cattle model fits. Region was removed as a fixed effect from the model. Fifty independent training and target datasets were constructed and both the cattle and LSU models were set to run 50 000 iterations on each dataset. This analysis aimed to identify if predicted LSU or cattle counts were more reliable in particular regions. The posterior prediction for each farm within each region was aggregated and regarded as the posterior prediction for each region. The RPS was calculated by comparing the predicted counts against the observed counts aggregated for the farms in each region (Figure 4). The mean RPS at the region level was the sum of all region-level RPSs divided by the number of regions.

v) As a final predictive performance step, the performance of the Bernoulli part of the model (which predicts presence or absence of LSUs and cattle on farms) was examined using ROC methodology (Zweig and Campbell, 1993). The conventional use of a ROC curve establishes the test characteristics (sensitivity and specificity) of a diagnostic test. Sensitivity is defined as (T+|D+) – the probability of a diseased animal (D+) returning a positive test (T+) and specificity is (T-|D-) – the probability of a non-diseased animal (D-) returning a negative test result (T-). The ability of the model to predict the absence of animals was regarded as T+ (or Pred Zero +), the absence of animals in the data as D+ (or Data Zero +), the prediction of animals present as T- (Pred Zero -) and the presence of animals on the property in the data as D- (Data Zero -). This can be summarised as follows:

Sensitivity = (Pred Zero +| Data Zero+)

Specificity = (Pred Zero -| Data Zero-) = (Pred Animals| Data Animals)

The x-axis of the ROC curve showed 1-(Pred Animals| Data Animals). The y-axis represented (Pred no animals| Data no animals). For each farm, the posterior predictive probability of absence of animals was calculated. A threshold was then imposed on this probability as the Bernoulli prediction of whether a farm had animals present or not. A ROC curve was constructed (Figure 5) to compare this prediction against the real data under different values of the threshold. This process was completed for both the LSU model and the cattle model using the whole dataset as a training dataset in each instance.

3.8 Acknowledgements

We thank the Ministry for Primary Industries in New Zealand for support of this project and the Centre of Excellence for Biosecurity Risk Analysis (CEBRA) at University Melbourne for funding. Thanks are due to AsureQuality and Robert Sanson for the use of AgriBase data for this project; to Hannes Calitz of the Institute of Fundamental Sciences at Massey University for assistance with grid computing; to Graeme Garner of the Australian Federal Department of Agriculture and Water Resources, Daan Vink of the Ministry for Primary Industries and Carolyn Gates of the EpiCentre at Massey University who provided input on early on drafts of this work.

CHAPTER 4

4 Does size matter to models? Exploring the effect of herd size on outputs of a herd-level disease spread simulator

4.1 Abstract

Disease spread modelling is widely used by veterinary authorities to predict the impact of emergency animal disease outbreaks in livestock and to evaluate the cost-effectiveness of different management interventions. Such models require knowledge of basic disease epidemiology as well as information about the population of animals at risk. Essential demographic information includes the production system, animal numbers and their spatial locations yet many countries with significant livestock industries do not have publically available and accurate animal population information at the farm level that can be used in these models. The impact of inaccuracies in data on model outputs and the decisions based on these outputs is seldom discussed. In this analysis, we used the Australian Animal Disease model (AADIS) to simulate the spread of foot-and-mouth disease seeded into high-risk herds in six different farming regions in New Zealand. We used three different susceptible animal population datasets: (1) a gold standard dataset comprising known herd sizes, (2) a dataset where herd size was simulated from a beta-pert distribution for each herd production type, and (3) a dataset where herd size was simplified to the median herd size for each herd production type. We analysed the model outputs to compare (i) the extent of disease spread, (ii) the length of the outbreaks and (iii) the possible impacts on decisions made for simulated outbreaks in different regions. Model outputs using the different datasets showed statistically significant differences, which could have serious implications for decision making by a competent authority. Outbreak duration, number of infected properties, and vaccine doses used during the outbreak, were all significantly smaller for the gold standard dataset when compared with the median herd size dataset. Initial outbreak location and disease control strategy also significantly influenced the duration of the outbreak and number of infected premises. The study findings demonstrate the

importance of having accurate national-level population datasets to ensure effective decisions are made before and during disease outbreaks, reducing the damage and cost.

4.2 Introduction

In countries that are historically free of significant livestock diseases such as foot-and-mouth disease (FMD), the outputs of disease spread models are a useful proxy for field information on disease behavior. This information may be used by the competent authority to compare the impacts of alternative disease control policy decisions (Thomas W Bates et al., 2003; Halasa et al., 2013; Harvey et al., 2007; Risk solutions, 2005; Roche et al., 2015, 2014; Sanson et al., 2017; Stevenson et al., 2007). Traditionally, the policies of FMD-free countries such as the United Kingdom, the United States, Australia and New Zealand rely on stamping out methods to eradicate outbreaks of FMD. This involves depopulation and thorough cleaning and disinfection of detected infected premises (IPs), tracing and bio-containment of all contacts, active surveillance to detect all clusters of infection, and intensive movement restrictions to limit disease spread. Areas of ongoing research include comparing the impacts of policies that would allow animals to be vaccinated, with policies that would cull animals on all affected farms. Furthermore, if a vaccination policy is considered, the impact of vaccinating cattle only compared with vaccinating all susceptible species (Roche et al., 2015) is of interest as there are seldom sufficient human resources and vaccine doses to target every animal.

These comparisons must consider the spread of disease in different regions, the effectiveness of a variety of control options, and the economic impact of the outbreak. This includes the cost of control measures and loss of trade due to restrictions implemented by the international community. The complexity of these decisions has driven the development of ever more complex disease spread simulation models that can now incorporate detailed information on within- and between- herd spread of disease. Some models, for example, use the count of animals on each farm to estimate infectivity according to latent periods, within-herd contact rates, and incubation periods, specific to the species and numbers of each species present in each herd on each farm. Although there is still much debate over the best modelling approach (Dubé et al., 2007; Probert et al., 2015; Roche et al., 2015; Sanson et al., 2011; Webb et al., 2017), a key requirement of any spatially enabled disease spread simulator is national (or district/state/county) level data of farm locations with susceptible animal populations. The model also requires data on the contact patterns of susceptible individuals and disease-specific information for each species represented. Few countries have publicly available and accurate animal population information at the farm level, which can be used in these models; however, the impact of inaccuracies in data on model outputs and decisions is seldom discussed.

This is an increasingly critical point as these modelling activities generally make use of centrally held datasets, the accuracy of which is rarely scrutinized (Honhold and Taylor, 2006; Jewell et al., 2015), while modelling becomes both progressively more complex and more highly valued by decision makers.

The objective of this study was to test the null hypothesis that uncertainty around farm level animal population sizes is not important when interpreting the outputs of within-herd spread FMD models. Whilst no dataset can be expected to have an exact representation of herd size at every point in time, our study is concerned with examining the performance of an FMD model which explicitly models within herd spread using a heterogeneous herd dataset with census-based herd sizes, compared with simplified herd datasets where herd size is estimated according to herd type. Three different herd datasets are used in simulations that cover six geographic areas in New Zealand, under three different disease control strategies. Each of the geographic areas have large populations of foot-and-mouth susceptible livestock in different densities. Impacts on outbreak size and duration are assessed, and the potential implications for decision makers and competent authorities of using inaccurate data discussed

4.3 Materials and Methods

To design an experiment that provides information on the impact of herd-level animal counts on disease modelling, a disease spread simulator that utilizes the susceptible population within a herd or farm was required. The Australian Animal Disease model (AADIS) (Bradhurst et al., 2015), which was developed in 2015 for use by the Australian Federal Government for disease response preparedness, was used for this study. AADIS is a hybrid model of livestock disease spread and control which is designed to support emergency animal disease planning. The disease simulator uses both population-based and individual-based modelling techniques. AADIS uses the herd populations to model within-herd spread with a deterministic model, and between-herd spread with a spatially-explicit stochastic agent based model (ABM). Our models use passive first IP detection which comprises two stochastic processes: detection and reporting. Detection is defined as inspecting stock (on a farm, at a sale yard or in an abattoir), noticing clinical signs and consulting a veterinarian. An infected herd is only a candidate for detection if it meets the minimum clinical prevalence level configured for the herd type. Reporting is defined as a veterinarian suspecting FMD, sending samples to a lab, FMD being confirmed and the Chief Veterinary Officer (CVO) being notified. The detection and reporting probabilities are defined per herd type and per premises type.

The AADIS model allows for representation of a situation where any type of herd may be present on any type of farm. AADIS allows disease to spread more effectively between herds on the same farm than between herds on separate farms. The ability to parameterise the spread of disease with respect to herd type and region captures the heterogeneous nature of seasonal management practices and contact patterns. For example, a beef herd on a non-commercial farm (referred to as a lifestyle block in New Zealand) will be subject to different management and marketing practices affecting the spread of disease when compared with a beef herd on a pastoral farm. AADIS also takes species and herd size into account when estimating herd susceptibility and infectivity reflecting Tildesley and colleagues' (2008) observation that a non-linear relationship between herd size and herd infectivity/susceptibility best described data from the 2001 UK FMD outbreak (Tildesley et al., 2008).

Model parameters

Herd location and herd size data were obtained from AgriBase, a commercial database of farm properties and animal populations maintained by AsureQuality, a New Zealand state owned entity (Sanson and Pearson, 1997). All AgriBase farms with no animals susceptible to FMD were removed, leaving 115,618 herds on 76,487 farms in the model (Table 5). Farms were categorised into four primary types (pastoral farming, dairy cattle farming, lifestyle farming or pig farming) based on what farmers reported in AgriBase as their main production activity. Ten secondary herd types were created based on other livestock species present on the farm: a large and small herd type for each of deer, sheep, pigs, dairy cattle and beef cattle, which are the farmed species in NZ susceptible to FMD (Table 5). Large or small herd type was allocated based on the farm type in AgriBase (which is specified by the farmer) to make allowance for management practices and then further divided based on the size of the herd. This made it possible for some of the effects of hobby or 'lifestyle' farming to be represented by assigning a 'small' herd type to those herds of any kind (or size) present on hobby farms and then allocating cut points as shown in Table 6. Cut points were chosen based on experience with farming practices in New Zealand.

To explore the importance of herd size, three different herd size model parameterisations were derived from the AgriBase data. The first used the actual herd sizes reported in AgriBase and represented the real or 'gold standard' dataset; the second assigned each herd a size equal to the median herd size for each of the 10 herd types. The third assigned each herd a size that was sampled from a beta- pert distribution generated from each of the 10 herd types. The herd type descriptive summaries used to generate the median and beta-pert distributions are shown in Table 6. Beta-pert distributions were selected to represent the herd size distributions based on testing of the gold standard herd dataset in a quantitative risk analysis software which identified this as being the best fit for the data (VoseSoftware, 2017). The minimum, maximum and mode (most likely) values were then selected to describe the best fitting beta-pert distribution for each of the ten herd types. Each simulation run sampled a different value for each herd from the constructed beta-pert distributions.

Table 5:*Counts of secondary herd types and primary farm types used to parameterise the Australian Animal Disease model (AADIS). Movement patterns and management activities vary by herd type and susceptibility to FMD varies by species.*

Secondary herd types on farm	**Primary farm type**				
	Pastoral	**Dairy**	**Pigs**	**Lifestyle**	**Total**
1 large sheep	17950				17950
2 small sheep	6772	1526	43	12086	20427
3 large pigs	114	33	126		273
4 small pigs	2044	624		1770	4438
5 large deer	3236	63			3299
6 small deer	302	22		294	618
7 large dairy	2010	11806	2		13818
8 small dairy	465	24		681	1170
9 large beef	23559	1203	40		24802
10 small beef	7907	2080	22	18814	28823
Total	64359	17381	233	33645	115618

Table 6: *Descriptive data for each of the herd types used to parameterise AADIS. The minimum, mode and maximum were used to create the beta-pert dataset, the median for the median dataset and remaining descriptors serve to describe the distribution of the gold standard data*

Herd type	Minimum	25th percentile	Mode	Median	Mean	75th percentile	Maximum
large sheep	50	214.2	100	1998.9	1080	2720	115600
small sheep	1	6	10	33.4	12	25	14450
large pigs	12	60	40	1168.3	300	1471	44000
small pigs	1	2	5	4.5	5	5	30
large deer	15	76	100	404.7	187	430	19249
small deer	1	2	1	14.5	6	12	1365
large dairy	15	193	200	400.9	304	497.8	10220
small dairy	1	2	1	10.5	4	10	650
large beef	15	31	20	168.1	73	195	14500
small beef	1	3	2	7.9	6	10	657

The disease specific parameterisation of the AADIS model was derived from the New Zealand Standard Model of FMD spread, which is represented in Interspread Plus, and models an outbreak of FMD serotype O pan PanAsia (Grubman and Baxt, 2004; Owen et al., 2011; Stevenson et al., 2013).

Outbreak Seeding

Using a random seed design across the whole of New Zealand in the AADIS model results in a large degree of heterogeneity in outbreak size with an insufficient number of large outbreaks to allow comparison of the different herd size scenarios equally. Furthermore, the population densities of farms and susceptible animals are known to influence the spread of highly infectious diseases (Thrusfield, 2007) as well as the efficacy of vaccination strategies for FMD (Flood et al., 2013; Thibaud Porphyre et al., 2013b). To address these known effects, as well as to make the simulations more representative of an economically severe outbreak in New Zealand, the territorial local authorities (TLAs) that were most likely to have FMD introduced were identified and then further ranked based on where an introduced outbreak would be most likely to spread.

The greatest risks of FMD introduction to New Zealand have been reviewed and published elsewhere (H J Pharo, 2002). Based on the FAO FMD contingency plans manual (Geering and Lubroth, 2002), the greatest risk for New Zealand appears to be through the feeding of FMD infected material to non-commercially kept pigs. As there was no data available for imported materials, we used the density of small pig herds as a proxy for the risk of introduction. The likelihood of spread was based on cattle and pig population density (Table 7).

Six TLAs were chosen to provide sufficient areas to give examples in both the North and South Islands but to limit the number of TLAs so that results are still intuitively comparable (Figure 6). The goal of the study is not to predict the distribution of outbreaks sizes, but to allow the comparison of the different data quality scenarios.

Figure 6: *Six New Zealand territorial local authorities (TLAs) selected for locations of simulation models to examine the effects of variation in herd size in the within-herd spread disease model of foot-and-mouth disease (FMD), AADIS.*

Table 7: *Description of the six New Zealand territorial local authorities (TLAs) used as disease index herd for hypothetical foot-and-mouth (FMD) outbreaks in the study.*

Territorial local authority (TLA)	**Area in (km²)**	**Farms with FMD susceptible animals/km²**	**Mean Nearest Neighbour distance (km)**	**Count of small pig herds**	**small pig herds/km²**	**cattle/km²**	**pigs/km²**
New Plymouth	2205	0.995	0.347	123	0.056	87	3.692
Auckland	4947	1.649	0.284	240	0.049	60	2.176
Whakatane	4474	0.27	0.454	161	0.036	38	1.504
Rangitikei	4484	0.264	0.700	110	0.025	41	0.727
Tasman	9650	0.201	0.534	148	0.015	13	0.049
Southland	30198	0.11	0.862	150	0.005	22	0.038

Control Strategies

The control of disease within AADIS is a part of the ABM. Measures include movement restrictions, surveillance and tracing, IP operations, resource management, and vaccination. The emergent behaviour of the ABM is the spatiotemporal spread of disease across the population and the subsequent activities to control and eradicate the disease. The disease spread pathways and control measures can be thought of as components of the ABM environment. Each component of the AADIS ABM environment operates independently (Bradhurst et al., 2015). Three control strategies were modelled. The first was 'stamping out' of identified infected farms (each FMD-susceptible herd on each of these farms is culled). The second applied vaccination to all susceptible species (with no restrictions on the number of doses). The third applied vaccination to cattle only.

Research is ongoing to identify alternative methods to controlling and eradicating the FMD virus, rather than automatic culling of sometimes healthy animals. The benefits of augmenting stamping out with vaccination for disease-free countries have been explored, and the strategy of vaccinating cattle only postulated as an effective alternative to vaccinating all susceptible species (J A Backer et al., 2012; J. A. Backer et al., 2012; Kahn et al., 2002; Roche et al., 2015; Sanson et al., 2017, 2013). Here we examine the effect of the accuracy of herd-level population information on the selection between two vaccination strategies, namely vaccinate cattle only, and vaccinate all susceptible animals (note that culling of animals on infected premises is still employed in these strategies). When considering the use of vaccination (vs stamping out only), decision makers must take into account the current World Organisation for Animal Health (OIE) regulations, which restrict international trade for a country for an additional time period if it employs vaccination compared with if it employs a stamp out strategy (Anonymous, 2017).

Model simulations

One thousand simulations were performed for each of the nine model parametrisations (three control strategies and three herd populations) in each of the six selected TLAs, giving 54,000 simulations in total. The study structure is represented in Table 8. Each outbreak simulation was seeded into a small pig herd (small pig herds were identified as the most likely entry point in NZ). The same seed lists were reused in each of the 54 different model parameterisations. Simulated outbreaks, which reached 365 days' duration (number of simulation days) were terminated.

Table 8: *Tabular representation of a study designed to test the null hypothesis that uncertainty around herd size is not important when interpreting the results of a within-herd spread FMD model. Three control strategies and three herd population datasets were used to create nine model scenarios each of which were run for 1000 iterations in each of six New Zealand territorial local authorities – each cell (1A – 3C) represents 6000 model iterations each using the same order of seed farms.*

		Control Strategy		
		Stamping out	**Vaccinate all species**	**Vaccinate cattle only**
Herd data set	**Beta-Pert**	**1A** 1000 iterations in 6 TLAs	**1B** 1000 iterations in 6 TLAs	**1C** 1000 iterations in 6 TLAs
	Real	**2A** 1000 iterations in 6 TLAs	**2B** 1000 iterations in 6 TLAs	**2C** 1000 iterations in 6 TLAs
	Median	**3A** 1000 iterations in 6 TLAs	**3B** 1000 iterations in 6 TLAs	**3C** 1000 iterations in 6 TLAs

Statistical analysis

Those outbreaks that failed to propagate were analyzed with a logistic regression model that included data type, TLA and control strategy as explanatory variables, and failure to propagate as the outcome variable. This analysis was performed to test the hypothesis that failure to propagate was independent of control type but was associated with herd data type and TLA.

For outbreaks that were eliminated within 365 days, outbreak duration and count of infected premises (IPs) on the last day of the outbreak were used as outcome variables as these are both important to decision makers choosing between control strategies. Further analysis was performed between the two vaccination scenarios using count of vaccinated animals (a proxy for vaccine doses) as the outcome variable with the same explanatory variables. We included interactions between each of these terms and a three-way interaction between all explanatory variables in all models.

AADIS outputs were described and analysed using the R statistical computing language, Cox Proportional Hazard (CPH) models were fitted using the R survival package (Hosmer et al., 2008; Team, 2017; Therneau, 2015; Therneau and Grambsch, 2000). The data were right censored because not all outbreaks had been eradicated within 365 days when the simulations were terminated.

Although the study design is balanced, the data are not, because many simulations generated outbreaks that were not detected, and were removed from further analysis (Table 9). Therefore, CPH models were run with all orderings of predictor variables to ensure that the explanatory ability of the variables was assessed conditionally on other variables in the model.

In order to simplify how we determined the relative contribution of each predictor variable to the response variable, we augmented comparison of p-values (which are all highly statistically significant), with a comparison of the deviance values that they index. Our reasoning is that under the null hypothesis of no term effect, the deviance follows a Chi-square distribution with set degrees of freedom. Ordinarily we compare these values with the index distribution to obtain a p-value. Because of the large numbers of simulations and the strength of the effects, all p-values are very small (Table 10). Our goal was to make a statement about the relative contribution of each of the experimental variables upon the response variable, but the uniformly very small p-values are difficult to interpret in this light. Therefore we augmented our consideration of the predictor variables by interpreting the size of the deviance values relative to their expectation under the null hypothesis of no effect, which is the same as the number of degrees of freedom. So, the relative importance of model co-variates was determined by averaging the deviance values from the output then dividing by the degrees of freedom for each term to estimate the relative deviance..

To assess the effect of herd type on choice of control strategy, the strategy that had the lowest operational cost for each seed herd was recorded. This process was repeated for each of the herd data types. This allowed the percentage agreement and Fleiss and Cohen's Kappa statistics on the lowest cost operational option to be calculated between the real, median and beta pert data sets (Cohen, 1960; Fleiss, 1971). When one of the herd type data sets did not have a completed simulation for that seed, the seed was deleted from the dataset used for comparison. This left 4292 of a possible 6000 data lines to compare between scenarios.

The operational cost of each strategy selected using the real herd data was compared with the operational cost that would have been incurred had an alternate strategy (based on the suggestion of the alternate herd data set) been followed. The cost of the alternate strategy was based on the cost generated by the real herd dataset. These amounts were examined as ratios rather than absolute amounts as we wish to demonstrate the value of this information rather than to predict outbreak costs which will change over time. As an example of how this comparison was made, consider a model iteration where the lowest fixed cost management strategy was to stamp out according to the real herd size dataset. This model iteration would be found in cell 2A in Table 8. Conversely, when using the median herd size data-set the lowest operational cost corresponds with the strategy to vaccinate all species in cell 3B in Table 8. To calculate the ratio between costs using the real dataset and median dataset, the real cost of the strategy to vaccinate all species (cell 2B) was divided by the cost of stamping out (cell 2A).

4.4 Results

Descriptive statistics for the number of simulated outbreaks that reached 365 days, the number of outbreaks that ended before they were detected, and the number of outbreaks that were detected and controlled are shown in Table 9 for each of the herd size scenarios. When the subset of simulations that ended prior to spreading were analysed in a logistic regression model, the explanatory variables representing data type and TLA were significant, but control strategy was not significant ($p > 0.05$).

The distribution of outbreak duration and counts of IPs between TLAs, control options (only stamping out, culling IPs and vaccinating all susceptible animals, and culling IPs and vaccinating cattle only), and data sets (beta-pert modelled herd size, gold standard herd size and median herd size) were compared and the results displayed in Figure 7 (duration) and Figure 8 (count of IPs). The CPH models demonstrated that for outbreak duration and number of IPs, all three explanatory variables are significantly associated with the outcomes. In addition, for each of the models a three-way interaction term was statistically significant, indicating that for each region both the control strategy and the data type are significantly associated with the outcome variable. This is evidenced by the relative deviance values shown in Table 10.

When assessing agreement between the three herd datasets on lowest operational cost strategy, 493 instances (12% of a total of 4292 observations where each of the three herd data sets could be compared) produced the same recommendation. In 1681 instances (39%), the simulations based on the beta pert dataset identified the same lowest operational cost strategy as simulations based on the real dataset. Similarly, in 1801 instances (42%), the median and real datasets identified the same strategy as being the most cost effective. A Fleiss Kappa statistic was 0.04 ($p < 0.0001$) indicating very low levels of agreement. Agreement on the most cost effective control strategy between the median and real dataset and the real and beta pert datasets had very low Kappa statistics of 0.06 ($p < 0.005$) and 0.02 ($p < 0.005$) respectively. The size of the median ratio between the lowest cost operational strategy and the strategy indicated by the alternate (median or beta pert) dataset was 7 (5th percentile 1.2; 25th percentile 2.5; 75th percentile 30.5; 95th percentile 613). This distribution is shown in Figure 9 at the log scale and stratified by datatype.

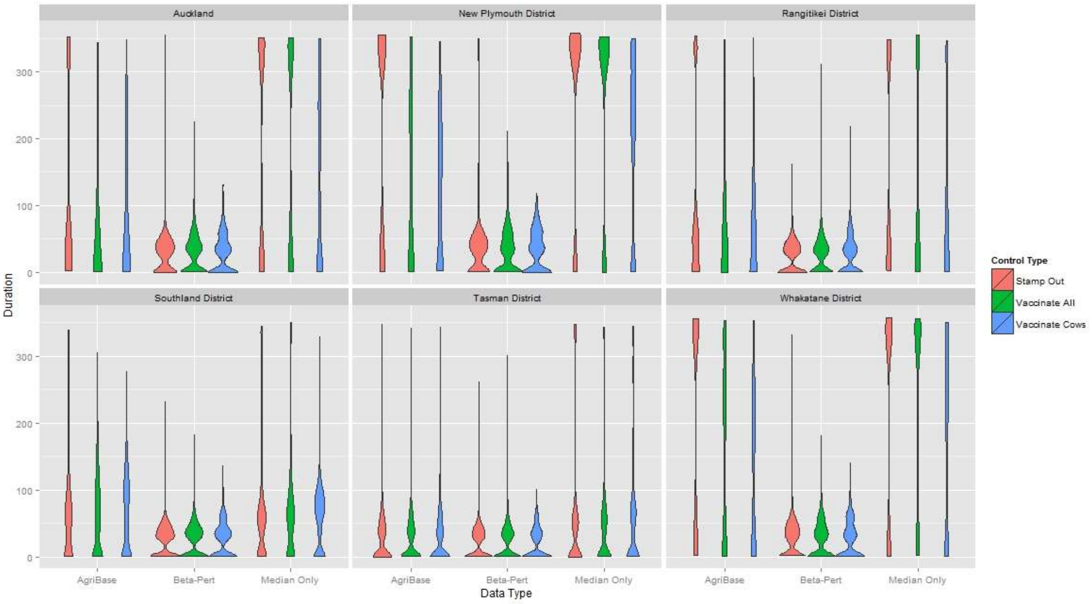

Figure 7: *Distribution of duration of outbreak for models using each of three herd size data sets across six New Zealand territorial local authorities (TLAs) for three control strategies.*

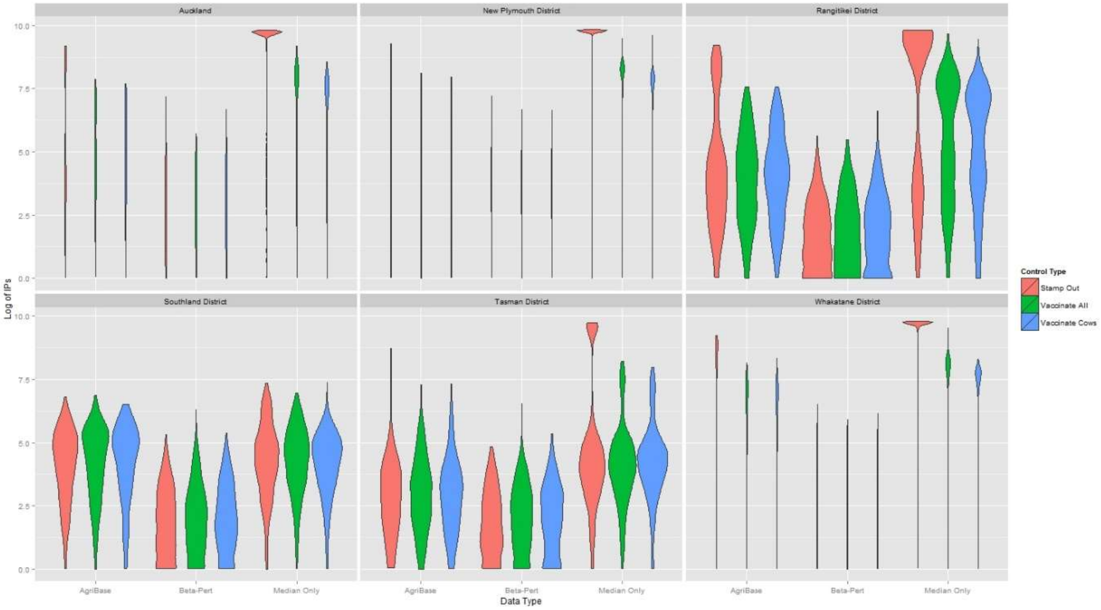

Figure 8: *Distribution of the log10 of final count of infected premises (IPs) for models using each of three herd size data sets across six New Zealand territorial local authorities (TLAs) for three control strategies.*

Table 9: *Number of simulations that generate outbreaks that are not detected (Not detected), outbreaks that last 365 days without being eradicated (Right censored), and number of simulations where the outbreak is detected and eradicated within 365 days (Detected and eradicated), by region and by dataset.*

	Beta-Pert			**Gold Standard**			**Median**		
	Not detected	**Right censored**	**Detected and eradicated**	**Not detected**	**Right censored**	**Detected and eradicated**	**Not detected**	**Right censored**	**Detected and eradicated**
Auckland District	1226	4	1770	1066	322	1612	1140	1037	823
New Plymouth District	1200	19	1781	745	679	1576	692	1576	732
Rangitikei District	1263	0	1737	997	173	1830	982	758	1260
Southland District	1256	0	1744	1012	5	1983	925	112	1963
Tasman District	1530	0	1470	1553	17	1430	1442	125	1433
Whakatane District	1337	1	1662	811	662	1527	928	1304	768

Table 10: *Analysis of variance (ANOVA) table for the Cox Proportional Hazards (CPH) models with infected premises (IPs) and duration as outcome variables. Explanatory variables were identical for both models. Given the large size of the dataset analyzed, the small p-values might be expected, however the large relative deviance for datatype provides an indication of the importance of this variable in the models.*

	Outcome variable : Count of Infected Premises				**Outcome variable : Duration**			
Explanatory variable	**Deviance**	**Chi2 Degrees of freedom**	**p value**	**Relative Deviance**	**Devaiance**	**Chi2 Degrees of freedom**	**p value**	**Relative Deviance**
Control type	656.4	2	<0.0001	328	519.2	2	<0.0001	260
Data type	9937.4	2	<0.0001	4969	10389.4	2	<0.0001	5195
Territorial local authority (TLA)	5997.6	5	<0.0001	1200	5353.5	5	<0.0001	1071
Control type * Data type interaction	574.6	4	<0.0001	144	596.0	4	<0.0001	149
Control type * TLA interaction	478.1	10	<0.0001	48	447.8	10	<0.0001	45
Data type * TLA interaction	2602.8	10	<0.0001	260	2526.5	10	<0.0001	253
Control type * Data type * TLA interaction	448.5	20	<0.0001	22	400.8	20	<0.0001	20

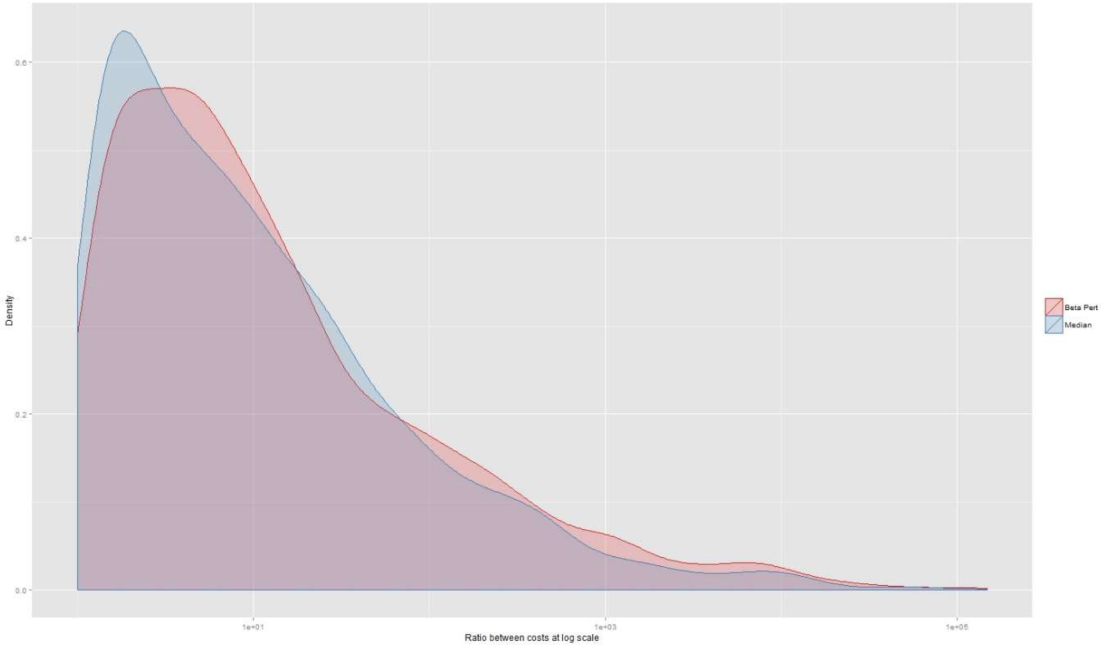

Figure 9: *Ratio between the operational costs generated by a comparison between the strategy indicated as being the lowest by real herd data-set and the costs of the strategy that would have been chosen should an alternate herd data set have been used. The x axis is displayed at the log scale. The median ratio is 7 (not at log scale) and the median of the cost ratio for those decisions using the beta pert data-set is 7.5 and the median ratio for the decisions made using the median data-set is 6.64.*

4.5 Discussion

Simulation studies have been performed in the UK to identify the effect of modelled farm location information on the performance of disease spread models (Tildesley and Ryan, 2012; Werkman et al., 2016). In these studies the optimal control policy is robust when using land cover but the epidemic sizes are not well predicted. In the Werkman et al 2016 paper, only small differences are seen when the scale of aggregation for the metapopulation model is very small. For larger sizes, the model does not perform well. These studies indicate that simulations using modelled farm locations have the potential to significantly differ from those for which farm locations are drawn from real data. To our knowledge there are no published studies in which model results using simulated animal counts are compared with results of models using real animal counts.

The accuracy of herd population data used in modelling can significantly influence the preparations for responding to disease outbreaks. In our study, the those model runs that use the gold standard herd sizes result in significantly different numbers of IPs and outbreak length when compared with those runs that used the modelled or median herd sizes. The size of the operational costs incurred based on decisions made using the median and beta pert herd sizes was seven times the cost of the decision made using the real herd data. This result suggests that accurate population datasets should be a priority to ensure effective decisions on the best available information in order to minimize the impact of disease outbreaks. The lowest operational cost was variable amongst the three datasets for the same seed farm (low values of Fleiss and Cohens Kappa). Although the seed herds are the same for each instance, AADIS introduces stochasticity into each run – so true agreement may be greater than what our calculations show. Another possible source of bias is that large outbreaks that run more than 365 days without finishing were removed from the comparative data analysis.

The number of model simulations that ran for unexpectedly long durations (resulting in right censoring) is strongly correlated with data type. Those simulations run with the median herd size are over represented in all TLAs, but those TLAs that have higher density of farms with susceptible species are most affected. Similarly, a beta-pert estimation of herd size results in an over representation of outbreaks that burn out before spreading further and remain undetected (Table 9). It is possible that as the model parameterisation of AADIS for New Zealand has not been fully tested and had as much time invested in it as the New Zealand Standard Model (NZSM), it could cause artefacts in these results (Owen et al., 2011). However, as the objective of this study was to examine the impact of the accuracy of herd-level populations on a disease spread model, we focus on these findings, comparing results between models to gather information on the importance of regionally representative herd size information, rather than making recommendations on specific disease control policy for FMD free countries. Particularly important in the context of our hypothesis is the large relative deviance for data types identified in the CPH models. The relative deviance for data types is approximately five times larger than the relative deviance for the next most influential variable in the model (TLA) when considering both duration and final count of IPs.

An underestimation of outbreak size or duration in a particular region based on incorrect or estimated herd and farm population information could be as damaging to response decision making efforts as an over estimation. Take the example of the duration in days in the Whakatane region (Figure 7 and Table 9). Here the beta-pert data would suggest that there is little benefit in augmenting stamping out with vaccination (median of 26 days duration regardless of strategy). The median herd size data suggest that stamping out will result in longer outbreaks when using the stamping out and stamping out augmented with vaccinating all species (a median of about 330 days) when compared with stamping out augmented with vaccinating cattle only (median of 215 days). Running the model with the gold standard dataset (a more accurate reflection of regional population heterogeneity), results in a more complex picture where stamping out results in longer median outbreaks (332 days) compared with either of the vaccination augmented strategies (152 days when vaccinating all species and 138 days when vaccinating cattle only).

Actual farm population information may not be available for a variety of reasons which include resource limitation and legal restrictions (Jewell et al., 2015). Where actual census data are not available, then a compromise may be to use modelled population data that are conditional on region as well as herd type rather than as a single function of herd type across the country, as has been done in this study. The optimal size for the regions that would best strike a balance between representing regional heterogeneity and best use of resource is not known. It might be argued that collection of representative samples from each locality to generate the conditional populations might require so much effort that the collection and use of actual data (which have other essential uses apart from disease spread modelling) would be a better use of resources. Our models indicate that population detail is more important in some areas of New Zealand than others and this is supported by previous work on farm animal populations in New Zealand (van Andel et al., 2017). The current study does not address the impact of subtle animal count biases on disease spread as it compares only the 'gold standard', beta pert distributed and median herd sizes for each of 10 herd types. The beta-pert representation of herd size was used based on the finding that the beta-pert distribution best fit the herd size distributions in the gold standard data. The herd type median was used as the final data set to include as InterSpread Plus, the current disease spread simulator used as the NZSM for FMD spread uses a probability of disease transmission based on the median herd size to drive disease spread in the model (18). While we have reasons for choosing both comparative datasets in our study, it is likely that inaccuracy in herd size data will, in practice not be distributed uniformly amongst all herds in the dataset and that some herd types will be more affected than others. We hope that our study will serve as a starting point for future, more nuanced studies which will explore the effect of sector specific data inaccuracies.

It is important to note that no resource constraints were applied in this set of model simulations. Our results reflect that (as expected) in the absence of constraints, it is preferable to vaccinate all susceptible species than to vaccinate cattle only. Note that in Table 11, there appears to be a paradoxical effect of vaccinating all species in the median dataset when compared with vaccinating cattle only. This is explained by the fact that model runs that exceeded 365 days were removed from the dataset. As shown in Table 9, there are more of these model runs in the parameterisations that use the median herd size dataset which leads to larger numbers of IPs. Further investigation of the response of the model to vaccine dose and human resource limitations will make interesting future work as will further investigation of impacts not limited to operational costs of an outbreak.

Our model findings are aligned with other published research that indicates that the value of vaccination is associated with the start location of an epidemic. Furthermore, while susceptible animal density does affect outbreak size, it does not alone predict infectiousness or infectivity of a herd (Keeling et al., 2009; T Porphyre et al., 2013). The presence of multiple species on a farm, the size of the holding and the distance to the closest infective property were risk factors identified in the 2001 UK epidemic (Paul R. Bessell et al., 2010). The authors point out that the accuracy of herd-level population and location information would be relevant to two (if not three) of these risk factors. In addition, should an actual outbreak occur, the strain type of the virus and its specific epidemiology would be hugely influential on the effectiveness of any chosen control strategy (Grubman and Baxt, 2004).

Our study indicates that when using a disease spread simulator that explicitly represents the spread of disease within farms the quality and origin of the data used to represent herd size has significant impacts on the model results. We recommend specific attention needs to be focused on national level animal population datasets that results in their alignment and more efficient utilization.

4.6 Appendix

Table 11: *Descriptive five number summaries (minimum, 25th percentile, median, mean, 75th percentile and maximum) for each of 2 outcome variables (count of infected premises and duration in days) and 3 explanatory variables for the models described in Table 8. These results are graphically represented in Figures 7 and 8.*

Data Type	Outcome variable	Region	Stamping Out (minimum, 25th percentile, median, mean, 75th percentile, maximum)	Stamping out augmented with vaccinating all susceptible species (minimum, 25th percentile, median, mean, 75th percentile, maximum)	Stamping out augmented with vaccinating cattle only (minimum, 25th percentile, median, mean, 75th percentile, maximum)
Gold standard herd size data	Infected Premises	Auckland District	0, 0, 19, 1658, 679.5, 10320	0, 0, 18.5, 264.2, 199, 2729	0, 0, 21.5, 275.9, 334.2, 2275
		New Plymouth District	0, 1, 2702, 3823, 8149, 11190	0, 1, 284, 591.6, 1073, 3394	0, 0, 284, 554.8, 994.8, 2961
		Rangitikei District	0, 0, 14, 644.5, 103, 10630	0, 0, 12, 119.3, 93.5, 1956	0, 0, 14, 122.2, 95, 1997
		Southland District	0, 0, 23, 83.08, 118, 928	0, 0, 20, 85.73, 129, 957	0, 0, 25, 89.42, 134.2, 692
		Tasman District	0, 0, 0, 59.53, 16.25, 6454	0, 0, 0, 24.3, 17, 1462	0, 0, 0, 29.41, 17, 1580
		Whakatane District	0, 0, 2120, 3531, 7187, 10460	0, 0, 245.5, 609.4, 1175, 3553	0, 0, 254.5, 561.9, 1079, 4320
	Duration in days	Auckland District	0, 0, 42.5, 109.5, 321.5, 355	0, 0, 40, 83.13, 123, 346	0, 0, 47.5, 85.71, 154, 349
		New Plymouth District	0, 27, 335, 201.2, 342, 357	0, 23.75, 128.5, 140.9, 247.2, 354	0, 0, 127.5, 126.8, 209, 347
		Rangitikei District	0, 0, 47, 87.34, 90, 355	0, 0, 45, 69, 103.2, 348	0, 0, 55, 69.79, 106.2, 352
		Southland District	0, 0, 54, 63.25, 89, 340	0, 0, 52, 59.38, 98, 306	0, 0, 61, 62.54, 106, 278
		Tasman District	0, 0, 0, 29.84, 47, 347	0, 0, 0, 28.34, 45, 342	0, 0, 0, 29.64, 46, 343
		Whakatane District	0, 0, 332, 193.6, 342, 357	0, 0, 152, 147.6, 258, 355	0, 0, 137.5, 130.5, 224, 354
Beta pert modelled	Infected Premises	Auckland District	0, 0, 3, 26.62, 24, 1350	0, 0, 3, 24.3, 26, 313	0, 0, 3, 25.49, 25.25, 796
		New Plymouth District	0, 0, 5, 53.37, 50.25, 1368	0, 0, 2, 33.94, 32, 805	0, 0, 4, 37.55, 39, 785

herd size data		Rangitikei District	0, 0, 1, 7.771, 7, 283	0, 0, 1, 10.51, 8, 244	0, 0, 1, 11.19, 8, 749
		Southland District	0, 0, 1, 8.901, 9, 206	0, 0, 1, 10.66, 9, 567	0, 0, 1, 10.1, 9, 222
		Tasman District	0, 0, 0, 6.898, 6, 129	0, 0, 0, 8.579, 7, 707	0, 0, 0, 7.982, 7, 214
		Whakatane District	0, 0, 1, 18.11, 16, 691	0, 0, 1, 16.52, 16, 373	0, 0, 1, 16.19, 14, 492
	Duration in days	Auckland District	0, 0, 30, 27.18, 43, 356	0, 0, 30, 27.78, 44, 226	0, 0, 29.5, 27.59, 45, 131
		New Plymouth District	0, 0, 34, 37.63, 50, 351	0, 0, 29, 30.9, 53, 212	0, 0, 34, 32.04, 57, 119
		Rangitikei District	0, 0, 27, 23.32, 40, 162	0, 0, 28, 25.63, 40, 312	0, 0, 27, 25.46, 41, 219
		Southland District	0, 0, 27, 23.4, 40, 232	0, 0, 28, 23.99, 40, 183	0, 0, 28, 24.88, 40, 137
		Tasman District	0, 0, 0, 18.96, 36, 263	0, 0, 0, 20.03, 37, 301	0, 0, 0, 19.52, 38, 101
		Whakatane District	0, 0, 26, 24.82, 42, 333	0, 0, 26, 25.65, 42.25, 181	0, 0, 25, 24.95, 43, 141
Median herd size data	Infected Premises	Auckland District	0, 0, 9510, 8281, 16750, 19160	0, 0, 207, 1365, 2816, 10220	0, 0, 173, 908.5, 1877, 5328
		New Plymouth District	0, 14, 17250, 12890, 17890, 19330	0, 13.5, 3108, 2660, 4128, 13800	0, 8.25, 2029, 1765, 2765, 14850
		Rangitikei District	0, 0, 33.5, 5657, 14700, 19220	0, 0, 27.5, 760.2, 1062, 16410	0, 0, 36, 465.6, 682.2, 13040
		Southland District	0, 0, 29, 114.7, 123.2, 1550	0, 0, 33, 98.43, 124.2, 1079	0, 0, 37, 77.89, 114, 1648
		Tasman District	0, 0, 4.5, 981.9, 63, 17980	0, 0, 2, 113.9, 59, 3877	0, 0, 2, 101.2, 68, 2969
		Whakatane District	0, 0, 15780, 10220, 17210, 18990	0, 0, 1902, 1843, 3334, 14150	0, 0, 1426, 1301, 2290, 4115
	Duration in days	Auckland District	0, 0, 326, 180.6, 341, 352	0, 0, 122, 167.5, 339, 353	0, 0, 108.5, 137.8, 261, 350
		New Plymouth District	0, 44, 340, 254.5, 343.2, 358	0, 57.25, 340, 247.3, 343, 352	0, 40, 230.5, 194.9, 294, 351
		Rangitikei District	0, 0, 58, 150, 336, 350	0, 0, 58.5, 130.4, 328, 357	0, 0, 66, 110.4, 213, 348
		Southland District	0, 0, 52, 71.41, 77.25, 346	0, 0, 57, 70.34, 86.25, 350	0, 0, 66, 56.13, 87, 330
		Tasman District	0, 0, 34, 56.34, 61, 348	0, 0, 28.5, 47.4, 63, 345	0, 0, 28.5, 46, 69, 345
		Whakatane District	0, 0, 337, 212.2, 343, 358	0, 0, 334, 206.6, 342, 358	0, 0, 215, 171.1, 284.2, 351

4.7 Acknowledgements

The authors thank Robert Sanson and AsureQuality for the use of the AgriBase dataset. Special thanks to Graeme Garner for his encouragement and help with parameterizing AADIS prior to this study. Further thanks are due to the Ministry for Primary Industries in New Zealand and to the Centre of Excellence for Biosecurity Risk Analysis at the University of Melbourne for funding.

CHAPTER 5

---

5 Evaluating the utility of national-scale data to estimate the local risk of foot-and-mouth disease in endemic regions

5.1 Abstract

Knowledge of the distribution of foot-and-mouth disease (FMD) is required if control programmes are to be successful. However, data on the prevalence and incidence of affected villages in developing countries with endemic disease is scarce. This is partly due to resource constraints as well as the logistical challenges of conducting intensive surveys and diagnostic testing in remote locations. In this study, we evaluated the performance of low resolution national scale data against high resolution local survey data to predict the FMD serological status of 168 villages in the Mandalay and Sagaing Regions of central Myanmar using both logistic regression and random forest modelling approaches. Blood samples for ELISA testing were collected from approximately 30 cattle per village in both the 6 to 18 month age range and in the over 18 month age range to distinguish between recent and historic exposure respectively. The results of the animal level tests were aggregated to the village level to provide the outcome of interest (village positive or not positive for FMD) and three explanatory datasets were constructed: using only nationally available data, using only data collected by survey, and using the combined survey and nationally available data. The true prevalence of FMD at the village level was 61% when only young animals were included, but increased to 87% when all animals were included. The best performing model was a logistic regression model using the combined national and survey data to predict recent infection in villages. However, this still incorrectly classified 40% of villages, which suggests that using national level data was not reliable enough for extrapolating prevalence in regions where conducting detailed surveys is impractical. Other methods for collected data on FMD such as the use of local reporting should be explored.

5.2 Introduction

Foot-and-mouth disease (FMD) is a rapidly spreading viral disease of ungulates that causes significant production losses due to high morbidity even though it causes limited mortality (Alexandersen et al., 2003b). The presence of FMD in a country limits the options for primary sector exports and so countries with chronic endemic FMD as well as countries that have historically been FMD-free but have experienced a recent outbreak have strong economic incentives to mount national eradication campaigns. To assist countries in managing FMD, the World Organisation for Animal Health (OIE) and the Food and Agriculture Organisation of the United Nations (FAO) have developed the progressive control pathway (PCP), which is a systematic multi-stage approach for countries with foot-and-mouth disease (FMD) to work toward becoming disease free in seven stages (Rweyemamu et al., 2008; Sumption et al., 2012). Depending on the specific goals of the country, activities may be aimed at reaching a specific stage rather than necessarily progressing to freedom. The first stage is to understand the country specific epidemiology of FMD including its prevalence, spatial distribution, and transmission pathways. Once this is adequately understood, the competent authority in the country can design and implement an FMD control programme for Stage 2 that may include measures such as movement restrictions, vaccinations, and ongoing surveillance leading to the establishment of disease-free zones. Stage 3 has the objective of supressing virus transmission by preventing clinical disease. During Stage 4, the country applies to the OIE to be recognised as free from FMD with vaccination in designated zones. In Stage 5 the objective is to transition the designated zone from free with vaccination to free without vaccination. This is followed by expansion of the free zones (Stage 6) and prevention of reintroduction of FMD (Stage 7).

One of the challenges with implementing the PCP is that there are often insufficient resources to conduct comprehensive sampling across the entire country and so the competent authority must determine how to prioritise limited available resource to make the greatest impact and gather the most useful knowledge. Given that local conditions, farming practices, and animal populations often vary greatly across a country, it is likely that certain regions will have a greater prevalence of FMD than others and would therefore be good candidates to target with increased disease control activities. Describing this local heterogeneity is best accomplished by performing representative surveys, but the cost and complexity of such a task can be daunting and prohibitively expensive. As most countries have some form of national census/land-use data and because remotely sensed data are increasingly available, the use of existing data is an attractive option particularly since these types of remotely sourced variables have provided encouraging results when used as predictors of FMD risk (Hamoonga et al., 2014; Nampanya et al., 2013; Sumption et al., 2008). However, although these national level datasets are available, estimates of their completeness and accuracy (in the case of census data) are lacking and it is not well known how well remotely sensed national level data predict local risk and what the implications of inaccuracies in this data are for planning and executing FMD control policies.

Aside from issues with data quality and scale, questions about the most appropriate statistical approach to identifying FMD risk factors and predicting FMD risk still remain. Logistic regression models are a traditional choice, but recently, random forest modelling has been proposed as an alternate approach. Little is published on the comparative performance of the two model approaches in veterinary epidemiology. Logistic regression is considered to be a standard approach to predicting a binary outcome (for example, outbreak or no outbreak in the period of interest), especially when modelling is undertaken, not only for prediction but also to describe and increase understanding of risk factors that could cause the outcome of interest (Langford et al., 2009). By contrast, the random forest prediction algorithm (a machine learning technique) is focused on prediction rather than explanation of risk factors. Furthermore, RF models have been found to out-perform logistic regression models in cases where prediction is the primary goal of the modelling activity (Breiman, 2001; Brownstein et al., 2005; Cutler, 2007; Nicolas et al., 2016; Prasad et al., 2006).

The South East Asia and China FMD (SEACFMD) project launched in 2015 to support regional progression on the PCP offers a unique opportunity to explore the impacts of different levels of information and different statistical models to predict the locations of FMD affected villages and the incidence in an endemic region. Our study makes use of PCP Stage 1 activities conducted in Myanmar and contributes to understanding of disease distribution and risk factors. The prevalence of FMD in Myanmar is not well documented but studies performed in 2010 showed endemic levels in Sagaing region (Kyaw, 2010); which is included in the current study area. Cattle in Myanmar are farmed in traditional small-scale systems with practices that include communal grazing of animals on common land (either tethered or free roaming), cut-and-carry systems and in some areas, migratory grazing practices. Most cattle are used for draft and meat with a small minority used for dairy. A previous study in the project area found that water was available year round in the area and that only 12% of farmers had to cart water between December and March each year (Kyaw, 2010). Traditional treatments for FMD in the study area include rubbing of banana and honey on the tongue lesions and walking the animals over hot sand for foot lesions (Kyaw, 2014, 2010).

Using the detailed serological and survey data collected for 168 villages in 25 townships in the central region of Myanmar as well as national scale data, our study evaluates the use of village-level risk factors collected by survey, national-level census data and national level remotely sensed data to estimate the local risk of foot-and-mouth disease in endemic regions using both logistic regression and random forest models. The predictive utility of the survey information can be compared with that of the available national data to evaluate performance. The implications for disease planning and implementing disease control measures are also discussed.

5.3 Materials and Methods

Selection of villages

According to the Myanmar Information Management Unit (MIMU), Myanmar is divided into 67 districts with 330 townships and 70,838 8 villages in total. MIMU (http://themimu.info/about-us) is a service under the management of the United Nations Resident, which aims to strengthen the coordination, collection, processing, analysis and dissemination of information to improve the capacity for analysis and decision making in Myanmar.

As part of a SEACFMD project involving 24 townships across 2 regions (Mandalay and Sagaing) in central Myanmar, a cross sectional survey of villages was conducted between 1 October and 31 December 2016 (Figure 10). The whole project area contains 4,615 villages. The primary goal of the survey was to establish the prevalence of FMD at the animal level at the start of the risk-based FMD vaccination programme in 2016. Sampling was conducted using a two-stage design first selecting 133 villages to visit and then 30 animals per village for a total sample of 4,000 cattle. Thirty animals per village was the minimum sample size required to detect a biologically meaningful difference in FMD prevalence at the village level if one was present on follow up sampling at the end of the programme. The estimated prevalence at the animal level at the beginning of the programme was 10% and the survey was designed to give a 90% chance of detecting a 50% reduction (prevalence decreased to 5% at the follow up survey) with 95% confidence (Dohoo et al., 2009). Villages were selected for inclusion in the survey from the MIMU register based on their location in the project area, logistical constraints identified by local experts, and willingness of the village to participate in the programme. Due to extra resources being made available to the project, a total of 168 villages were included by the end of the study.

Classifying village FMD status

Villages were visited once each by the field teams in the study period in 2016 during the months of October, November, or December. During the visit, the teams collected blood samples from a total of 30 animals distributed across 6 households in each village. The households within villages were chosen by systematic random sampling. The animals within each household were selected at random from the two different age groups: 6 to 18 months of age and animals older than 18 months of age. The purpose was to allow comparison in seroprevalence between age classes to make inferences about both recent and historical infections with FMD.

Blood was collected by jugular venipuncture in the field by regional veterinary staff. The samples were placed in the shade at a 45 degree angle and allowed to settle in disposable syringes for an hour until the serum and red cells had separated. Serum was then decanted into 1.8ml plastic tubes (Cryovial by Sigma Aldritch) and stored in the regional departmental office refrigerators before being delivered to the national laboratory in Nay Pyi Taw. The time between blood collection and delivery to the national laboratory varied between 1 and 3 days depending on distance from the regional office to the national laboratory. Serological testing was performed using a commercial non-structural protein (NSP) ELISA test kit (PrioCheck FMDV by Prionics) using a cutpoint of 50% inhibition as specified by the manufacturer (Sørensen et al., 1998). The PrioCheck FMD ELISA has published specificity in cattle of 95 to 99% depending on the vaccination status of the cattle in question (Fukai et al., 2013; Sørensen et al., 1998). The sensitivity of this test in experimentally infected animals nears 100% (Brocchi et al., 2006).

The outcome variable in our models was the serological prevalence in young stock (6 – 18 months of age) in the survey villages. This was represented as a binary outcome where the cut point for assigning the outcome of positive or negative was the presence of more than 10% of stock between 6 and 18 months in a village with more than 50% inhibition on NSP ELISA. The unit of interest for our models was the village.

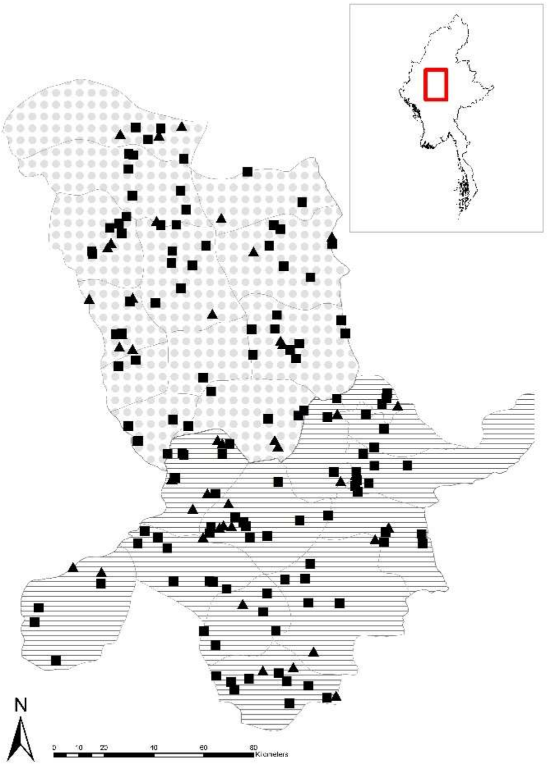

Figure 10: *Twenty four project townships in Myanmar (inset top right). Striped townships are in Mandalay Region and spotted townships in Sagaing Region. Both the Mandalay and Sagaing regions are in central Myanmar. Survey villages are indicated on the map as triangles (used in the model testing dataset) and squares (used in the model training dataset).*

<u>Village-level risk factor data</u>

At the same visit, a questionnaire was delivered to the village headman of each village to collect detailed information on any actions taken if FMD was previously identified in a village, as well as data on the village level household and animal demographics. Additional questions were asked about grazing practices, contact with other villages, the presence of cattle traders in the village and previous vaccination (Table 12). For each village listed by MIMU as being located in the project area, a list of exposure variables was compiled to create a data frame used for further analysis.

The proportion of households in each village that owned FMD susceptible stock was calculated by simple division. The regionally common grazing practices of communal grazing, cut-and-carry grazing and migratory grazing were recorded for each month of the year and aggregated to represent a proportion of the year in which the practice was undertaken in each village. Village headmen were asked which neighbouring villages had bovines, which shared communal grazing areas with bovines from the survey villages and a count of in-contact villages was generated. The reported FMD control strategies were examined and aggregated into a single categorical variable with two levels: villages vaccinated against FMD in the face of an outbreak and villages that undertook other intervention measures. New explanatory variables were generated for those villages that said that they reported FMD outbreaks to a central authority (yes or no) and those that reported that they limited fomite contact routes within the village (yes or no). The free text answers to the questions about the actions of farmers in the face of an FMD outbreak were analysed and binary explanatory variables created for if farmers sold clinically affected FMD animals, or if they reported tried to treat them. The free text answers to the question about the fate of dead FMD animals were similarly categorised into villages where these animals were burned, buried, sold for meat or eaten by the owner.

Table 12: *Survey questions presented to the village headmen of 168 villages in the project area*

Question	**Variable type**	**Answer type**
How many households are in the village?	Numeric	Count of households
How many of these households keep cattle, buffalo, goats, sheep or pigs?	Numeric	Count of households
What is the total number of cattle (dairy and draft separate), buffalo, goats, sheep, pigs in this village today?	Numeric	A count of each species
During which months do you do the following grazing practices of cattle and buffalo? Repeat question for each of common grazing, restricted grazing (cut and carry) and migration	Categorical	An indication by month of grazing type for each of the 3 types
Is livestock from your village in contact with livestock from other villages as a result of grazing practice or watering	Boolean and a list of villages which were counted and	Yes or No and if yes, a list of village names

points? If yes, please list the villages	treated as numeric	
How do you control FMD outbreaks in your village?	Categorical	Open question with common strategies as tick boxes and a space to record 'other'. Options are: treat using antibiotics, vaccinate, treat with traditional medicine, slaughter, report, limit sharing of equipment, do nothing, up to the farmer
If farmers have animals that are sick with FMD, what do they do?	Categorical	Open question with common strategies as tick boxes and a space to record 'other'. Options are: sell, treat, walk on hot sand, do nothing, don't know
If farmers have animals that die of FMD, what do they do?	Categorical	Open question with common strategies as tick boxes and a space to record 'other'. Options include: bury, burn, sell the meat, eat the meat, don't know.
Did you vaccinate any livestock in the village for FMD in the last 12 months? If yes how many each of cattle and buffalo.	Boolean and numeric	Yes or No, and if yes, counts of species vaccinated
Are there livestock middlemen or traders in your village?	Boolean	Yes or No

National-level risk factor data

Census information relating to bovines (cattle and buffalo), pigs and small ruminants (sheep and goats) were available from the LBVD officers. LBVD also supplied the locations of cattle markets in the project area which had been compiled by experts within the department, for animal management purposes. Exposure variables considered during national data model building are detailed in Table 13. Haversine distances from each village to various features were calculated using the rgdal package in R (Bivand et al., 2014b). Where variables were available as densities held in raster format, values were extracted to the villages using the raster package in R (Hijmans, 2015). Livestock units were calculated based on the national GLOW data according to the definition used by the European Union (Anonymous, 2013b) which allocates 0.8 LSU to a beef cow, 0.1 LSU to sheep and goats and 0.4 LSU to a pig. All cattle were assumed to be beef cattle in our context as dairy cattle in the village level data were sparse and could not be separated out in the national modelled data. Data visualisation was done in ArcMap 10 (ESRI, 2018).

Table 13: *Data sources and spatial data layers used to build a data frame of explanatory variables used to explore the usefulness of national level data for modelling FMD prevalence at the village level in 25 townships in central Myanmar. Sources include the Livestock Breeding and Veterinary Department in Myanmar (LBVD) and the Myanmar Information Management Unit (MIMU)*

data	**original source**	**variable derived (unit of measure)**
Sale yards	LBVD	distance from each village to the nearest cattle market calculated (km)
village locations	MIMU	used in calculations to roads, railways, borders etc
Any road	MIMU	distance from each village to the nearest road calculated (km)
Tracks	MIMU	distance from each village to the nearest track (secondary or tertiary road in km) calculated
railroads	MIMU	distance from each village to the nearest railroad calculated (km)
human population density	MIMU	each village assigned a human population value based on a density surface
pig populations	LBVD	each village pig population collected by LBVD livestock officers
bovine populations	LBVD	each village cattle and buffalo population collected by LBVD livestock officers
small ruminant populations	LBVD	each village sheep and goat populations collected by LBVD livestock officers
Distance to an international border	MIMU	Distance from each village to the closest international border was calculated (km)
Susceptible Livestock Units	LBVD	Calculated by adding the count of bovines, 0.1 of the count of small stock and 0.4 of the pig population of each village (LSUs)

Statistical Analysis and Model Building

Descriptive statistics

The true and apparent prevalence estimates at the village level were calculated in an online calculator (Sergeant, 2019) using standard statistical methods to estimate apparent and true prevalence and their confidence limits (DasGupta et al., 2001; Reiczigel et al., 2010; Rogan and Gladen, 1978). Descriptive statistics were provided on the distributions of exposure variables displayed in Table 14 (national variables), Table 15 (survey variables – continuous) and Table 16 (survey variables – categorical).

Model data

Three sets of explanatory variables were generated for testing in the models. The first used village-level survey variables only, the second used nationally available variables only, and the third model was built using the combined village and national level data. Our hypothesis was that the latter would have the greatest explanatory power. The data were imported into the R statistical software package (Team, 2017) for processing and analysis. Survey village data from the project area were partitioned into a training (0.75) and testing (0.25) dataset by random allocation. The same training dataset was used to construct all the models and the testing dataset was used to define the comparative performance of the different models. The training dataset contained 126 survey villages. Models were built using the 126 survey villages and predictions made to the remaining 42 villages for comparison with actual survey results.

Logistic regression models:

A Spearman correlation ($|r|$) matrix was used to identify significant pairwise collinearity amongst the predictor variables when considering a linear regression framework. This was defined arbitrarily as a correlation of $|r| > 0.7$. Our choice to use the Spearman's correlation coefficient rather than a variable selection process within the modelling framework allowed the analysts to make variable selections based on consultation with local experts on the biological system under examination. As the distances from villages to all roads or tracks were collinear, it was decided to include the distance from each village to the closest track in the testing phase of logistic regression model building. Local experts confirmed that cattle movement on foot commonly takes place via these routes, so inclusion made biological sense. Continuous variables were categorised and tested in the models when they were not normally distributed. Variables were then included or excluded in the model building on the training dataset, based on stepwise backward exclusion. Biologically relevant interaction terms were tested for significance. The final model on the combined data sets (survey and national) included all the explanatory variables from the survey data model, with the addition of distance to a cattle sale yard and distance to a railway. An interaction term was tested between the presence of traders in a village and distance to a sale yard, but found to be statistically insignificant ($p > 0.3$). Variables regarded as highly biologically significant (for example cattle sale yards), but which were not statistically significant at a p value of 0.05 in the model building process, were forced into the final models. A random effect for Township was tested for significance and the residuals of the final model were checked and examined for spatial clustering prior to the models being run on the testing dataset. Comparison between models during the model building process was made by Chi squared test to compare the residual reduction in the sum of squares and by McFadden's pseudo R squared to compare the variability explained by each model (Cochran, 1954; McFadden, 1974).

Random forest analysis

A random forest (RF) machine learning method was used. The optimal parameters were assessed prior to fitting the final models. RF modelling does assume that the sample is representative of the population, a common statistical assumption. RF is a bagging algorithm that creates decision trees by repeatedly selecting bootstrap samples from a training set, and fitting trees to each replicate. Bagging randomly generates a set of data from the original with replacement. A random sub-sample of the predictors is used to split each node in a tree (Vinet and Zhedanov, 2010). The given number of trees is averaged to obtain the estimates. Fitting variables can be changed in the model, including the number of trees to be grown and the number of randomly selected variables to be used as candidates at each split. RFs were built using the 'randomForest' package in R with the number of trees set to 2000 (Liaw and Wiener, 2002; Team, 2017) . The random forest algorithm identifies the most important variables in the model. The most reliable method is a comparison of the error obtained from a model with the real value of an explanatory variable with a model built with a random permutation of this value (Strobl et al., 2007). The value is expressed as a percentage change in mean squared error (MSE). Larger values indicate more important variables in the model and variables with negative values add no predictive ability to the model.

Comparing models

Models were compared based on their ability to correctly predict the outcome variable in a compartmentalised subset of the project area data (the partitioned testing dataset). Predictions from models and the true serological status for each of the villages were tabulated to compare performance, true positives and negatives and false positives and negatives were recorded. True positives and true negatives were counted together, but false positives and false negatives were counted separately

5.4 Results

Descriptive statistics

Of the 168 villages surveyed, only 2 had no evidence of circulating antibodies against FMD in any of the animals tested. A further 24 had less than 10% of the animals sampled test positive on ELISA. When only the 6 to 18 month old animals were considered, 67 of the 168 villages had no evidence of seroconversion in the young stock and 4 had seroprevalence in young stock of under 10%. This yields an apparent prevalence (AP) of 56% (CI 48-63%) and a calculated true prevalence (TP) of 61% (CI 51-72%) in young animals and an AP of 79% (CI 72-84) and TP 87% (CI 78 – 94%) if the whole population is considered.

Table 14: *Six number summaries of explanatory variables derived from national level data in a project region of central Myanmar.*

	minimum	**25%**	**50%**	**mean**	**75%**	**maximum**
Distance to nearest sale yard (km)	10	63.5	91	130	137	282
Distance to any road or track (km)	0	1	3	4	6	13
Distance to a road (km)	0	2	5	6	10	18
Distance to a track (km)	0	3	7	8	12	31
Distance to a railroad (km)	0	4	8	10	15	39
Human population density (people/km squared)	86	138	184	220	250	637
Pig population at the village level	0	6	22	40	50	311
Sheep and goat population at the village level	0	0	69	152	200	1379
Cattle and buffalo population at the village level	38	129	237	322	435	1647
Livestock units at the village level	38	144	253	353	499	1671
Distance to an international border (km)	0	50	93	94	131	193

Table 15: *Six number summaries of continuous explanatory variables derived from the survey questionnaire data in a project region of central Myanmar.*

	minimum	**1st quartile**	**median**	**mean**	**3rd quartile**	**maximum**
Households	7	102	165	256	273	5873
Households with livestock	0	55	100	168	180	3725
Proportion of households with livestock	1	36	71	68	100	100
Draft cattle	0	130	250	432	500	3500
Dairy	0	0	0	12	0	400
Buffalo	0	0	0	5	0	327
Goats	0	0	60	170	200	2450
Sheep	0	0	0	44	30	716
Pigs	0	10	29	51	60	516
Common grazing proportion of the year	0	0.2	1	0.7	1	1
Restricted grazing proportion of the year	0	1	1	0.87	1	1
Migratory grazing proportion of the year	0	0	0	0.01	0	1
Count of in-contact villages	0	1	3	3	4	10
Vaccinated cattle in previous year	0	0	0	83	53	2500
Vaccinated households in previous year	0	0	0	41	16	3000

Table 16: *Counts of categorical explanatory variables derived from the survey questionnaire data in a project region of central Myanmar. Aside from these variables with binary outcomes, when village headmen were asked if any treatment was undertaken for FMD, 3 answered that nothing was done, 3 claimed that FMD affected animals were slaughtered, 43 said they vaccinated in the face of an outbreak and the remaining 119 applied other treatments. Results have been stratified by the village level status showing if the village is regarded as a case of FMD (serological evidence of exposure in young stock).*

	Yes			**No**		
Village level status	**case**	**not case**	**total**	**case**	**not case**	**total**
Direct contact with other villages	72	64	136	21	11	32
Report FMD	21	21	42	71	55	126
Limit fomite contact in the face of an outbreak	20	15	35	69	64	133
Sell clinical FMD animals	8	4	12	85	71	156
Treat clinical FMD animals	74	67	141	16	11	27
Bury dead animals with FMD	23	18	41	66	61	127
Burn dead FMD animals	5	0	5	84	79	163
Sell the meat of dead FMD animals	6	9	15	86	67	153
Eat the meat of dead FMD animals	4	9	13	89	66	155
Vaccinated in the previous year	37	26	63	54	51	105
Traders present	42	42	84	51	33	84

Risk factors identified

In the logistic regression model using only survey data, the proportion of households with livestock, the counts of draft cattle, counts of goats, counts of pigs, the presence of traders, the count of contact villages, the proportion of the year spent on communal grazing, the practices of burning FMD carcasses, selling clinical FMD animals and burying dead FMD animals were found to be significant (Table 17). An interaction term between the proportion of households with susceptible stock and the count of goats and another interaction term between the proportion of the year spent on communal grazing and the count of contact villages were added to the model as they provided additional explanation of variability (McFadden’s pseudo R squared statistic 0.14 vs 0.11 for the model without interactions). A logistic regression model using only national data found that only the cattle population was useful for predicting the outcome, but the distance to the closest sale yard was retained in the model based on its perceived significance. The McFadden R squared statistic for this model was 0.04.

Combining both the national and the survey datasets increased the McFadden R squared statistic only marginally to 0.17.

When the survey data were used to build a random forest model, the most important explanatory variable based on the impact on MSE was the presence of buffalo in a village (6.16%), followed by the sale of meat from animals that died with clinical signs of FMD and attempts to treat animals for FMD. The next most important variables were the goat and pig populations (Table 18). When only national level data were used in a random forest model, the most important were: township (5.35%), pig populations (1.65%), distance to cattle sale yards (0.18%) and bovine populations (0.16%). When data were combined, township name was the most important variable (20.3%), followed by counts of draft cattle and buffalo (7.3 and 6.7% respectively).

Model performance

The model that performed best was a logistic regression model using a combination of survey and national data. This model correctly predicted the FMD status of 25 of the 41 test villages. Eight villages were identified as false positives and another 8 were identified as false negative. Results of the other 5 models are presented in Table 19.

Table 17: *Regression table of coefficients and p values for three logistic regression models where the outcome was the presence of FMD in young cattle (6-18 months old) in central Myanmar. The three models use survey data, national data and a combination of survey and national data to model the serological presence of FMD in young stock at the village level.*

Model	**Questionnaire model**	**National model**	**Combined model**
	coefficient (p value)	**coefficient (p value)**	**coefficient (p value)**
Distance to nearest sale yard (km)		-0.006 (0.07)	- 0.004 (0.017)
Count of bovids		- 0.002 (0.05)	
Proportion of households with susceptible livestock	- 0.11 (0.9)		- 0.18 (0.7)
Count of goats	-0.007 (0.6)		- 0.006 (0.6)
Presence of traders (no is referent)	-0.9 (0.03)		- 1.02 (0.02)
Does the village sell animals clinically affected with FMD (no is referent)	1.4 (0.07)		1.45 (0.06)
Dead animals with clinical FMD are sold for meat (no is referent)	- 1.6 (0.03)		-1.47 (0.05)
Dead animals with clinical FMD are buried (referent is yes)	-0.8 (0.13)		- 0.96 (0.06)
Interaction between the proportion of households with susceptible stock and the count of goats	0.05 (0.07)		0.004 (0.1)
Count of in-contact villages			- 0.17 (0.2)
Proportion of the year where common grazing is practiced			-0.18 (0.7)
Interaction between the proportion of the year where common grazing is practiced and the count of contact villages			0.16 (0.03)
Distance to nearest railway (km)			0.03 (0.07)

Table 18: *variable importance in three random forest models predicting FMD at the village level in central Myanmar. One model uses only questionnaire data, one uses only nationally available data and the third uses data combined from both of these sources.*

variable description	**% MSE combined model**	**% MSE questionnaire model**	**% MSE national model**
Count of bovids	-	-	-0.39
Count of combined sheep and goats	-	-	0.68
Township name	20.32	-	21.74
Count of draft cattle	7.27	1.94	-
Count of buffalo	6.66	6.16	-
Distance from an international border	6.24	-	7.53
Proportion of the year where common grazing is practiced	5.12	-1.41	-
Human population	4.7	-	3.72
Direct contact with other villages	3	3.09	-
Does the village treat cases of FMD	2.95	3.89	-
Distance from secondary roads and tracks	2.65	-	-1.4
Proportion of households vaccinated for FMD in previous year	2.61	2.1	-
Distance from a sale yard	2.38	-	3.18
Count of pigs	2.24	3.41	5.83
Dead animals with clinical FMD are sold for meat	1.98	5.32	-
Count of goats	1.84	3.46	-
Contact via fomites is limited in the face of an outbreak of FMD	1.12	1.45	-
Does the village report cases of FMD	0.35	-4.51	-
Dead animals with clinical FMD are buried	0.34	1.12	-
Count of vaccinated buffalo	0	0	-
Does the village sell animals with clinical FMD	-0.13	0.37	-

Dead animals with clinical FMD are eaten	-0.18	0.09	-
Count of vaccinated cattle	-0.21	-1.74	-
Count of in-contact villages	-0.22	-1.1	-
Proportion of households with FMD susceptible livestock	-0.41	0.72	-
Dead animals with clinical FMD are burned	-0.59	1	-
Presence of traders	-0.94	0.62	-
Proportion of the year where restricted grazing is practiced (cut and carry)	-1.22	-4.63	-
Count of sheep	-1.26	-4.19	-
Vaccinated in the previous year	-1.75	-2.46	-
Proportion of the year where migratory grazing is practiced	-1.78	-1.19	-
Count of dairy cattle	-1.92	-5.98	-
Distance from a railway	-1.92	-	3.15
Treatment options for FMD	-2.98	-5.54	-

Table 19: *Counts of false negative, false positive and correctly identified villages from the test dataset identified by 2 model types (random forest models and logistic regression models) over 3 datasets (questionnaire data, nationally available data and combined data). The testing data consist of 41 records. Percentages shown in brackets with a 95% CI around the estimates following the comma.*

	Logistic regression model			**Random forest model**		
	Correct	**FN**	**FP**	**Correct**	**FN**	**FP**
Questionnaire data	20 (48%, 33-65)	6 (14%, 6-29)	15 (37%, 22-53)	20 (48%, 33-65)	6 (14%, 6-29)	15 (37%, 22-53)
National data	20 (48%, 33-65)	11 (26%, 14-43)	10 (24%, 12-40)	23 (56%, 40-72)	4 (9%, 3-23)	14 (34%, 20-51)
Combined data	25 (61%, 45-76)	8 (20%, 9-35)	8 (20%, 9-35)	21 (51%, 35-67)	5 (12%,4-26)	15 (37%, 22-53)

5.5 Discussion

In evaluating the use of different data sources and statistical models for predicting FMD positive villages, our study identified the best performing model as being a logistic regression model using both the national data and survey data to predict the presence of serological reactors to FMD in 6 to 18 month old stock. However, as many as 40% of villages were still incorrectly classified, which could result in suboptimal allocation of resources if this model was used as a tool for predicting the distribution of FMD across Myanmar. It is interesting to note that the logistic regression model including only the variables of cattle density and distance from a sale yard, obtained from the national level data, provided similar predictive ability to the logistic regression model built with the survey data alone. The combined datasets, in a logistic regression framework, performed only a slightly better than the random forest model, built with only national level data. This result was somewhat surprising given that random-forest models are generally considered to have greater predictive accuracy, as they easily fit complex relationships and interactions without the need to explicitly specify them, they do not over fit and are stable with noisy data (Cutler, 2007; Elith et al., 2006; Hollings et al., 2017; Robinson et al., 2014). However, it should be noted that although the random forest model had a large number of false positive results (34%), it also had the lowest number of false negatives produced by any of our models (9%).

The risk factors identified by our models all make biological sense given the current knowledge surrounding the epidemiology of FMD (Ferguson et al., 2001). In particular, we found that the risk of a village having seropositive cattle increased with the number of contacts with other villages via common grazing and the greater the proportion of the year in which common grazing occurred. An interesting finding is the interaction between the size of the livestock population as a proportion of households in a village and the number of goats in a village. While the role of small ruminants has been widely discussed in historically disease-free countries, these species are not a current area of focus for efforts in Myanmar (Kitching and Hughes, 2016). Our results highlight the need for further investigation into seroprevalence in these species, that are highly mobile and have limited or absent clinical signs of FMD (Barnett and Cox, 1999).

The significance of the exposure variables associated with how villages respond to FMD should be interpreted with some caution since it is not clear from the data if the burying and burning of FMD carcasses and the selling of FMD affected animals are causal factors in disease prevalence or if they are the inevitable result of having an occurrence of disease. Further investigation with more carefully recorded temporal information is advised before comment on the importance of these practices for disease spread can be made. We found that the further a sale yard was from a village, the smaller the risk of infection with FMD (negative coefficient). While this is biologically plausible, it is unclear how the presence of cattle traders in a village could provide a protective effect (positive coefficient) according to our regression model. We are aware of ongoing studies on cattle movements and trading patterns in Myanmar and believe that this will provide valuable insight into this area.

Our study uses serological status to allocate the FMD status of a village. Importantly we limit our interpretation to those animals between 6 and 18 months of age in an attempt to quantify recent viral spread and to avoid the effect of maternal antibodies (Hammant et al., 1984; Kitching, 2002). Despite these measures, the sensitivity and specificity of ELISA testing is not perfect and the performance measures of the test as specified by the manufacturer are based on experimental infections under controlled conditions. In an endemic situation under field conditions, both false positive and false negative results at the animal level can be expected. As the number of animals sampled in any given village was around 30 and approximately one third of these were in the 6 to 18 month age group, the impact of a single false negative or false positive result is potentially high. That said, similar to our results where only 2 villages had no seropositive animals, a study in 2010 mentions that every village sampled in the Sagaing region had at least one seropositive animal (Kyaw, 2010).

Another issue with our method of indicating the best performing model is that the size of our testing data set is limited, so that the 95% confidence intervals around the proportions of correct predictions, false negatives and false positives overlap, which precludes a definitive statement on the best performing model. In addition to this, there are sources of bias in our study. These include the fact that because interviews were performed by regional LBVD staff, significant interviewer bias could have been introduced at the township level, which could explain the importance of township as an explanatory variable. Furthermore, the questions relating to if FMD was reported and if fomite spread was limited in the face of an outbreak, may have been confusing and so it is not clear if the low level of reporting indicated by the survey (25% of villages indicated they would report an outbreak of FMD, 95% CI is 19% - 32%) is in fact correct. The question was posed as an open ended question about actions taken to control an outbreak at the village level and one of the actions that was mentioned was that disease was reported to LBVD. Despite possible bias in this estimate, other studies have also reported severe under reporting to central authorities in the general SE Asian region and in other endemically affected developing countries (Gleeson, 2002b; Kenton L. Morgan et al., 2014; Sumption et al., 2008).

Our results suggest that while a combination of nationally available data and survey data provide the best predictive ability, extrapolating prevalence estimates from these models is not likely to provide reliable risk information for all the villages in the population. A method to collect information from public reports of FMD (as opposed to centrally collated disease reports) could be explored as a way to collect additional data points from more villages. The use of mobile technology to collect disease reports is successful in developing countries in the context of human health programmes and for zoonosis like avian influenza (Déglise et al., 2012; Madder et al., 2012). Its application for collection of FMD reports in Myanmar would be heavily influenced by the ability of farmers to identify FMD in an endemic situation and on the perceived importance of reporting (Bellet et al., 2012b; Calba et al., 2015; Mariner et al., 2011; Kenton L. Morgan et al., 2014; Sumption et al., 2008). Further research into alternative ways to collect FMD prevalence information and ways to address under reporting of FMD in endemic situations is required.

5.6 Acknowledgements

The authors acknowledge members of the OIE-LBVD FMD Project Team including Kyaw Thinn, Assistant Director, Myo Thiha, Research Officer, and Kyaw Htay, Research Officer, for their efforts in collecting data from the field to be able to complete this study. In addition, we greatly appreciate the direct contributions made by LBVD regional staff in Mandalay and Sagaing who directly participated in this study and openly shared available date and knowledge from their respective townships. We gratefully acknowledge the support provided by the DLBD of the Ministry of Agriculture, Govt. of Myanmar, and the permission to use national census data. This report is generated for the Myanmar section of the South East Asia and China Foot and Mouth Disease (SEACFMD) Campaign run by the OIE (World Animal Health Organisation).

CHAPTER 6

---

6 Estimating foot-and-mouth disease (FMD) prevalence in central Myanmar: comparison of village headman and farmer disease reports with serological findings

---

6.1 Abstract

The impacts of foot-and-mouth disease (FMD) on food security in developing countries are difficult to quantify due to the scarcity of accurate data on the prevalence and incidence of affected villages. This is partly due to resource constraints as well as the logistical challenges of conducting regular diagnostic testing in remote locations. In this study, we used descriptive analysis and latent class analysis (LCA) models to analyse data collected during a field survey of 160 villages in central Myanmar in the Mandalay and Sagaing Regions over the 2012 to 2016 time period. We evaluated the performance of verbal reports made by village householders and headmen against serological data to retrospectively determine the FMD-infection status of our study area and to identify factors contributing to underreporting. Blood samples were collected from approximately 30 cattle per village in both the 6 to 18 month age range and over 18 month age range to distinguish between recent and historic exposure. Village householders were asked to identify pictures of FMD affected cattle amongst pictures of cattle affected with other common endemic diseases to assess the accuracy of their verbal reporting. The serological results confirmed that FMD is endemic in central Myanmar with village level prevalence estimated at 56% for animals 6-18 months of age and 80% when all age groups were considered together. Most village householders were familiar with the clinical signs of FMD affected cattle (72%). Based on the results from the LCA models, the village headman had a sensitivity of 77% and specificity of 75% for identifying FMD outbreaks in their village, whereas individual householders had a higher sensitivity and lower specificity of 80% and 56%, respectively. The level of disagreement between the different sources was correlated with the

total number of cattle in the village and may potentially be worse in villages where endemic FMD may have led to a high level of natural immunity in cattle and subsequent masking of clinical signs. However, other regional effects such as the intensity of FMD extension efforts cannot be ruled out. Overall, the results suggest that verbal reports of FMD outbreaks from village headmen may be a useful tool to integrate into active FMD surveillance programmes in developing countries.

6.2 Introduction

Food security in subsistence farming communities is adversely affected by infectious diseases of production animals and in South East Asia, foot-and-mouth-disease (FMD) has been identified as a transboundary disease of particular importance. The losses attributed to FMD in South East Asia include lost income because draft animals are unable to work and lost production at the household level; both of which negatively affect food security and household income (Nampanya et al., 2016, 2015; Perry et al., 1999; Young et al., 2013). In 2010, it was estimated that the benefit to cost ratio of eradication of FMD from South East Asia was 3:1 (Mcleod, 2010). For these reasons, control of FMD has been identified as a priority for the livestock production sector in Myanmar. Myanmar is taking part in the Food and Agriculture Organisation of the United Nations (FAO) and World Organisation for Animal Health (OIE) Progressive Control Programme for FMD (PCP): a programme which aims to provide epidemiological support and benchmarking to those countries that wish to progress toward FMD freedom (Jamal and Belsham, 2013).

One of the requirements for participation in PCP is that the country seeking to progress toward disease freedom demonstrate a good understanding of disease epidemiology. This includes estimating the prevalence of FMD affected households and livestock in regions where the disease is known to be endemic and estimating the probability of disease freedom in regions where the disease is believed to be absent. In general, there are three main methods used to derive these estimates: (i) obtaining estimates from the literature and extrapolating these to local conditions, (ii) eliciting estimates from expert opinion, or (iii) collecting new data as a part of active or passive surveillance activities. Although the latter method should theoretically provide the most accurate estimates, resource constraints often make it impossible to census test every epidemiological unit and so various sampling strategies are employed to control for bias with the aim of producing the most internally and externally valid estimates possible.

FMD can broadly be diagnosed in three ways: PCR (or Ag ELISA) to detect the presence of the virus, Ab ELISA to detect a susceptible host’s immune response at some time in the past to the virus, and the observation of clinical signs (Alexandersen et al., 2003b; Longjam et al., 2011). Diagnosis by PCR and Ab ELISA requires samples either submitted to laboratories or analysed using penside tests (Longjam et al., 2011; Reid et al., 2001). As we are using the ELISA test to classify the disease status at the village level, we can expect the sensitivity of ELISA to be higher at the individual animal level than at the village level because we used a random sample of animals from each village rather than a census. If FMD prevalence in the village is low, there is a higher likelihood that the sample will fail to include a truly positive animals and therefore the village could be assigned a false negative status despite having positive animals. For developing countries, resource to perform disease testing on large scale and the logistical difficulty of obtaining reagents, and getting samples to laboratories can prove to be prohibitive. In these limited conditions, verbal disease reports may also be used as a source of surveillance information and may provide additional insights about the extent of clinical disease that cannot be gained from serology alone (Goutard et al., 2015; Muellner et al., 2016; Robertson et al., 2010; Sawford, 2011). However, it is unknown how well verbal reports represent the true disease status of the population given issues such as who the disease report is made by, how well they are able to correctly classify FMD, what the drivers for reporting are, and where the reporter obtains their information (Bellet et al., 2012a; Kenton L Morgan et al., 2014; Vergne et al., 2012). Disease reports made by human observers are also by no means perfect indicators of disease presence or absence – they vary based on the disease under investigation, the individual surveyed (e.g. knowledge and experience etc.) and the production system under examination. While FMD exhibits dramatic and recognizable clinical signs in the acute phase, the expression of clinical signs in populations where large numbers of carriers are present and endemic strains are circulating are less well described (Bertram et al., 2018b). This is important when considering how verbal reports compare with serological evidence of infection.

As none of the above diagnostic test methods are perfect, statistical approaches such as latent class analysis (LCA) can provide a valuable tool for estimating the relative performance of different tests and for estimating the true prevalence. LCA is based on the concept that observed results of different imperfect tests for the same condition are influenced by a latent common variable, the true status, which cannot be directly measured. In basic LCA models, the observed variables are assumed to be conditionally independent – this means that if a particular diseased individual has a positive result in one test, this does not influence the probability that the other test is positive (and similarly for non-diseased individuals). In a group of hosts where the true infection state is not known but for whom results from several diagnostic tests are available, LCA can be used to model the probability of each combination of test results on the latent class and thereby provide an estimate of sensitivity and specificity for each of the diagnostic tests evaluated (Hui and Walter, 1980; Rindskopf and Rindskopf, 1986). LCA has been widely used in veterinary epidemiology to compare test results of various diagnostic tests as well as comparisons of disease reports with serological testing in endemic disease situations (Bellet et al., 2012a; Bronsvoort et al., 2006; Kenton L Morgan et al., 2014; Stringer et al., 2013)

As part of the South East Asia and China Foot and Mouth Disease Project (SEACFMD) supported by OIE and the New Zealand Ministry for Foreign Affairs and Trade (MFAT) and the New Zealand Ministry for Primary Industries (MPI), researchers collected serological data from livestock in 160 villages across central Myanmar and conducted interviews with village headmen at the time of sample collection to get information on disease occurrence at the household level. This dataset therefore offers a unique opportunity to compare two types of village level disease reporting against serological results to determine how each can contribute to disease surveillance efforts. As neither of the village level reports (headman or householder) nor serology can be regarded as a "gold standard" for determining the presence of FMD, this study uses Bayesian LCA to estimate the performance of each of the three methods for correctly classifying the village FMD status. The implications of these findings for designing surveillance programmes in FMD endemic countries are discussed.

6.3 Materials and Methods

6.3.1 Survey design, respondents and serology:

As part of a SEACFMD project in central Myanmar, a cross sectional survey of villages was conducted between 1 October and 31 December 2016. The primary goal of the survey was to establish the prevalence of FMD at the animal level at the start of the risk based FMD vaccination programme in 2016. Sampling was conducted using a two-stage design first selecting 133 villages to visit and then 30 animals per village for a total sample of 4,000 cattle. Thirty animals per village was the minimum sample size required to detect a statistically significant difference and clinically meaningful reduction in FMD prevalence at the village level if these truly existed when follow-up sampling was conducted at the end of the programme. We used the following assumptions to define statistical and clinical significance: the estimated prevalence at the animal level at the beginning of the programme was 10% and the survey was designed to give a 90% chance of detecting a 50% reduction (prevalence decreased to 5% at the follow up survey) with 95% confidence (Dohoo et al., 2009). Due to extra resources made available to the project, a total of 160 villages were included by the end of the project.

Villages were selected from the Myanmar Information Management Unit (MIMU) register based on the perceived risk of FMD, logistical constraints identified by local experts, and the willingness of the village to participate in the programme. The project area overlaps 2 regions, Sagaing in the north and Mandalay in the south. Villages were visited once each by the field teams in the study period in 2016 during the months of October, November, or December. During the visit, the teams collected blood samples from 30 animals from 6 households in each village. The households within each villages were chosen by systematic random sampling. The animals within each household were selected at random from the two different age groups: 6 to 18 months of age and animals older than 18 months of age. The purpose was to allow comparison in seroprevalence between age classes to make inferences about both recent and historical infections with FMD.

Blood was collected by jugular venipuncture in the field by regional veterinary staff. The samples were placed in the shade at a 45 degree angle and allowed to settle in disposable syringes for an hour until the serum and red cells had separated. Serum was then decanted into 1.8ml plastic tubes (Cryovial by Sigma Aldritch) and stored in the regional departmental office refrigerators before being driven to the national laboratory in Naypyidaw. The time between blood collection and delivery to the national laboratory varied between 1 and 3 days depending on distance from the regional office to the national laboratory. Serological testing was performed using a commercial non-structural protein (NSP) ELISA test kit (PrioCheck FMDV by Prionics) using a cutpoint of 50% inhibition as specified by the manufacturer (Sørensen et al., 1998). The PrioCheck FMD ELISA has published specificity in cattle of 95 to 99% depending on the vaccination status of the cattle in question (Fukai et al., 2013; Sørensen et al., 1998). The sensitivity of this test in experimentally infected animals nears 100% (Brocchi et al., 2006). Estimates of the duration of detectable antibodies from natural infection with FMD are between 32 weeks and 5.5 years in the literature (Doel, 2016).

Two sets of interviews took place at each village during the sampling visit. The first interview administered survey questions to the village headman (VHR) and took place in a group setting. This collected information on the animal populations in the villages based on the knowledge of the headman and the assembled group. The second more detailed survey was performed with each householder (HHR) who owned the animals that had blood samples collected and tested for evidence of antibodies to FMD. To assess the ability of HHRs to accurately diagnose clinical FMD, each HHR was shown a series of photographs of animals displaying clinical signs of common infectious diseases including FMD, Black quarter (caused by *Clostridium chauvoei*), Anthrax (caused by *Bacillus anthracis*) and Haemorrhagic septicaemia (caused by *Pasteurella multocida*) and asked to identify the FMD infected animal.

The VHRs and HHRs were both asked if they had noticed animals with clinical signs of FMD during each of the calendar years between 2012 and 2016. Data from 2015 and 2016 represented the period in which animals that were between 6 and 18 months old at the time of sampling would have potentially been exposed to FMD. A comparative seroprevalence estimate was calculated by including serology from older animals and paring these with observations from 2012 to 2014. In the latent class models we used all animal serology to generate the prevalence estimates and the young animal results (6-18months old) paired with the 2015 and 2016 survey results to generate the incidence estimates.

6.3.2 Data analysis

Descriptive statistics

Animal populations in the surveyed villages were described (a. participant demographics and location). Serology results for tested cattle were aggregated to the village level to estimate circulating antibodies in different age groups. The recognition of FMD by householders (b. HH FMD recognition) was summarized by village, township and region and the variance within and between villages plotted.

Estimating village-level FMD incidence and prevalence based on serology

To approximate disease incidence, we used the serology results in animals six to 18 months of age and compared these with disease reports by village headmen and householders in 2015 and 2016. For this analysis we regarded serology as the gold standard. Disease reports from 2012 to 2014 and serology results from older animals were described. Finally, all years and all serology results were used to estimate overall disease prevalence. The results in the 6-18 month age group were compared using t-tests with results in the older animals and this was compared between Sagaing and Mandalay Regions which comprise the north and south of the project area respectively. A linear regression model was run with the prevalence in young animals at the village level as the outcome variable and region and prevalence in old animals as explanatory variables. True and apparent prevalence estimates were calculated in an online calculator (Sergeant, 2019) using standard statistical methods to estimate apparent and true prevalence and their confidence limits (DasGupta et al., 2001; Reiczigel et al., 2010; Rogan and Gladen, 1978).

Comparing serology and verbal reports for classifying village FMD status

The multi-level structure of our data made it possible to examine how householder, village headman and ELISA test results performed at identifying FMD in an endemic context at the village level. Simple descriptive tables and mapping of results were used to describe the relationships between the ELISA testing at the village level and what was reported by the householders and the village headmen (ESRI, 2018). For the purpose of these plots, if more than 10% of animals in a village tested disease positive, the village was regarded as positive. Agreement between VHR and HHR, HHR and ELISA, and VHR and ELISA was calculated by simple percentage agreement as well as by Cohen's kappa statistic (Cohen, 1960). Cohen's kappa (K) measures agreement between two categorical outcomes relative to an expectation under complete randomness: if both datasets are in complete agreement then K=1 and if there is no agreement between the two categories other than what would be expected by chance then K=0. By convention, interpretation of the statistic between 0 and 1 is that values between 0 and 0.2 indicate slight agreement, 0.21–0.40 is fair agreement, 0.41–0.60 is moderate agreement, 0.61–0.80 is substantial agreement and 0.81–1 is almost perfect agreement (Landis and Koch, 1977). Classification and regression trees with the Gini information index (Therneau et al., 2015) were used to identify a level of seropositivity at which clinical signs were reported.

Factors influencing the agreement between villagers and village headmen

Verbal reports from the village headmen were compared with village householder reports. Village householders were asked about clinical signs in their own animals and village headmen about clinical signs in their villages. If any of the householder reports were positive, this was identified as a positive identification at the village level for the HHR data. Where village headman and householder reports were different, the village was assigned a 1 and if the same, a 0 creating a binary outcome. This binary outcome was regressed against available exposure variables in a logistic regression model and examined in a random forest model using a minimum bucket size of 10 and Gini splitting algorithm (Liaw and Wiener, 2002; R Core Team, 2014; Therneau et al., 2015; Vinet and Zhedanov, 2010). Exposure variables that were examined included the numbers of draft and dairy cattle, the number of households with susceptible livestock in the village and the ability of the householders to identify FMD as ascertained by the survey questions (measured as a percentage at the village level).

Latent class models

As neither the serology nor verbal reports are entirely accurate methods of classifying village FMD status, we chose to use a series of latent class models to estimate performance of the different tests and explore prevalence of disease in our study area. In the first latent class model (Model 1) all three tests have binary outcomes and allowance is made for correlation between VHR and HHR. This model is directly adapted from Branscum et al 2005 (Branscum et al., 2005). VHR is a binary outcome as each village has only one headman and the questions were based on the recollection of the occurrence of clinical signs of FMD in each survey year. HHR were manipulated to create a binary outcome from 6 household level reports for each village: a HHR in a particular year was regarded as positive if any one of the 6 householder surveys reported seeing clinical signs of FMD in that year. Interpretation of the NSP ELISA test at the village level was required to create a binary outcome for NSP ELISA. If more than 10% of animals tested returned a result of greater than 50% inhibition then the village was classified as positive in Model 1.

Models 2 and 3 were modified so that the ELISA result was modelled as a continuous variable, while VHR and HHR were considered to be binary as in Model 1. In Model 2, the corrected optical density (OD) NSP ELISA result for each village was the average corrected OD for all animals in that village i.e. a continuous variable. Model 3 represented the ELISA results as the proportion of animals with NSP results greater than 50% inhibition in each village. This model allowed the number of ELISA positive animals from a village to be a binomial distribution with the probability (P) that a tested animal was ELISA positive modelled as a logit-normal distribution with different parameters for infected and non-infected villages.

The two time periods and the three latent class models resulted in 6 estimates of sensitivity, specificity and disease frequency. Receiver Operator Compliance (ROC) Curves were generated for the continuous ELISA outcomes. These curves are interpreted according to convention with the curve which is closest to the top left corner of the graph representing the most useful test

Sensitivity and specificity of each of the three ‘tests’ (householder reports, headman reports and ELISA at the village level) as well as prevalence estimates at the village level were then examined using the outputs of the LCA models. Models were built in OpenBugs and outputs evaluated in R version 2.15 with a burn-in period of 20000 runs for each of the models and retaining 80000 iterations for estimates of the values and credible intervals (Branscum et al., 2005; Lee and Wagenmakers, 2010; Lunn et al., 2000; Ntzoufras, 2011; R Core Team, 2014). OpenBugs code is provided as an appendix.

6.4 Results:

Descriptive statistics

a. Participant demographics and location

A total of 160 villages were surveyed from 24 townships in a project area situated in two central districts (Sagaing and Mandalay) of Myanmar (Figure 11). The villages surveyed were distributed relatively uniformly across the study region. Within the 160 villages, a total of 4,075 animals were sampled. The surveyed villages were approximately representative across the project area when the underlying population of villages was taken into account. The size of the villages surveyed varied both in relation to how many susceptible animals were present and the species distribution present (Table 20).

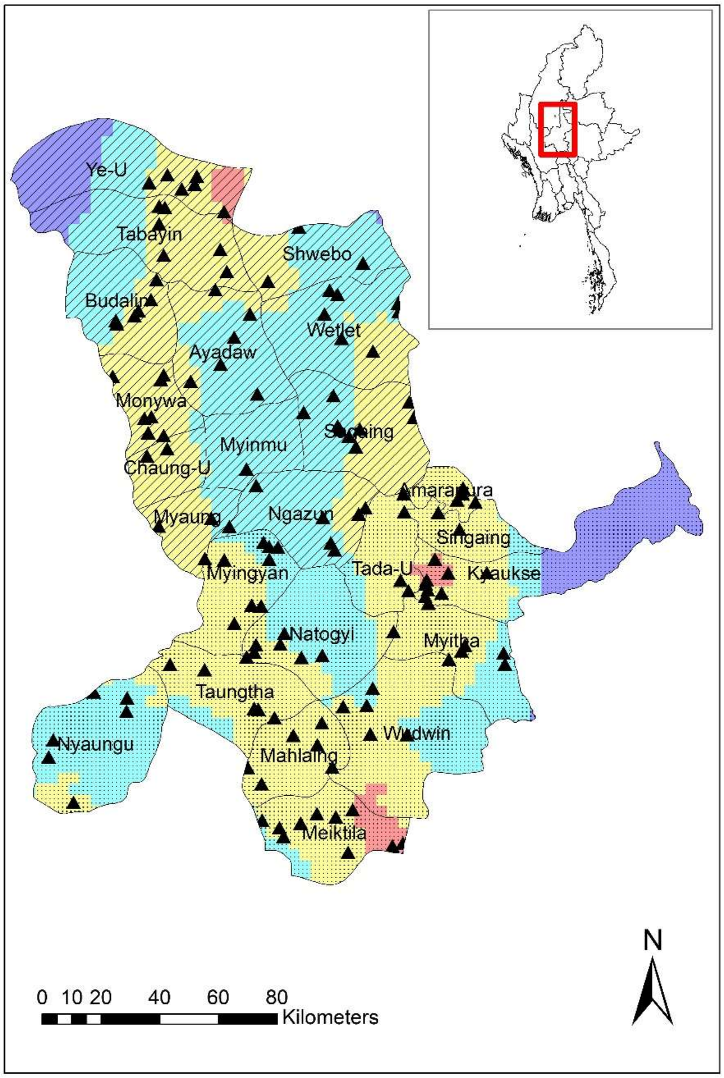

Figure 11: *Distribution of project regions, townships and villages. Twenty four project townships are part of our project area in central Myanmar (inset top right). Striped townships are in Sagaing Region and speckled townships in Mandalay Region. Both the Mandalay and Sagaing regions are in central Myanmar. Villages surveyed are represented by black triangles and the background colour gradient represents the density of villages with blue representing the lowest densities (less the 1 village per 10 square km) and red the highest (of more than 3 per square km). Yellow areas have a density of between 2 and 3 villages per square km and turquoise areas between 1 and 2 villages per square km). Spatial representation adjusted by underlying population can be estimated from examination of this figure.*

Table 20: *Distribution of the counts of households and FMD susceptible animals present in 160 villages included in a serological survey performed in the Sagaing and Mandalay regions of central Myanmar.*

	Minimum	**25%**	**Median**	**75%**	**Maximum**	**Mean**
number of households	7	104	167	275	5873	160
households with livestock	0	55	100	182	3725	170
count of draft cattle	0	145	254	500	3500	442
count of dairy cattle	0	0	0	0	400	11
count of buffalo	0	0	0	0	327	4
count of goats	0	0	60	200	2450	176
count of sheep	0	0	0	34	716	46
count of pigs	0	10	28	60	516	52

b. HH FMD recognition

When asked to identify the clinical signs of FMD, 72% of householders were able to correctly identify photographs of FMD from a line-up of four endemic diseases. The 4 pictures were shown in sequence and each interviewee was asked of each one if it was FMD. The ability of the HHs to identify FMD was not statistically different in Mandalay Region when compared with Sagaing Region (t = 0.02, p value = 0.98). The variation at the township level was notable (Figure 12) and was included as an exposure variable in the regression analysis exploring agreement between householders and village headmen. Of the remaining 28% of interviewees, 2% thought that FMD was not included in the pictures shown and 18% said they did not know. The remainder identified other endemic diseases as FMD (2.1% black quarter, 2.1% anthrax and 3.5% haemhorragic septicaemia). Interpretation of the boxplots is at the township level, so in the instance of Ngazun township in Figure 12, most of the villages had a 100% recognition of FMD clinical signs, and one village had 83% recognition – shown as the outlier (pink dot). Given that only 6 householders were surveyed per village, this would have been the result of one householder not being able to correctly identify the clinical FMD picture.

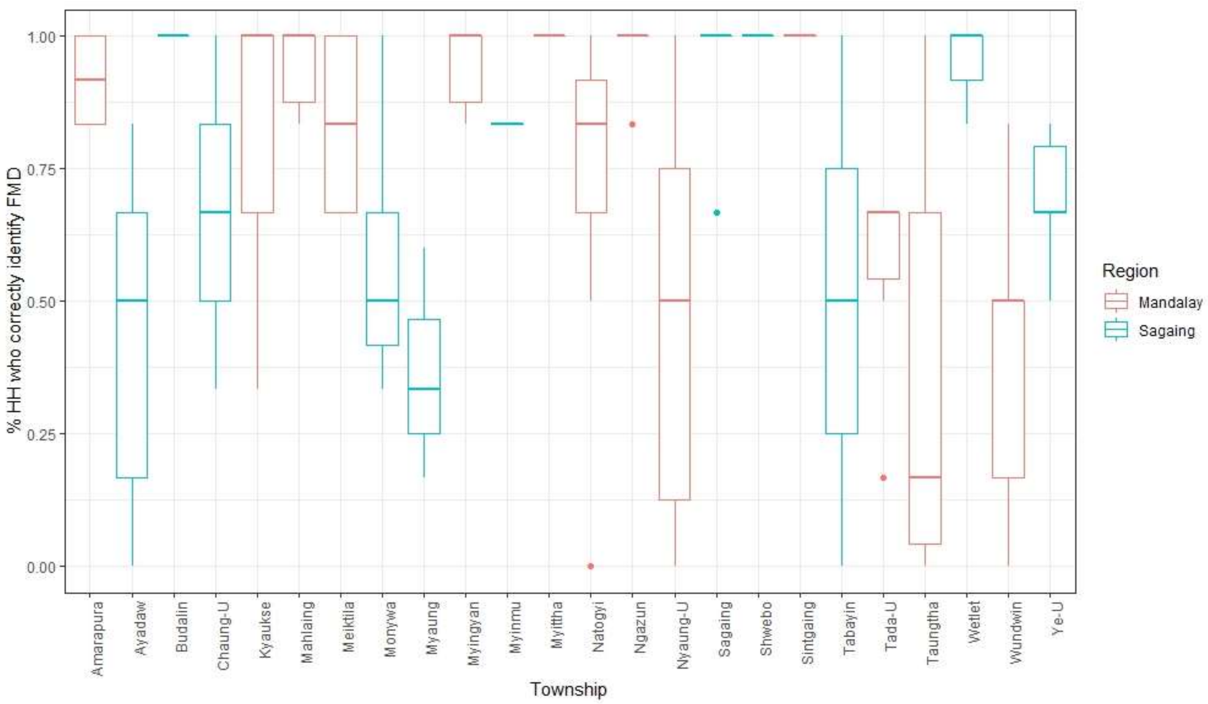

Figure 12: *Variation in the percentage of village householders able to correctly identify a picture shown to them depicting the clinical signs of FMD in a lineup of 4 endemic diseases. Householder responses have been aggregated to the township level (x axis) and Regions are indicated by colour (Mandalay in pink and Sagaing in blue).*

Estimating village-level FMD incidence and prevalence based on serology

Of the 160 villages sampled, 89 had serological evidence of exposure (defined as more than 10% seropositive animals) in the 6-18 month old animals (interpreted as incidence) and 126 had evidence of exposure in animals of any age. The geographic distribution by township of these villages is shown in Figure 13. This yields an apparent prevalence (AP) of 56% (CI 48-63%) and a calculated true prevalence (TP) of 61% (CI 51-72%) in young animals and an AP of 79% (CI 72-84) and TP 87% (CI 78 – 94%) in the whole population.

Figure 13: *Geographic distribution of total prevalence and prevalence in young stock at the township level. Village level seroprevalence in the sampled population aggregated to the township level (panel A) compared with the same measure for animals 6-18 months of age (panel B). Light grey townships have less than 15% seropositivity, mid grey 15% - 25% and dark grey greater than 25% seropositivity.*

When these results were stratified by region and age group, a t-test identified a statistically significant difference in the older animals between Mandalay and Sagaing (t =2.24, p =0.02). Prevalence rates in the older age group were 34% in Mandalay and 27% in Sagaing. Prevalence rates were the same in the 6-18 month age group. Figure 14 shows this relationship. A linear regression model with prevalence in young animals as the outcome and prevalence in old animals as an explanatory variable with region as a fixed effect had an Rsquared value of 0.2.

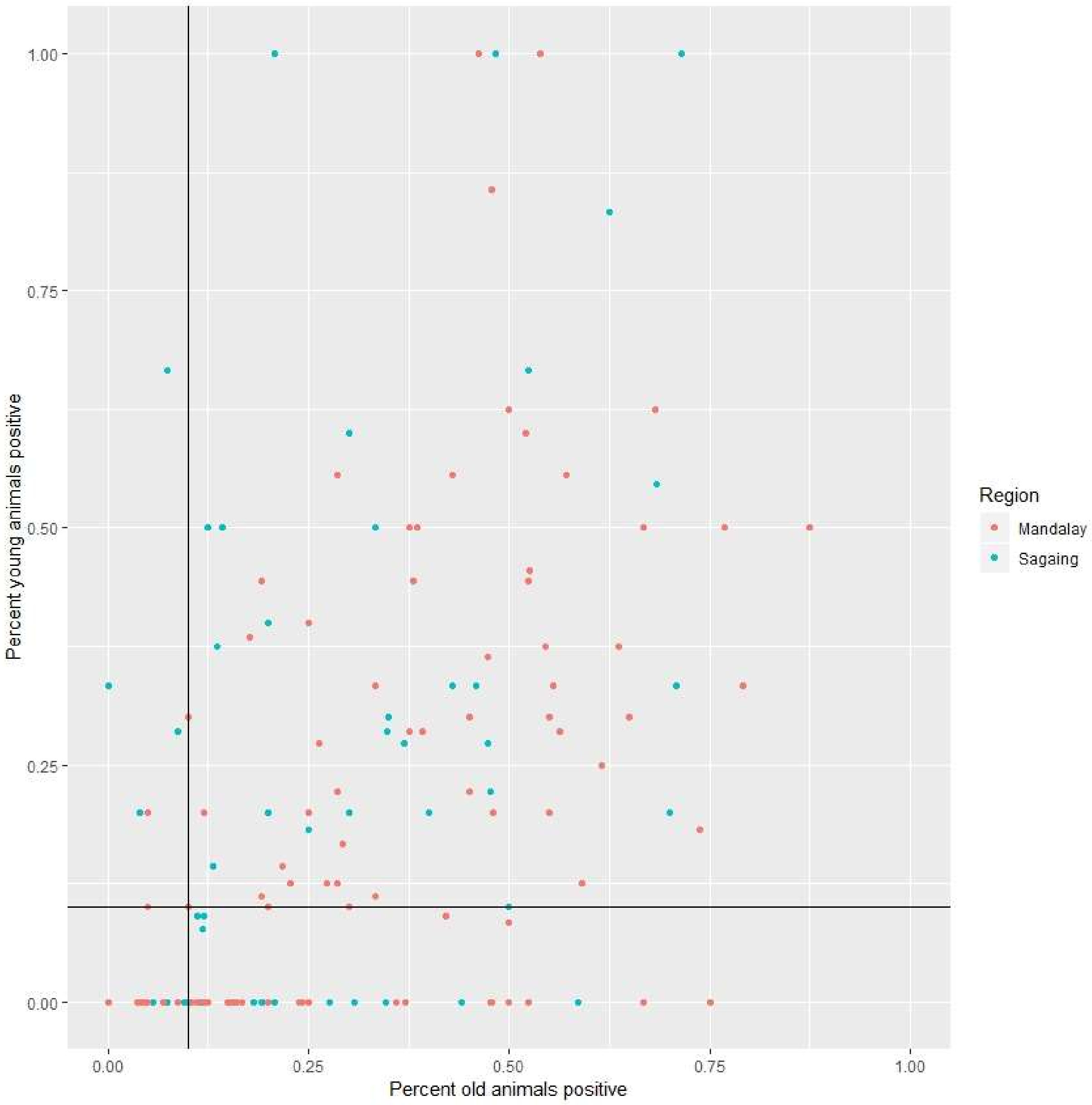

Figure 14: *Scatter plot of the relationship between the percentages of 6-18 month old animals with positive serological results and the percentage older animals at the village level in 2 regions of central Myanmar. The reference lines indicate the cut points for considering a village to be positive (greater than 10% of animals positive on serology). Villages that are negative and remained negative fall in the bottom left quadrant. Villages negative in older animals and positive in young stock are found in the top left quadrant and villages with negative young stock but positive older animals in the bottom right quadrant. Those villages with persistently infected old and young stock appear in the top right quadrant.*

Comparing serology and verbal reports for classifying village FMD status

Simple tabular measures of agreement (Table 21) show that householders and village headmen were in agreement 77% of the time in their estimates over the last 2 years (68 + 55 / 160) and 82% of the time over the whole study period (20 + 111 / 160). Agreement for the period between 2012 and 2014 was 68%. Adjusting for the effect of chance using Cohen's Kappa indicates only slight agreement between serology and survey results from both householders and village headmen (Table 22).

Table 21: *comparison of how village headman and householder interviews reported outbreaks in 2015 and 2016 (2A) and for 2012 to 2014 (2B) and for all records (2C) displayed as raw counts.*

2A 2015 and 2016	**Householder negative report**	**Householder positive report**	**Total**
Village head negative report	68	27	95
Village head positive report	10	55	65
Total	78	82	160

2B 2012 to 2014	**Householder negative report**	**Householder positive report**	**Total**
Village head negative report	43	42	85
Village head positive report	9	66	75
Total	52	108	160

2C 2012 to 2016	**Householder negative report**	**Householder positive report**	**Total**
Village head negative report	20	23	43
Village head positive report	6	111	117
Total	26	134	160

Table 22: *Comparison of the agreement between village headman reports, householder reports and ELISAs both as simple agreement and using Cohen's Kappa in three different time periods. K=0 is no agreement, 0–0.2 is slight agreement, 0.21–0.40 is fair agreement, 0.41–0.60 is moderate agreement, 0.61–0.80 is substantial agreement and 0.81–1 is almost perfect agreement*

	VHR and HHR		**VHR and ELISA**		**HHR and ELISA**	
	% agreement	**K (95% CI)**	**% agreement**	**K (95% CI)**	**% agreement**	**K (95% CI)**
animals 6-18 months and 2015 2016 reports	77	0.54 (0.41 - 0.67)	59	0.19 (0.04 - 0.34)	55	0.08 (0 - 0.23)
older animals and 2012-2014 reports	68	0.37 (0.23 - 0.51)	48	0	47	0.02 (0-0.21)
All animals and all reports	82	0.47 (0.29 - 0.64)	75	0.30 (0.11 - 0.49)	74	0.14 (0 - 0.36)

Based on classification and regression tree models, 21% of animals had to be seropositive on ELISA testing for village headman or householders to respond that FMD was present in the time frame under discussion.

Factors influencing the agreement between villagers and village headmen

Logistic regression models suggested a trend that for every 1000 bovines present in a village, the odds of the village headman and the village householder disagreeing increased 2 fold. The p value for this finding was 0.1. The percentage of households with livestock in the village and the ability of the householders to correctly identify clinical signs of FMD were retained in the model but did not have significant statistical effects on the outcome in the logistic regression models run. Random forest models identified that the percentage of households in the village with livestock made the largest impact on the node purity and mean squared error of the model (11.5 and 68%) and that the count of cattle (in increments of 1000) had the second largest impact at 3.2 and 5.4%. The random forest models again confirmed that ability to correctly identify FMD did not contribute to disagreement between the HH and VH.

Latent class models

The model that used all the individual ELISA testing results and modelled the outcome as a percentage positive at the village level (Model 3) provided the best performing model (Figure 15). All three models estimated a prevalence at the village level of between 70 and 80 % in all cattle and around 60% in the 6-18 month old age group (Table 23). The sensitivity of ELISA using our sampling strategy was estimated at 83% for the 6-18 month old age group by the binary model and close to 100% for the 10% cut point we estimated for our calculations (based on the ROC curve in Figure 15). In our models, neither village headmen nor householders are a sensitive source of disease reports when only the binary outcome model is considered. When the model that makes the fullest use of the data is considered (prevalence model 3), householders are the most sensitive source of disease reports (VH 77%, HH 80%), but less specific than village headmen (HH 56%, VH 75%).

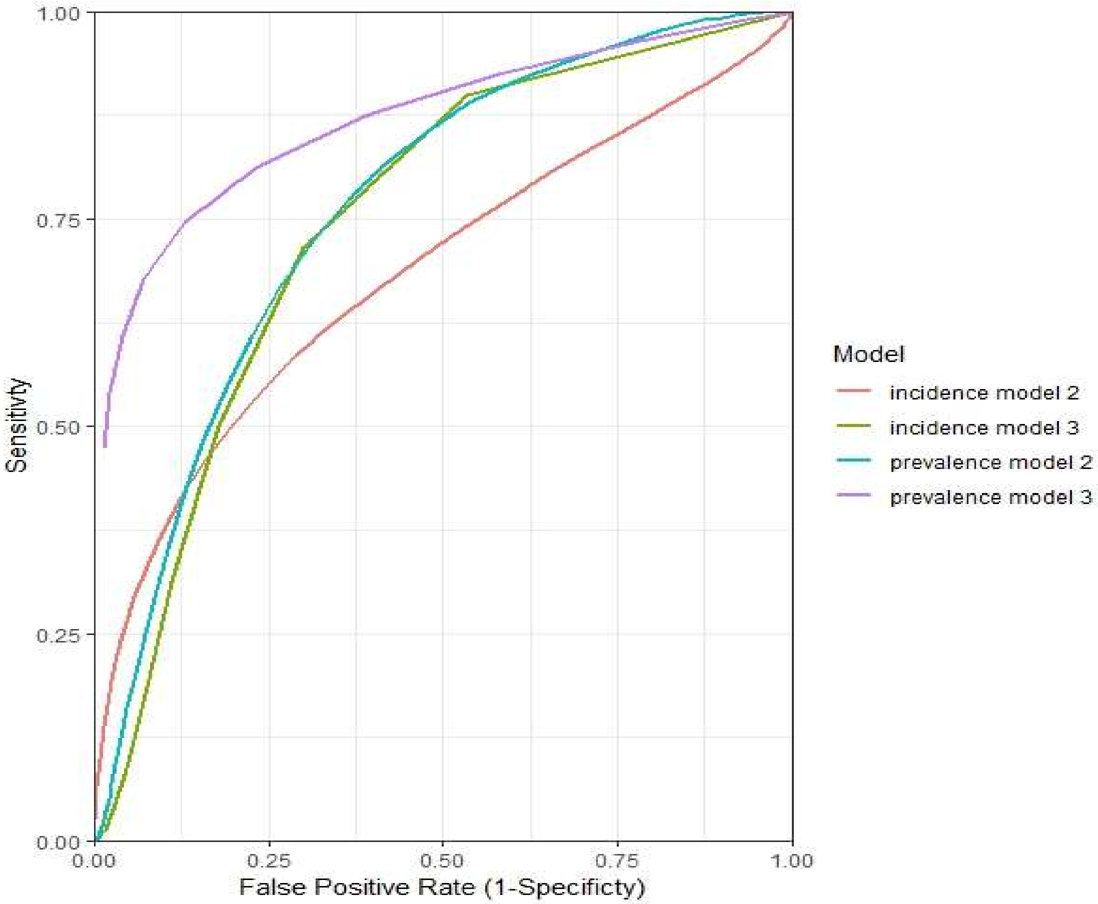

Figure 15: *ROC curves for the 2 latent class models which represent the ELISA testing data (including sampling uncertainty) as continuous variables. The purple line includes serology results from all animals tested and represents the continuous ELISA outcome as the proportion of corrected OD values over 50% in the sampled population of each village. The green line uses the same model but incorporates less samples as it includes only those animals between 6 and 18 months of age (the incidence group). The blue line modelled the ELISA results as the average of the corrected OD value at the village level and included all animals sampled. The pink line uses the average corrected OD value at the village level but included only those animals aged between 6 and 18 months old.*

Table 23: *model results of 3 latent class models representing 3 ways to identify FMD at the village level (ELISA testing of a sample of animals, survey questions to village headmen and survey questions to householders) and the estimates of incidence, prevalence, sensitivity and specificity derived from each model*

	Model 1: 3 tests modelled with binary cut points (credible intervals)		**Model 2: ELISA has continuous outcome based on average corrected OD (credible intervals)**		**Model 3: ELISA has continuous outcome based on average prevalence of corrected ODs over 50% (credible intervals)**	
	6-18 month olds	**all cattle**	**6-18 month olds**	**all cattle**	**6-18 month olds**	**all cattle**
prevalence estimate	0.63 (0.49 - 0.77)	0.84 (0.76 - 0.93)	0.58 (0.41 - 0.74)	0.87 (0.78 - 0.96)	0.57 (0.4 - 0.74)	0.81 (0.71 - 0.91)
Sensitivity of ELISA incorporating sampling uncertainty	0.83 (0.69 - 0.97)	0.89 (0.81 - 0.97)	see Figure 15	see Figure 15	see Figure 15	see Figure 15
Sensitivity of Householder reports	0.56 (0.48 - 0.64)	0.71 (0.65 - 0.78)	0.64 (0.54 - 0.74)	0.79 (0.73 - 0.85)	0.64 (0.55 - 0.75)	0.80 (0.74 - 0.85)
Sensitivity of Village Headman reports	0.56 (0.48 - 0.64)	0.81 (0.74 - 0.86)	0.63 (0.53 - 0.74)	0.75 (0.68 - 0.81)	0.64 (0.54 - 0.75)	0.77 (0.71 - 0.84)
Specificity of ELISA incorporating sampling uncertainty	0.89 (0.84 - 0.94)	0.89 (0.84 - 0.94)	see Figure 15	see Figure 15	see Figure 15	see Figure 15
Specificity of Householder reports	0.78 (0.66 - 0.92)	0.75 (0.56 - 0.93)	0.71 (0.55 - 0.86)	0.64 (0.40 - 0.89)	0.71 (0.56 - 0.86)	0.56 (0.33 - 0.77)
Specificity of Village Headman reports	0.61 (0.47 - 0.76)	0.53 (0.32 - 0.73)	0.83 (0.72 - 0.94)	0.76 (0.60 - 0.91)	0.84 (0.74 - 0.94)	0.75 (0.59 - 0.89)

6.5 Discussion:

In this study, we evaluated the use of three different imperfect methods for estimating FMD prevalence at the village level in central Myanmar. Specifically, we focused on comparing the performance of village headman (VHR) and householder reports (HHR) against serological data at the village level to identify the presence of circulating FMD during specified time periods. It was hoped that by quantifying the diagnostic test performance of verbal reports, this would provide authorities with an additional tool to estimate the prevalence of FMD without the expense or logistical challenges of conducting serological testing in each village. Using the available data, we have identified that village headmen are a more specific source of FMD disease reports, but appear slightly less sensitive than householders (confidence intervals overlap in Table 23) and that the level of disagreement between the two sources can be correlated with the size of the cattle population in the village. These trends are borne out both by the descriptive Kappa results in Table 22 and the LCA in Table 23. We do not expect that our findings around diagnostic performance can be directly extrapolated to other countries. The low levels of agreement indicated by the Kappa results are cause for further investigation and we regard this study as a first step to quantifying the usefulness of using village level reports to establish FMD prevalence in Myanmar and establishing a methodological framework for other countries to use in evaluating the use of village reporting for FMD surveillance.

Our results showed that the level of agreement between the village reports and serological data varied across the different townships in Central Myanmar. One possible explanation is that officials in some regions are more proactive with their FMD communication campaigns leading to greater awareness of FMD amongst livestock owners, but we also cannot rule out the potential role of interviewer bias during the face-to-face interviews with village headman and householders. The interviews within each township were carried out by staff from the township and no single interviewer performed interviews in more than one township. It is also possible that clinical expression of FMD in livestock varies from township to township particularly if there is a high underlying prevalence of other endemic cattle diseases that could mimic or mask the presence of FMD (Bellet et al., 2012b). Most householders were comfortable with identifying FMD from pictures making it less likely that they would have missed obvious clinical cases in their animals. However it is unclear what priority was placed by the householders on the reporting of the disease to officials. Findings in other FMD endemic countries indicate that diseases that have high mortality rates are prioritized higher by farmers and that this leads to under-reporting of FMD (P C Cleland et al., 1995;

Perry et al., 2002, 1999). Further field studies collecting a wider range of pathogen data and using an independent team of interviewers might help to identify the reasons for differences in reporting performance between townships, which is important for being able to adjust future prevalence estimates based on the underlying township demographics. The authors accept that an ideal study design would have seen the study performed in an area where vaccination was never done or excluded previously vaccinated animals from the analysis. Unfortunately this was not possible in our context.

It was also interesting to note that the overall agreement of the village headman reports with the serology was good when all five years of historical observations were included, but poorer when only recent observations from 2015 and 2016 were included. In villages where the disease has reached endemic stability or where there have been aggressive vaccination campaigns, we would expect the livestock populations to have a high degree of immunity against FMD, which may suppress the expression of clinical signs (Bertram et al., 2018a). Another possibility is that FMD may have died out in particular villages if the population of susceptible hosts was depleted due to the combination of rapid viral transmission and low livestock population turnover through births or migrations. The former scenario would be expected to reduce the sensitivity of verbal reporting while the latter scenario would be expected to reduce the specificity. A further complicating factor is that as antibodies can persist for many years, using serological status to infer current infection status is not possible. Verbal reporting of FMD is therefore more likely to be unreliable in countries with a high endemic prevalence of FMD or where the epidemiological situation is not well quantified. Our study did not conduct any molecular studies on the circulating viral strains to determine if there were regional differences in virulence and/or transmissibility. Such information could be valuable for future disease control efforts to better understand how the virus circulates within and between villages, townships, and regions.

Our analysis treated a single village as the epidemiological unit of interest for surveillance. Given that the agreement between village headman and household reports was worse in villages with more cattle, it is possible that village headman may not always be able to keep accurate track of the disease status of all households in the village over time. This is supported by the findings from our random forest model showing that number of cattle in the village and the number of homes in the village were significantly correlated with differences of opinion on disease occurrence, and we also found that agreement between the two groups on whether FMD was present in a given year became progressively worse the further back in time participants were asked to recall disease events. Our analysis allocated a binary outcome for agreement between householders and village headmen –

allocating a 1 to positive HH, positive VH and negative HH, negative VH cases and a 0 to positive HH, negative VH and negative HH, positive VH cases. An interesting future analysis could focus on further examining the differences between the cases in which village headmen and householders disagreed. This may lead to deeper understanding of any differences between positive HH, negative VH and negative HH, positive VH cases. Despite these differences, a village headman report is still expected to be more sensitive at the village level since they have contact with a greater number of households making it more likely that they will hear of at least one affected household if FMD is truly actively circulating through the village. Sensitivity at the household level is expected to be lower as only the household cattle and potentially a small number of neighbouring cattle are being observed. Due to the relative scarcity of villages with dairy cattle we were not able to draw conclusions on the role of dairy versus draft cattle as well as other variations in cattle management between villages, but this would be an interesting area of future study. Our study has broadly comparable results with a similar study performed in Cambodia where VHR was estimated to have a Se of 0.87 and Sp of 0.3 (Bellet et al., 2012b); although our estimate of sensitivity for the VHR group is slightly lower (0.77) and the specificity higher (0.56), credible intervals approach the estimates from the study in Cambodia. The relatively low specificity of verbal reporting in the region has been noted in the past (Madin, 2011).

Incorporating both the serological data and verbal reports into our statistical models, our most conservative estimate is that 57% of all surveyed villages truly experienced FMD disease events within the two years prior to sampling thereby confirming the endemic nature of FMD in central Myanmar (Bellet et al., 2012b). Our analysis has found some evidence of regional heterogeneity in disease prevalence. Serological prevalence in older animals in the Sagaing region was statistically different (27%) from the same group in the Mandalay region (34%) while seroprevalence in the 6-18 month old animals was the same in both regions (21-22%). Some selection bias is inherent in our study design as very remote villages, those not recorded in the central data repository or those that are unwilling to participate are underrepresented – this could falsely extrapolate higher levels of education and awareness of FMD to areas where this is not actually the case.

Although we have not performed a formal sensitivity analysis, we have presented 3 alternate latent class models. The agreement between the different methods is encouraging but does not provide proof of the external validity of our approach.

It is unsurprising that ELISA was predicted to have the best sensitivity and specificity out of the three information sources in the latent class models. However, some assumptions were necessary to be able to use latent class models to represent our situation and we present them here as per the Standards for the Reporting of Diagnostic accuracy studies that use Bayesian Latent Class Models

(Kostoulas et al., 2017) (STARD-BLCM). The first of these is that the householder and the village headman reports are conditionally independent of the ELISA results given the status of the village but that household and village headman reports are correlated (Branscum et al., 2005). The second is that the sensitivity and specificity of the 'tests' are not interpreted at the level of the individual animal but represent the sensitivity and specificity of the system under assessment. This includes design features such as sampling uncertainty which is implicit if every animal in every household is not sampled in each village and the upward aggregation of household reports to the village level. These issues affect the interpretation of our models. Furthermore, each village must be regarded as a unit in which homogenous mixing occurs and the 6 householders picked from each village must be accepted to be representative of the population of householders in the village. This is a reasonable assumption given communal grazing practices in central Myanmar (Cocks et al., 2009). A comparison of the three models described shows that modelling the proportion of animals which have NSP ELISA results (corrected OD) over 50% inhibition is a better use of the data than modelling the average corrected OD ELISA results at the village level.

Future surveillance in the project area should incorporate more participatory epidemiology and focus on combining the local knowledge of village headmen and householders. This should be augmented with investigation of what factors motivate and inhibit the reporting of FMD cases.

6.6 Acknowledgements

The authors acknowledge members of the OIE-LBVD FMD Project Team including Kyaw Thinn, Assistant Director, Myo Thiha, Research Officer, and Kyaw Htay, Research Officer, for their efforts in collecting data from the field to be able to complete this study. In addition, we greatly appreciate the direct contributions made by LBVD regional staff in Mandalay and Sagaing who directly participated in this study and openly shared available date and knowledge from their respective townships. We gratefully acknowledge the support provided by the LBVD of the Ministry of Agriculture, Livestock and Irrigation, Govt. of Myanmar. Further thanks are due to Flavie Goutard of CIRAD for sharing her experience with a similar undertaking in Vietnam.

CHAPTER 7

7 Integrating multiple data sources to evaluate the quality and completeness of national animal identification and traceability systems

7.1 Abstract:

In countries with national level animal movement databases, those that prepare for and respond to outbreaks of infectious disease in animal populations must make assumptions about data completeness. It may be impossible to assess the degree of missing data present in a system without intensive interviews and labour-intensive amalgamation of alternative data from multiple sources. The objective of this study was to make use of existing work done as a part of New Zealand's *Mycoplasma bovis* eradication programme to assess the level of data missing from New Zealand's national animal movement database. Two analyses were completed, the first examined movement data missing at the farm level for all case farms identified in a period of the programme and the second examined missing information at the animal level at a sample of abattoirs over consecutive years during particular months. Various descriptive statistics and simple models were used to explore factors associated with missingness of information. The results showed at the farm level that a significant number of dairy cattle movements are missing and that these are mostly related to the movements of calves, bulls and grazing stock. When animal level missing movements were examined, it appeared that compliance with movement recording improved over time and that time of year and region were both statistically significant factors that influenced missingness. Overall, these findings highlight the need for improved compliance with New Zealand's animal movement recording system and provide a useful starting point against which to measure compliance in the future.

7.2 Introduction:

Animal identification and traceability systems are an important part of national disease control programmes because they allow regulatory authorities to rapidly trace the movement history of infected animals during a disease outbreak to identify other animals and/or farms that may have been exposed to the disease. During the immediate response to a disease outbreak, the focus is typically on quantifying the extent of disease spread to estimate the magnitude of the control measures required as well as the potential economic and social impacts on the industry(Caporale et al., 2016a). Aside from this essential operational use, these databases also provide the opportunity to characterise the location and health status of animal populations as well as national, regional and enterprise type specific movement patterns and how these patterns influence the risk of disease transmission (Buhnerkempe et al., 2014; Gates et al., 2014; M. Carolyn Gates et al., 2013b; Gopal et al., 2006; Kiss IZ Kao RR, 2006; Schärrer et al., 2015). This type of information is often used in disease preparedness activities to explore how the competent authority might best respond to an outbreak of an exotic disease or design the most feasible control programmes for endemic diseases (M C Gates et al., 2013; M. C. Gates et al., 2013c; M. Carolyn Gates et al., 2013b; Gates and Woolhouse, 2015).

Recognising the importance of traceability systems for trade assurance and long-term sustainability of the livestock industries, New Zealand implemented the National Animal Identification and Tracing (NAIT) system for cattle and deer in 2012. At the time of writing June 2019), this system requires each bovine or cervid to be tagged with a radio frequency identity tag (RFID) at 6 months of age or when it is moved from the location of its birth (whichever happens first). This age limit was chosen since many extensively reared beef calves in New Zealand pastoral production systems are often not handled until weaning at approximately 6 months of age and it would be logistically challenging for farmers to yard cows with calves at foot at an earlier time for tagging. The current NAIT regulations also account for the common practice of farmers moving animals between grazing locations on farm by allowing a single NAIT number to apply to all land belonging to a single enterprise that is within a 10km radius of a main location. To increase the likelihood of farm to farm movements being captured in the database, both the sending and receiving parties are required to record the movement. In the NAIT database, movements that were only logged by one party are flagged as being ‘incomplete’ until the discrepancy has been resolved. The exception to this rule are movements to abattoirs, which are a one step process where the movement is logged as complete regardless of whether or not the farmer records the movement. The NAIT numbers associated with movements to abattoir are taken from the paper-based animal status declaration (ASD) document which is required by law to accompany the animals to slaughter. The ASD documentation does not record animal level identification (the RFID) and this is collected by a scanner as the animal arrives.

Worldwide, there are known limitations to national livestock traceability databases that affect their utility in supporting disease control activities. For example, there is often little information on data quality or the rate of under recording of movements since this requires more in-depth follow-up surveys to accurately estimate (Enright and Kao, 2016; Green and Kao, 2007). There are inevitable time delays between farmers recording animal births, deaths, and movements. Depending on the legislative requirements of the country, the point location of the farm recorded in the database may not accurately reflect the true spatial location of where the animals are grazed or housed. This results in difficulties when linking the traceability data to other national animal health databases. This becomes particularly inaccurate if there are inconsistencies in how each system records the identity of individual animals and farms (Gates et al., 2014; M. Carolyn Gates et al., 2013b; Jewell et al., 2015).

For New Zealand, the limitations of the existing NAIT data and the complexities of integrating from other national animal health databases with NAIT have been described(Jewell et al., 2015) and became readily apparent during the response efforts to the incursion of *Mycoplasma bovis* which was first reported in New Zealand in July 2017(Rawdon, 2018). Previous surveillance for the disease had yielded negative results, which suggests that *Mycoplasma bovis* was absent from New Zealand up until this time (McDonald et al., 2009). Due to the heavy reliance of New Zealand's cattle industry on moving large volumes of cattle over large distances both for trade and for grazing, by the time the first case was identified in the South Island of New Zealand, animals that had previously been in contact with infected stock had already been distributed widely across operations in both the North and South Islands.

As with other *Mycoplasma* species, the primary route of spread for *Mycoplasma bovis* is through the movements of infected animals or by feeding calves infected milk purchased from infected farms and so the tracing of animals moved onto and off of case farms was a significant focus of the early response operations. This was performed both by interrogating the NAIT system for historical movement data as well as collating data from multiple other sources to create a census of movements for each farm. These sources included one-on-one semi-structured interviews with affected farmers, cross checking of Animal Status Declaration (ASD) documents, and rescanning the RFIDs of all animals present on a farm. It is currently unknown to what extent not having complete, accurate, and timely data on all movement contacts available through the NAIT system when *Mycoplasma bovis* was first detected has contributed to the silent spread of disease.

Quantifying the current limitations in the NAIT data is vital to help New Zealand better prepare for potential future disease incursions. The aim of this study was to quantify the magnitude of missing cattle movement data in the New Zealand NAIT system. We used two different methods, in the first method, we use the census of animal movements collected during the course of the *Mycoplasma bovis* response activities. However, given the considerable amount of effort it took to collect these data, we were also interested in exploring other options for conducting routine monitoring of data quality. The second part of our study was undertaken separately from the *Mycoplasma bovis* programme operations and examined the completeness of the animal movement history available for animals which arrived at 3 abattoirs in 6 two-month periods. This dataset was chosen since the scanning of RFIDs at slaughter is part of the production chain in New Zealand and compliance at this point is known to be high. These data were also used to explore farm level risk factors for the underreporting on movements. The implications of data quality on the course of the *Mycoplasma bovis* outbreak and response are discussed.

7.3 Materials and Methods:

Mycoplasma bovis case farms:

Sample population

This study included data from all 160 case farms that were present in the eradication programme database on 15 March 2019. Each of these farms were identifiable by farm type, cattle population and location. Locations are not displayed due to privacy requirements, however coordinates were used to calculate the distance between farms.

Data acquisition and analysis

Information related to the movements of cattle onto and off case farms involved in the *Mycoplasma bovis* eradication programme was obtained from the eradication programme database. Movements to slaughter were excluded from the database since they did not present a viable pathway for onward disease transmission from case farms. The movement data for each farm included the existing NAIT movement records, which were then augmented with multiple additional sources of animal movement information as described below. When a movement was identified that was absent from NAIT, the movement was added and designated with the name of the data source it came from. If a movement was present both in NAIT and in another data source, the movement was recorded as having been sourced from NAIT. This resulted in a census of animal movement information from multiple sources with NAIT as the baseline data and four additional data sources used to supplement NAIT.

a. NAIT data: When a farm in the programme met the case definition of being infected, NAIT records of all movements onto and off the farm for a specified period were queried in the NAIT database. The period of interest was established by assessment of existing intelligence around the suspected infection date of the farm and the assessment of a veterinary epidemiologist.

b. Interview data: A one on one interview between a veterinary epidemiologist or appropriately trained large animal veterinary practitioner and the owner/manager of any newly detected case farm was conducted as soon as possible after identification of a case farm. The purpose of the interview was to validate the recorded movements in the NAIT database and to supply any additional missing information surrounding animal movements as well as to describe the farming operation in detail and provide clinical and on farm management history and detail.

c. Animal Status Declaration documentation (ASDs): The ASD is a requirement under the Animal Products Act of 1999 which must accompany a group of animals when they are moved to any location. As these documents detail the sender of the animals in question, it was possible to attempt to match these with NAIT records to identify any movements that were missing from NAIT.

d. Public reports: Due to the large amount of media attention caused by the initial identification of *Mycoplasma bovis* in New Zealand and the subsequent eradication programme, many famers who had received animals from known case farms voluntarily reported these to MPI.

e. Animal censuses on farms: A census of all animals present on case farms was conducted to determine compensation for farmers under the Biosecurity Act of 1993. However, an essential additional benefit of this activity was generating a list of RFIDs for all animals on the farm, which could be queried in the NAIT system to identify if any movements in their lifetime had not been recorded (i.e. animals that were previously known to be located on a different farm with no recorded movement onto the case farm). This activity was particularly valuable for identifying many back traces to farms which could have been the source of infection

All movements were manually populated into the response database by tracing team members and cross-checked against other data sources to confirm the identity of sending and receiving farms. This was necessary due to the lack of a single farm identifier which is applied across all properties in New

Zealand. Each movement record in the response database included the sending and receiving farms, the number of animals, the stock class (calves, bulls, cows, beef or dairy) where available, the date of the movement and the means by which it was collected.

Data were imported into a standard statistical package for processing and analysis (Team, 2017). Farms were identified by their AgriBase numbers and NAIT numbers matched spatially to the NAIT number via LINZ during data manipulation as a part of the eradication programme (Jewell et al., 2015). To explore whether the distance of the movement influenced the likelihood of it being omitted from NAIT, distances between farms were calculated by Euclidian geometry and descriptive analysis of the distances represented by each data source displayed. The relationships between distance and missingness from NAIT, completeness of NAIT recording and production types and completeness of NAIT recording and North or South Island were tested using Mann-Whitney_U or Kruskal Wallis tests (Hodges and Lehmann, 2007). Demographic information related to the type of animals moved (dairy, beef and age group) was not available in sufficient numbers of records to allow statistical tests of the relationship between animal demography and missingness from NAIT. The farming operations were classified based on their self-identified main activity. Any farms raising animals for meat were classified as beef operations, both dairy platforms and dairy support properties were classified as dairy and those farms aggregating calves from dairy operations and selling them on to beef operations after weaning were classified as calf rearers.

Movements onto abattoirs:

Data acquisition

An initial pilot analysis was performed involving a single abattoir for a limited time period. Data on all animals processed at this facility in a 35-day period between 9 January 2017 and 1 March 2017 were acquired from the NAIT database for analysis. Taking advantage of the high level of RFID scanning compliance at slaughter, we chose to examine those animals which had only one movement recorded in their lifetime – from a NAIT number to slaughter and to compare the farmer sending the animal to slaughter with the farmer who purchased the RFID tag. In cases where the sender of the animal and the tag purchase were the same, we assumed that no movement information was missing from the particular animal’s lifetime history. However, if the tag purchaser and the sender of the animal to slaughter were different, we assumed that movement data were missing from NAIT.

Following initial analysis of the pilot data, there were concerns about the quality of NAIT data and it was deemed useful to design a larger study, which would examine records over a larger range of time periods and in additional geographic locations to assess the magnitude of the variation by region and time of year. Three abattoirs were identified for inclusion based on geographic location, one from the Waikato region in the North Island (WO), and two from the South Island, in Canterbury (MC) and Southland (SL) (exact locations are not given to prevent the identification of individual abattoirs). For each abattoir, lists of RFID identifiers were acquired for the period between 15th of January and 15th of May and the period between 1st October and 1st December for both 2017 and 2018. These time periods were selected based on knowledge of regional New Zealand production systems for both milk and beef production and were optimized to include dairy and beef bull culling post mating, the annual dairy cattle cull and the activities of New Zealand's extensive beef fattening system. For analysis purposes, we have categorized the time periods into 2 month blocks as follows: A17 is the period from 15th Jan to 15th March 2017; B17 is the period from 16th March to 15th May 2017 and C17 is between 1 October and 1 December 2017. The same time periods in 2018 are denoted as A18, B18 and C18.

The RFIDs were then queried in the NAIT database to recover the associated lifetime movement history for each of the slaughtered animals. Each line of movement history data has a sending NAIT number, a receiving NAIT number, associated contact information and a movement date. For those animals with only one movement recorded in their lifetime, the NAIT number of the farmer that purchased the tag was recovered from the NAIT database.

Data processing and analysis

Data were imported into a standard statistical package for processing and analysis (Team, 2017). Counts of animals processed, counts of consignments arriving at each abattoir and counts of supplying NAIT numbers at each facility in each time period are presented. As a single NAIT number was able to supply animals across more than 1 time period in our dataset, we calculated a count of unique supplying NAIT numbers by period and abattoir. We then separately calculated the number and distribution of consignments by aggregating RFIDs based on abattoir, date and period. Animals arriving at a particular abattoir on a particular date from a particular supplying NAIT number were designated as being in a consignment.

The lifetime movement history for each animal was analysed and the count of recorded movements described as frequency histograms by season and processing facility. All RFID tags which had only one movement in their lifetime history (to the abattoir) were retained for further analysis. This choice was made based on the fact that an animal with only a single movement in its lifetime from its birth farm to an abattoir should have the same farmer recorded as the tag purchaser as the sending farm to the abattoir. The NAIT numbers of the farmers that purchased these single movement tags were queried in the NAIT database. The purchasing NAIT number was then compared with the NAIT number that sent the animal to the abattoir. Mismatches between the tag purchaser and the sending NAIT number were recorded and described by abattoir and by time period. The odds of the tag purchaser and the sending farmer being one individual was adjusted by abattoir and time period and tested in a simple logistic regression model. As NAIT was made a legal requirement in 2012, a binary outcome was assigned for if the RFID tag was purchased before or after this date and tested for significance in the regression model.

7.4 Results:

Movements onto and off case farms:

Of the 160 case farms 62 were beef farms, 75 were dairy farms and 23 were lifestyle and other farm types. There were 17730 animal movements recorded in the programme database for these farms over the time period from 21 July 2017 when the database was first established to 13 March 2019 when the data were extracted for this analysis. Of these 17730 movements, 13768 (78%) had data source recorded and the rest had either blank or null as their data source. NAIT was the source of 90% of the recorded movement events (Table 24). A smaller number (587: 4.3%) were obtained by interview. While 587 movements is a small proportion of the movements examined, it remains a large number of animals to follow up. Of all movements where cattle type was available (Table 24), interviews were the source of 34% of calf movements and 42% of bull movements. Where cattle type was available, NAIT accounted for 91% of the beef animal movements, but far less of the dairy cow (47%), calf (49%) and breeding bull (33%). The distance moved between sites was left-skewed with a median of 68.14 km, a mean of 175km and a maximum of 1328km. There was no significant statistical difference between the distances recorded for movements in NAIT when compared with the movements captured by other data sources (interviews, ASDs, censuses and public reports) when these were examined using a Mann-Whitney-U test (p=0.056). However the p value approaching 0.05 indicates the presence of a trend. The mean percentage of movements that were recorded in NAIT was similar for both islands, and slightly right-skewed 72% (median 77%) in the North Island compared with 58% (median of 72%) in the South Island. The difference between the North and South Islands was not significant when tested using a Kruskal Wallis test (p= 0.17). A Kruskal Wallis test of the difference in NAIT recording completeness between cattle enterprises did identify a difference in rates of NAIT completeness between calf rearing operations, beef farming operations and dairy operations (p= 0.0003) (Table 25).

Table 24: *Counts and percent of the total of animal movement records that had data source classifications in the Mycoplasma bovis database and where available the type of animals involved in the movement event. Percentages in brackets show the proportion of animal type reflected in each collection method.*

	Beef cattle	**Dairy cattle**	**Bulls**	**Calves**	**Cattle type blank**	**Count of records**	**Percent of Total**
NAIT	155 (91.18%)	93(47.45%)	53 (32.52%)	135 (48.91%)	12114 (93.45%)	12550	91.2%
Interview	13 (7.65%)	65(33.16%)	69 (42.33%)	95 (34.42%)	345 (2.66)	587	4.3%
Census	1 (0.59%)	12(6.12%)	4 (2.45%)	3 (1.09%)	403 (3.11%)	423	3.1%
Public reports	1 (0.59%)	15(7.65%)	19 (11.66%)	29 (10.51%)	46 (0.35%)	110	0.8%
ASDs	0 (0%)	11 (5.61%)	18 (11.04%)	14 (5.07%)	55 (0.42%)	98	0.7%
Total	170	196	163	276	12963	13768	100%

Table 25: *For those farms investigated as a part of the Mycoplasma bovis eradication campaign, the proportion of all farm movements which were present in NAIT was calculated by treating the total movements from all sources as the denominator and the movements recovered from NAIT as the numerator. These proportions were further stratified by farm type. Counts of farms in each category are in brackets. Percentages indicate summary statistics for completeness of NAIT recording in each farm category.*

Farm type (count in sample)	**minimum**	**25 percentile**	**50 percentile**	**mean**	**75 percentile**	**maximum**
Beef raising and finishing farms (56)	0	0.6364	0.8583	0.7312	0.9906	1
Calf rearers (11)	0	0.2519	0.6667	0.5744	0.899	1
Dairy platforms and dairy support (51)	0	0.032	0.5254	0.4711	0.8449	1
All farms (118)	0	0.28	0.74	0.603	0.92	1

<u>*Cattle abattoir study:*</u>

Pilot Data: A total of 10988 animals were received for slaughter at the single abattoir during the 35-day trial period. The minimum number of animals received on any single day was 148; the mean was 313 and maximum was 377 animals per day. Animals were sent by 615 unique suppliers and most animals were sent in small consignments of 1 to 10 animals from a single supplier. Only 7 consignments exceeded 100 animals. Of the 10988 animals, 2628 had no movement history prior to arrival at the abattoir (23.9%). Of these 2628 animals, 38% had been registered on the sending farm (no missing movement information). The remaining 1629 animals (62%) had RFID tags which had not been registered by the farmers that sent the animals to the abattoir and so the intervening animal movement history for these animals was unknown.

Full study Data: The total number of animals that were sent to the three facilities in the specified time periods was 384185 animals in 26746 consignments. These came from 6423 unique supplying NAIT numbers and the consignment size varied by season and abattoir as shown in Figure 16 and Table 26. Seven large consignments (ranging between 243 and 527 animals) were removed from Figure 16 to allow display. All 7 of these were from the SL abattoir and in period B18. There were 170307 animals with only one recorded movement in NAIT (Figure 17). Of these, 53 257 could be shown to have properties missing from their lifetime movement history in NAIT (Table 26). The number of locations that a single animal could have been on prior to slaughter according to the data available from NAIT varied from 1 to 22 with a mean of 2 (median: 2.3). The variation by abattoir and time period is shown in Figure 18. Table 27 shows the results from the logistic regression analysis predicting the level of missing data based on abattoir and time period. When the referent category was time period A17, all other time periods had statistically less missing data than the referent category when we adjusted for the effect of region using processing plant as a proxy (Table 27). The coefficient point estimates trend toward an increase in data quality over time, and the confidence intervals around the odds ratios overlap only for the period between January and March 2018 and the period between March and May 2018. A term indicating if the tag was purchased before or after NAIT became a legal requirement in 2012 was not statistically significant when tested in the model.

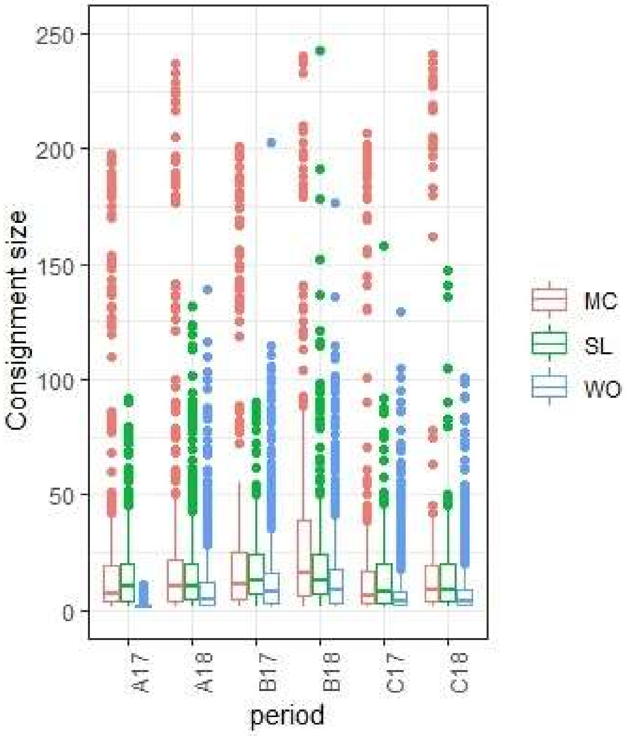

Figure 16: *Boxplot of the distribution of consignment sizes amongst three abattoirs in 6 time periods in New Zealand. Animals arriving at a particular abattoir on a particular date from a particular supplier were allocated to a consignment. A17 is the time period from 15th Jan to 15th March 2017; B17 is the period from 16th March to 15th May 2017 and C17 is between 1 October and 1 December 2017. The periods in 2018 are denoted as A18, B18 and C18. MC, SL and WO are abattoirs in the Canterbury, Southland and Waikato regions respectively.*

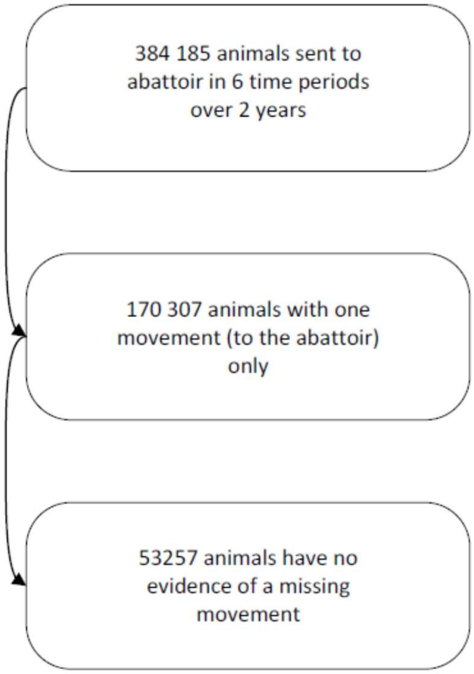

Figure 17: *Diagram showing the selection process of animal movement records used to estimate the extent of missing data from NAIT over six time periods in 2 years at 3 abattoirs.*

Table 26: *Description of the available data from 3 abattoirs in 3 time periods in New Zealand. The counts and proportions of instances of missing information and complete information (Tag purchaser and sending NAIT number are the same) are shown along with the counts of suppliers and consignments.* [1] *Proportion calculated with a denominator of those animals with one recorded movement.* [2] *Proportion calculated with the total number of animals slaughtered as the denominator. It is important to note that a single animal with missing information could have more than one missing farm in its lifetime movement history. A17 is the time period from 15th Jan to 15th March 2017; B17 is the period from 16th March to 15th May 2017 and C17 is between 1 October and 1 December 2017. The periods in 2018 are denoted as A18, B18 and C18.*

		Animals with missing information (count)	**Tag purchaser and sending NAIT number are the same (count)**	**Only 1 recorded movement (count)**	**% with missing information[1]**	**Total Animals slaughtered (count)**	**% with 1 recorded movement only[2]**	**count of suppliers**	**count of consignments**
Pilot Data		1479	900	2379	62.17	10988	21.65	615	not calculated
Study Data		116980	53327	170307	68.69	384185	44.33	14964	26746
MC	A17	3915	1178	5093	76.87	18842	27.03	548	988
	B17	5153	1812	6965	73.98	20006	34.81	434	855
	C17	2784	802	3586	77.64	13896	25.81	411	645
	A18	4101	2028	6129	66.91	21305	28.77	497	1009
	B18	4308	2426	6734	63.97	20585	32.71	370	690
	C18	815	546	1361	59.88	9277	14.67	240	346
SL	A17	7580	2320	9900	76.57	22948	43.14	875	1565
	B17	15318	5071	20389	75.13	36163	56.38	993	2107
	C17	3272	1382	4654	70.31	12343	37.71	561	840
	A18	9027	4754	13781	65.5	30305	45.47	1026	1990
	B18	11559	6126	17685	65.36	34125	51.82	925	1758
	C18	2587	1831	4418	58.56	10782	40.98	525	715
WO	A17	851	293	1144	74.39	2199	52.02	797	1139
	B17	15852	5650	21502	73.72	34759	61.86	1411	2759
	C17	4465	2304	6769	65.96	17340	39.04	1367	2227
	A18	7078	3933	11011	64.28	24128	45.64	1322	2366
	B18	15573	8498	24071	64.7	41701	57.72	1516	3067
	C18	2742	2373	5115	53.61	13481	37.94	1146	1680

Figure 18: *distribution of the number of recorded movements that a single animal had recorded in its lifetime in different time periods and at different abattoirs in New Zealand. A17 is the time period from 15th Jan to 15th March 2017; B17 is the period from 16th March to 15th May 2017 and C17 is between 1 October and 1 December 2017. The periods in 2018 are denoted as A18, B18 and C18.*

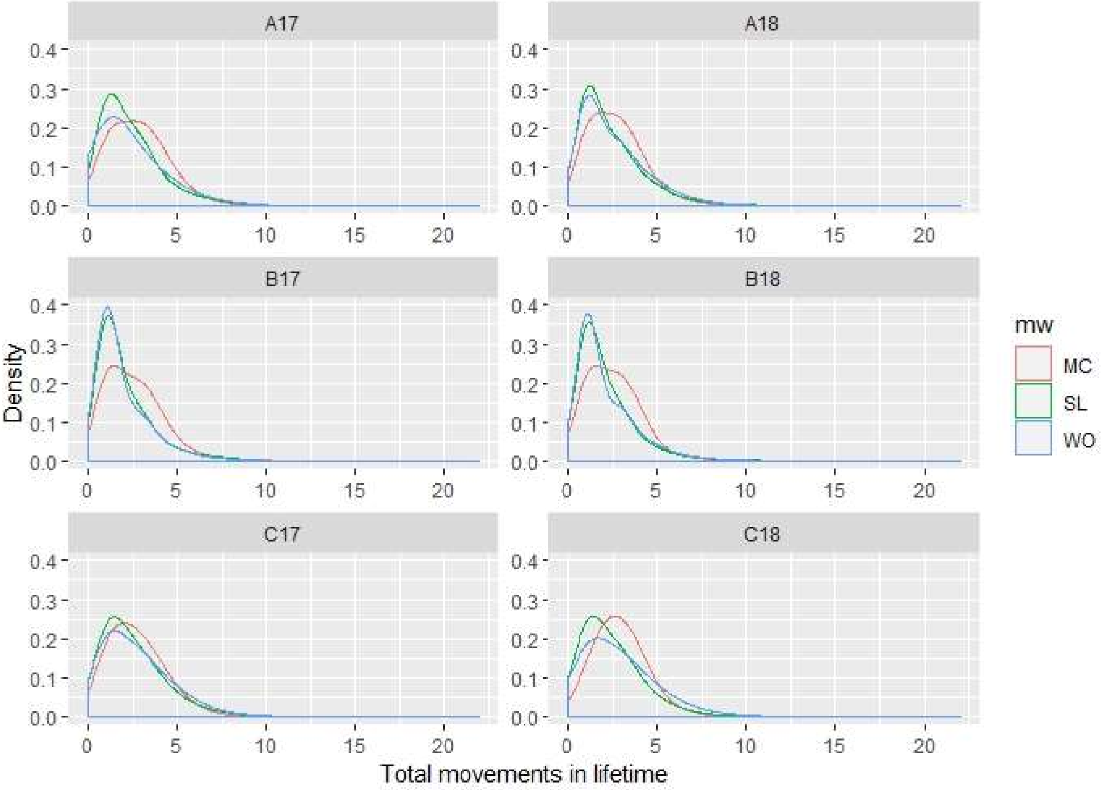

Table 27: *Logistic regression model table showing the role of time period and region against movements with no missing data at three New Zealand abattoirs. The time period between 15th of January and the 15th of March 2017 (A17) is the referent category for time and the abattoir designated as MC is the reference category for the abattoir fixed effect. The outcome variable in our model was a 1 if the animal had no missing information identified in the step from supplier to abattoir.*

	coefficient	standard error	p value	OR	2.50% CI	97.50% CI
intercept	-1.21	0.021	< 0.001			
16th March -15th May 2017 (B17)	0.08	0.022	< 0.001	1.085	1.04	1.13
1st October – 1st December 2017 (C17)	0.30	0.026	< 0.001	1.345	1.28	1.42
15th January – 15th March 2018 (A18)	0.52	0.022	< 0.001	1.682	1.61	1.76
16th March – 15th May 2018 (B18)	0.53	0.021	< 0.001	1.698	1.63	1.77
1st October -1st December 2018 (C18)	0.89	0.027	< 0.001	2.425	2.30	2.56
SL	0.03	0.015	0.04	1.032	1.00	1.06
WO	0.11	0.015	< 0.001	1.117	1.08	1.15

7.5 Discussion:

This analysis described two different methods for examining the extent of missing data in a national animal movement database: one relying on intensively collected data from an outbreak response and the other relying on passively collected data from routine slaughtering of cattle at abattoirs. Both methods revealed that the NAIT data was incomplete. When the farm level data were examined, 5% of the data were not present in NAIT but were identified by other means. We can never know the full extent of the missing information in this analysis as even with the addition of adjunct data sources, there is no way to sure that a census of movements for each farm was obtained. For the abattoir analysis, 62% of those animals that had only one registered movement could be proven to have missing movement links records. It is important to bear in mind that the 62% percent does not indicate a proportion of missing movements, but rather a proportion of animals with incomplete movement records when only those animals with a single recorded movement are assessed. It is likely that some of these animals have more than one missing movement and this situation is only compounded when those animals with more than one movement in the data are examined.

There are additional caveats to consider when interpreting the meaning of the 5% and 62% findings. Firstly, different units of observation were used for each method. For the outbreak response data, data are missing at the farm-level and the investigations focus on filling in gaps in the farm level network connected to the identified case farm. Here the focus is on proportions of movements. In addition where missing data were detected using advanced analysis of the NAIT data, these movements were was classified as sourced from NAIT. Although the source of the data was NAIT, the farmers did not necessarily populate the data in the NAIT system.

For the abattoir study, missing data are at the animal level, and the proportion of animals with only one recorded movement (to the abattoir) was the unit of interest These animals were identified when the sending farmer was not the same person who purchased the RFID tag despite the fact that the animal was apparently sent from its birth farm to slaughter. It is possible that the case farms are not representative of the general population as the process of being diagnosed as a case of *Mycoplasma bovis* takes several weeks to months and involves multiple interactions with animal health officials, some of which may actively encourage farmers to update their NAIT records. Since farmers can log on to NAIT at any time and complete historic movement records, it is likely that these interactions may have motivated them to update their NAIT data before the in-depth contact tracing interviews were conducted, which would lead to a significant underestimation of the extent of missing data. We regard our study using the movements to slaughter as less biased as the vast majority of the sending farms would not have been exposed to a large number of direct interactions and reminders related to animal movement recording. Finally, abattoirs have better compliance with RFID scanning and ASD documentation than farm-to-farm movements due to greater regulatory oversight of these facilities compared with farmers.

Through examining the outbreak response data, we were able to estimate the relative contributions of each data source towards building a complete census of historical movements from a farm. NAIT was the largest contributor with 91% of all movement records coming directly from this source. Given the large amount of records that were missing data on animal type, it is not possible to provide definitive evidence of which demographics are under recorded in NAIT. However, it appeared that movements of beef cattle were more completely captured than movements of calves, dairy cattle, and bulls that were more frequently identified through other data sources.

One of the challenges in being able to fully quantify the extent of missing NAIT data using the abattoir records was that we could not identify the number of locations that any individual animal was present on prior to its arrival at the processing plant. This includes both the animals where the last known location prior to the movement onto the abattoir was different than where the tag was registered and the animals where the locations were the same since these may have movements to and from the farm in the time period between registration and slaughter. Therefore, the 62% of animals missing movement data could grossly underestimate the total number of movements missing from NAIT. A larger and more detailed analysis of animal lifetime movement data comparing it against other data sources would provide more insights but would present practical limitations. The extrapolation of data missing from those animals with only a single step in their lifetime history to those animals with more than one movement in their lifetime movement history will make interesting and important future work. Even without further analysis, these findings show that even for those animals that appear to have gone direct from farm of origin to an abattoir according to their NAIT records, have a 1 in 3 chance of having an unregistered movement to at least one other property in their lifetime movement history. Thus, the possibility that an animal could have been on another farm and spread disease prior to slaughter cannot be ruled out.

The logistic regression analysis indicated a trend towards the data quality of NAIT gradually increasing over time. This improvement in data quality could be attributed to the increased awareness that the New Zealand farming community has of the importance of NAIT for biosecurity purposes in the wake of the *Mycoplasma bovis* eradication announcement and subsequent depopulation of farms. Tags purchased prior to NAIT becoming a legal requirement in 2012 did not show a statistically different rate of having incomplete data than those purchased after that date. This seems to be in contrast with findings related to the UK animal movements database where a marked improvement in data quality was identified once recording became mandatory, a possible reason for this is that we were not able to adjust the model for the age of the animals but expect that the majority of animals would have been younger than 6 years of age (Green and Kao, 2007). The processing plant where the animal was slaughtered had a smaller but still statistically significant effect, which could be interpreted as a proxy for regional variation in farmer compliance with recording NAIT data, variation in farming activity and the regions where the *Mycoplasma bovis* eradication programme is ongoing.

The increased awareness of the importance of recording animal movements and the change in farmer behaviour in the face of an outbreak have impacts on the future of the *Mycoplasma bovis* eradication campaign (Hidano et al., 2018; Hidano and Gates, 2019). It can be expected that as *Mycoplasma bovis* is a relatively slow-moving disease which can have a latent carrier condition, the odds of an animal being sold will not be influenced by infection status. In addition, the highly linked structure of the New Zealand farming systems is unlikely to change but the relative importance of different nodes will alter based on the placing of movement controls and the attitude of the community to farms being investigated even if they are found to be not infected (van Segbroeck et al., 2010). The drivers of animal movement under-reporting must be fully understood in the New Zealand context so that strategies to improve compliance can be deployed. This is essential so that New Zealand can realise the full potential of its animal movement recording system to support disease response.

7.6 Acknowledgements

This study would not have been possible without the hugely dedicated *Mycoplasma bovis* tracing team. Special thanks are due to Karyn Froud, Kara Dawson, Linda Collard, Fanny Leduc and Dale Kaio.

CHAPTER 8

8 Conclusion

The overarching aim of this thesis was to explore the impact of poor quality national animal demographic and movement data on the ability of competent authorities to prepare for and respond to animal disease outbreaks. Firstly, accurate predictions of the size and location of farm animal populations in New Zealand were attempted by combining information from multiple national databases, which highlighted numerous challenges in uniting New Zealand's fragmented data landscape (Jewell et al., 2015). Using an established disease simulation model, the effects of using inferred data on herd sizes in New Zealand to predict the course of an FMD epidemic in a disease free situation were explored. This was followed with a comparison of the use of local-level disease reports from villages and national-level indices in Myanmar against serological data to predict the FMD status of villages in an endemic country. Finally, data collected through the intensive outbreak investigation activities for *Mycoplasma bovis* in New Zealand and animal movement data available in the national livestock movement database were compared to estimate completeness of the national database.

The disease control activities of veterinary epidemiologists and policy-makers are profoundly limited by having fragmented, incomplete, and out-of-date national-level animal demographic and movement databases. This is largely because the starting conditions and assumptions made about animal populations in disease simulation models can influence the predictions and potentially lead to the deployment of suboptimal disease control strategies under emergency conditions (Van Andel et al., 2018). Although there have been many studies looking at the impact of changing specific disease spread parameters on model predictions, this thesis is one of the first pieces of work to thoroughly explore the impacts of biases and errors in demography. For example, the models in Chapter 3 showed that when using cost as a way to measure outbreak impact, only 39% of simulation runs based the inferred dataset identified the same lowest operational cost strategy as simulations based on the real dataset, which could potentially to lead to millions of dollars of extra cost to the industry as well as delays in being able to resume normal farming and trade activities.

Aside from the impacts on the choice of strategy, the operational delays caused by inaccurate data would create further delays to resuming international trade.

A major challenge with the New Zealand studies in this thesis was the amount of time and effort it took to get the national datasets linked together and in a format that could be used for analysis. In a real outbreak situation, it would take significant time, manual work, and many confirmatory phone calls to transform the multiple existing sources of imperfect data into a cohesive databases that provides epidemiologists and policy-makers with useful intelligence to inform decisions. This time will not be available in the case of a fast moving infectious disease like FMD and these significant delays to the government being able to respond effectively highlights the importance of New Zealand investing resources now to address this problem.

Unfortunately, it is often expensive and logistically challenging to maintain up-to-date farm population data since this relies on farmers to provide accurate records in a timely fashion on top of balancing the many other demands on their time. In Chapter 2, a Bayesian model was built to predict the presence and count of cattle and livestock units on New Zealand farms to see if this was a viable alternative for filling in missing or incomplete data in the national databases. The Bayesian approach was chosen over traditional regression models since the posterior distributions provides a mechanism to incorporate uncertainty around a population distribution into a disease spread simulator. Concurrently, machine learning models were also trialled with this analysis and investigations were made into the use of remote sensing as a way to aggregate animal population data (Hollings et al., 2018, 2017). Although this approach is commonly used in wildlife ecology, using these estimates for biosecurity activities could introduce systematic error. As the results showed, the predictions of the model were more reliable at the larger spatial scale, but not suitably accurate at the farm level for use in a disease outbreak response. In the case of the large New Zealand population models, the random forest models were most predictive and outperformed the traditional statistical approaches at predicting the presence and size of populations (Hollings et al., 2017). For the smaller predictive models in Chapter 5, there was little difference between the approaches in terms of predictive performance. The key conclusion from these different farm animal population modelling projects is that using predicted data on animal populations is no substitute for real data. Investment must be made in improving the New Zealand data landscape by designing databases and database linkages which combine the many sources of essential information which are present in New Zealand. A single source of true information is essential, achievable, and responsible use of tax payer money.

Another large and concerning gap in the New Zealand animal demographic databases is accurately determining the location and number of lifestyle block owners, which are typically peri-urban and dominated by hobby farms not engaged in primary production activities driven by profitability. These farms were very poorly predicted by our models as the presence of suitable high quality pasture does not predict the presence of livestock, nor does its absence ensure that no FMD susceptible animals are present. This group presents a unique challenge because the drivers of their farming behaviours are unpredictable and undocumented and they have no single sector representative body. This sector has been implicated in the introduction of transboundary animal diseases in many countries worldwide and represent a significant risk to New Zealand due to their peri-urban locations, which are geographically close to potential points of disease entry into New Zealand. Aside from the potentially risky location of these farms and the potentially non-standard husbandry practices which could place them at higher risk for a transboundary animal disease, these owners might have different awareness and motivations for reporting diseases to veterinarians than farmers in the commercial sector, which could lead to delays in the recognition of disease incursions. These small and very diverse animal populations are also virtually impossible to model because even the five yearly animal census in New Zealand and the annual surveys do not capture this population since these only include those farming businesses registered for general sales tax. Even if lifestyle block owners were captured in the census surveys, the farm-level data is cannot be used for biosecurity purposes under the Statistics Act of 1975. Possible solutions to this problem include the an expansion of the agricultural production census target population to capture information relating to lifestyle farmers and making the registration of any livestock mandatory against a single New Zealand farm identifier.

These challenges with data availability are not unique to New Zealand. In 2017, a short unpublished qualitative survey of the baseline data (animal movement and animal location and population data) used by the FMD spread modelling teams in 6 countries was completed. These countries included the signatory countries to the Quadrilateral agreement on biosecurity [10] as well as the United Kingdom and Sweden. The purpose of the study was to assess the level of data availability and currency amongst the participating countries and to identify areas where shared work could be undertaken. All countries identified ongoing challenges with data linking across multiple data sources as well as data accuracy, data gaps associated with the lifestyle sector, and data currency. This is an important area for epidemiologists to be aware of as disease-spread models are evolving rapidly and require more complex population and movement data. Aside from this, the data are

---

[10] namely Australia, New Zealand, Canada and the United States of America

essential for disease response and disease control programmes. No country except for the United Kingdom, was able to link their animal census information at the actual (rather than simulated) farm location level or allowed to use non-aggregated farm animal information (from agricultural census) as population data for their models. No country except Sweden had regulations that included the registration of all animals including those on lifestyle farms.

In comparison with most of the participants, the New Zealand situation appears favourable as the country has national level farm location and animal movement databases. This is in contrast with Australia, Canada and the United States each have different levels of data recording amongst their states or regions and must contend with the difficulties of combing the state or regions level data to be used at the federal or national level . That said, New Zealand is a small island country with a single national (rather than state level and a federal level) animal health and biosecurity authority which could be considered as more comparable with one of the Australian states than to the whole of Australia. Even though the situation is favourable in New Zealand, it is fundamentally more important that New Zealand invest in this area as primary industry exports are essential for the economy. An FMD outbreak in New Zealand would result in a large loss of national income with most of the impact falling on the primary sector and agricultural processors (Forbes and van Halderen, 2014). Given New Zealand's international market share, cessation of overseas trade would depreciate New Zealand's exchange rate. Trade would only be able to resume following successful eradication from the whole country (or from a zone as defined by the OIE) and the timing of return to trade depends on trading partners (although the minimum time to resume trade is specified by the OIE). Anything that slows the disease eradication programme will exacerbate these effects. This affects all FMD countries to come degree, but according to the CIA world fact book, New Zealand has the highest GDP from agriculture of all the participating countries (4.9%), as well as the smallest total land area under use in agricultural systems of all 6 countries that participated. This indicates that an animal disease outbreak with trade implications would have a disproportionately large impact on the New Zealand economy when compared with countries with similar biosecurity priorities.

Chapters 5 and 6, then explored the situation in Myanmar, which has endemic FMD and no centralised database of animal counts and animal movements. Two alternate sources of national data were used: village livestock census lists collected by district veterinary staff as part of in-depth survey for the subset of townships in central Myanmar included in the FMD seroprevalence study and also national modelled population data from the Gridded Livestock of the World (GLOW) resource maintained by FAO (Nicolas et al., 2016; Robinson et al., 2014). From the latter dataset, both cattle population size and proximity to animal movement hubs were found to be predictive of FMD status at the village level and these variables performed just as well as data collected through the detailed survey. Proximity to cattle markets which act as movement hubs was available using simple geospatial methods. The role of population data accuracy at the village level was not assessed in this thesis, but was subsequently examined by other researchers on the Myanmar project team and found to be more useful for predicting than models using population information derived from GLOW. This again highlights the importance of being able to collect up-to-date information rather than making inferences from other passive data sources.

Chapter 6 focussed on the FMD endemic disease situation in Myanmar through the lens of outbreak reporting and disease recognition. Here the local knowledge of village headmen and householders was compared. The question at issue was their perception of the FMD status of their village compared with the results from serological screening of different age groups of cattle. Despite the high prevalence and endemic nature of FMD in the region, enough animals were still showing clinical signs that village householders were able to relatively accurately identify the presence of disease. This finding is apparently at odds with the low rates of centralised disease reports of FMD which are collated by the OIE, which suggests a behavioural gap between identify FMD and taken action to report the disease to central authorities. Furthermore, the level of agreement between the village reports and serological data varied across the different townships in Central Myanmar, which suggests the need for further research to understand these variations and develop interventions to improve reporting in certain areas. One possible explanation is that officials in some regions are more proactive with their FMD communication campaigns. This increased activity would lead to greater awareness of FMD amongst livestock owners thereby increasing recognition of disease and reporting to authorities. An examination of the behaviors that drive the reporting of FMD cases at every level (householder to headman, headman to district veterinary office, district office to national office) would be very useful to assess if a sustainable way to monitor disease prevalence could be designed. Collated village headman reports or householder reporting using mobile technology are both valuable ways to collect ongoing disease distribution information which may be less resource intensive that serological surveys once initial baselines of disease prevalence are established.

Even though the previous analyses focused on Myanmar, the issue of understanding how local conditions and farmer behaviours act as barriers to compliance with regulatory programmes is highly relevant to New Zealand's passive surveillance system for exotic animal diseases. This system currently has private veterinarians reporting cases of disease which cannot be identified as endemic to a central team of field epidemiologists trained in field investigation of transboundary animal diseases. This system relies heavily on the NZ veterinary profession being aware of which livestock diseases are of interest and being comfortable with reporting the cases to the central authority. Risks inherent in this system include the conflict of interest presented when a veterinarian reporting the disease status of a client to the biosecurity authority has impacts in the form of regulatory controls for the client and is deemed by that client as a breach of trust despite the legal obligation to report exotic diseases and unwanted organisms under the Biosecurity Act of 1997. This situation was observed during the *Mycoplasma bovis* response where farmers that were aware that they had stock which had movement links to infected farms voluntarily restricted their farming business practices for fear of infecting other properties. Despite their good intentions, these farmers did not qualify for compensation if regulatory control was applied under the Biosecurity Act of 1997 and so some were disadvantaged. This could have an effect on farmer willingness to report disease to the regulatory authority in New Zealand in the future thereby compromising the efficiency of passive surveillance.

Even though it is well known that national livestock movement and demographic databases have missing or inaccurate data, it is usually difficult to estimate the magnitude of these errors due to the absence of a gold standard database with the true census of animal movements, farm animal population data, and land use information. An example was presented in Chapter 7 where the minimum extent of missing data for animals with only one recorded movement prior to slaughter could be estimated. To accomplish this, the person that sent the animal to the abattoir was compared with the person registered as the RFID tag purchaser. If all records were correctly recorded and no lifetime movement steps were missing, the tag purchaser should be the same as the person that sends the animal to the abattoir if the animal has only one lifetime movement. Using this method, 62% of animals with only one movement recorded in their lifetime could be identified as having missing information indicating either unrecorded locations in their history or misuse of RFID tags by the producers. It is likely that some of these animals have more than one missing movement and this situation is only compounded when those animals with more than one movement in the data are examined. This volume of missing data is sufficiently large to severely impact on the activities of a disease control programme. Many of the shortfalls in the New Zealand NAIT system have been anecdotally ascribed to the system being difficult to navigate and failing to interact seamlessly with other data sources. Furthermore, when the system was originally designed, the focus was on traceability in the event of an emerging disease outbreak rather than normal compliance activities or endemic disease control. With no additional market value was associated with animal lifetime traceability and no robust mechanisms for punishing farmers who fail to comply, it is also not surprising that there have been significant issues with getting farmers to input data in a timely and accurate manner. While compliance with the system does appear to be improving in the wake of the *Mycoplasma bovis* outbreak, there is still a significant issue with not having enough historical records from before the system was introduced in 2015 to support contract tracing efforts for slow moving diseases with long latent periods like bovine tuberculosis.

Another limitation is that the NAIT system does not currently record much information about the characteristics of the animals being moved or the purpose of the movement. New Zealand has a highly mobile dairy cattle population and some of those movements might be considered to have more risk associated with them than others. In particular the co-grazing of dairy cattle with other stock away from the milking platform, the raising of calves sourced from multiple farms and the introduction of lease bulls could be identified as risk activities that are underrepresented in NAIT. In dairy farms where animal movement information had been collected by interview, Animal Status Declaration (ASD) examination and compared with existing NAIT records, the median farm had 50% of movements missing from NAIT. These missing movements are made up of production movements which occur repeatedly over the farming calendar in New Zealand and many of which may have occurred between the farmers annually over a long duration. Aside from those movements which make up the farming calendar and are regarded as business as usual, the NAIT system does allow for those farms within a 10km radius of a central farm to have the same NAIT number and to not require recording if the farms are used for the same purpose and if the stock have the same person in charge of them. This creates difficulty as the recording of these movements is not required under legislation, but they remain extremely important for understanding disease spread and for disease control activities.

None of the studies presented made use of existing value of information (VOI) frameworks to quantify the value of national animal population and movement databases to government and industry, which is an interesting and important area for further work. The VOI framework provides decision makers with a mechanism for assessing the trade-off between increased accuracy and the costs of achieving that accuracy (Keisler et al., 2014; Yokota and Thompson, 2004). VOI has been applied in the context of epidemiology in the areas of import risk analysis (Disney and Peters, 2003) and decision making under foreign animal disease outbreak conditions (Cox et al., 2005; Probert et al., 2015; Shea et al., 2014). Further applications have been published for decision making in human global health (De Gourville et al., 2006), fisheries management (Forsberg and Guttormsen, 2006; Hansen and Jones, 2008), the utility for the public of labelling of meat products (Klain et al., 2014) and management of natural resources (Williams et al., 2011), invasive species (Moore and Runge, 2012) and the abundance of endemic species (Williams et al., 2011). Animal health, movement and location databases are expensive and difficult to maintain. Having a dollar figure associated with a failure to maintain such resources would be useful for decision makers to understand how much money should be invested to make systems that work. This future research work could include experiments to quantify the effect of animal counts of disease control operations in FMD endemic countries. This will provide insight into the value that animal population data at the farm level has over and above exotic disease response. In the New Zealand case, information held by the *Mycoplasma bovis* programme might be useful to quantify the impacts of incomplete and inaccurate data on disease management and control efforts during the actual response. Based on personal experience, very large numbers of casing and tracing staff have had to be employed to manually match and cross check information gathered from disparate and unlinkable national datasets.

As a result of political pressure generated by the national interest in the *Mycoplasma bovis* eradication programme, it is likely that changes to legislation (in the case of animal movements) and changes to departmental policy will be made. The changes required are conceptually simple: improved compliance with the requirements to record animal movements, an ability to record attributes including health status against a unique animal identifier and a single current farm identifier for all farms which is used at the national level and which industry, government and researchers may use to link the information required for a robust animal health and biosecurity system. Achieving these goals, however, is much more complex. Even when more than one source of information exists, comparison and or combining of various data sources may not be a fruitful exercise when neither data source can be regarded as up to date or accurate. Some value might be derived from the exercise when the two data sources are in agreement or in subsets of the data. In both New Zealand and Myanmar, the drivers which prompt producers to record and report animal movements and the presence of clinical disease have to be fully understood before anything can be done to improve the available information. Once data are flowing in, data linkages must be created and made sustainable. The primary producers and their representatives hold essential information that, when combined with accurate nationally held data provides the key to successful disease management and control programmes as well as robust production and endemic disease research.

9 Appendix to Chapter 6: Latent Class Model Winbugs Code

MODEL 1

```
{
y[1:K, 1:K, 1:K] ~ dmulti(p[1:K, 1:K, 1:K], n)
p[1,1,1] <- pi*SeE*(SeHH*SeVH+covDp) + (1-pi)*(1-SpE)*((1-SpHH)*(1-SpVH)+covDn)
p[1,2,1] <- pi*SeE*(SeHH*(1-SeVH)-covDp) + (1-pi)*(1-SpE)*((1-SpHH)*SpVH-covDn)
p[1,1,2] <- pi*(1-SeE)*(SeHH*SeVH+covDp) + (1-pi)*SpE*((1-SpHH)*(1-SpVH)+covDn)
p[1,2,2] <- pi*(1-SeE)*(SeHH*(1-SeVH)-covDp) + (1-pi)*SpE*((1-SpHH)*SpVH-covDn)
p[2,1,1] <- pi*SeE*((1-SeHH)*SeVH-covDp) + (1-pi)*(1-SpE)*(SpHH*(1-SpVH)-covDn)
p[2,2,1] <- pi*SeE*((1-SeHH)*(1-SeVH)+covDp) + (1-pi)*(1-SpE)*(SpHH*SpVH+covDn)
p[2,1,2] <- pi*(1-SeE)*((1-SeHH)*SeVH-covDp) + (1-pi)*SpE*(SpHH*(1-SpVH)-covDn)
p[2,2,2] <- pi*(1-SeE)*((1-SeHH)*(1-SeVH)+covDp) + (1-pi)*SpE*(SpHH*SpVH+covDn)
ls <- (SeHH-1)*(1-SeVH)
us <- min(SeHH,SeVH) - SeHH*SeVH
lc <- (SpHH-1)*(1-SpVH)
uc <- min(SpHH,SpVH) - SpHH*SpVH
rhoD <- covDp / sqrt(SeHH*(1-SeHH)*SeVH*(1-SeVH))
rhoDc <- covDn / sqrt(SpHH*(1-SpHH)*SpVH*(1-SpVH))
pi ~ dbeta(13.322, 6.281) ## Mode=0.70, 95% sure > 0.50
SeHH ~ dbeta(42.008,28.339) ## Mode=0.60, 95% sure > 0.50
SpHH ~ dbeta(9.628, 3.876) ## Mode=0.75, 95% sure > 0.50
SeVH ~ dbeta(20.996, 11.767) ## Mode=0.65, 95% sure > 0.50
SpVH ~ dbeta(9.628, 3.876) ## Mode=0.90, 95% sure > 0.50
SeE ~ dbeta(14.484, 1.709) ## Mode=0.95, 95% sure > 0.75
SpE ~ dbeta(130.706, 15.411) ## Mode=0.90, 95% sure > 0.85
covDn ~ dunif(lc, uc)
covDp ~ dunif(ls, us)
}
```

```
list(n=161, K=2)

list(pi=0.70, SeHH=0.60, SpHH=0.75, SeVH=0.65, SpVH=0.75, SeE=0.95, SpE=0.90)

y[,1,1] y[,1,2] y[,2,1] y[,2,2]

39 16 5 5

10 17 35 34
```

MODEL 2

```
#hhS is the status (0/1) based on householders; vhR based on village head reporting
#mE is the ELISA score
model{
for (i in 1:n)      {
z[i]~dbern(pi)                          # z=1 for infected
phh[i]<-z[i]*Sehh+(1-z[i])*(1-Sphh)    # Apparent prevalence of hhS
hhS[i]~dbern(phh[i])
h1[i]<-hhS[i]+1                                   # 1 if hhS=0, 2 if hhS=1
pvh[i]<-z[i]*Sevhc[h1[i]]+(1-z[i])*(1-Spvhc[h1[i]])   #Se/Spvhc is conditional Se/Sp given hhS
vhR[i]~dbern(pvh[i])
z1[i]<-z[i]+1
mE[i]~dnorm(mu[z1[i]],tau[z1[i]])          # normally distributed conditional on z
}
for (i in 1:2)      {
tau[i]<-1/sig[i]/sig[i]
}
#Derived parameters:
Sevh<-(1-Sehh)*Sevhc[1]+Sehh*Sevhc[2]                              #unconditional Se/Sp
Spvh<-Sphh*Spvhc[1]+(1-Sphh)*Spvhc[2]
rhop<-Sehh*(Sevhc[2]-Sevh)/sqrt(Sehh*Sevh*(1-Sehh)*(1-Sevh))      #hh-vh correlations
rhon<-(1-Sphh)*(Spvh-Spvhc[2])/sqrt(Sphh*Spvh*(1-Sphh)*(1-Spvh))
#ROC and AUC:
mEp~dnorm(mu[2],tau[2])        # random draw from mE|infected
```

```
mEn~dnorm(mu[1],tau[1])        # random draw from lnE|uninfected
for (k in 1:60)    {
c[k]<- -25+k*2                     # grid of points on mE scale
se[k]<-step(mEp-c[k])             # 1 if lnEp > cutoff, mean gives se
fp[k]<-step(mEn-c[k])
}
AUC<-step(mEp-mEn)            # 1 if lnEp > lnEn, mean gives AUC
#Priors
pi ~ dbeta(13.322, 6.281) ## Mode=0.70, 95% sure > 0.50
Sehh ~ dbeta(42.008,28.339) ## Mode=0.60, 95% sure > 0.50
Sphh ~ dbeta(9.628, 3.876) ## Mode=0.75, 95% sure > 0.50
Sevhc[1] ~ dbeta(20.996, 11.767) ## Mode=0.65, 95% sure > 0.50
Spvhc[1] ~ dbeta(9.628, 3.876) ## Mode=0.90, 95% sure > 0.50
Sevhc[2] ~ dbeta(20.996, 11.767) ## Mode=0.65, 95% sure > 0.50
Spvhc[2] ~ dbeta(9.628, 3.876) ## Mode=0.90, 95% sure > 0.50
mu[1] ~ dunif(-20,40)
mu[2] ~ dunif(mu[1],80)
sig[1] ~ dunif(0,30)
sig[2] ~ dunif(0,30)
}
```

MODEL 3

```
#hhS is the status (0/1) based on householders; vhR based on village head reporting
#mE is the ELISA score
model{
for (i in 1:n)      {
z[i]~dbern(pi)                             # z=1 for infected
phh[i]<-z[i]*Sehh+(1-z[i])*(1-Sphh)    # Apparent prevalence of hhS
hhS[i]~dbern(phh[i])
h1[i]<-hhS[i]+1                                        # 1 if hhS=0, 2 if hhS=1
```

```
pvh[i]<-z[i]*Sevhc[h1[i]]+(1-z[i])*(1-Spvhc[h1[i]])   #Se/Spvhc is conditional Se/Sp given hhS

vhR[i]~dbern(pvh[i])

npos[i]~dbin(P[i],nsamp[i])

z1[i]<-z[i]+1

logit(P[i])<-x[i]

x[i]~dnorm(mu[z1[i]],tau[z1[i]])          # normally distributed conditional on z

}

for (i in 1:2)       {

tau[i]<-1/sig[i]/sig[i]

}

#Derived parameters:

Sevh<-(1-Sehh)*Sevhc[1]+Sehh*Sevhc[2]                                 #unconditional Se/Sp

Spvh<-Sphh*Spvhc[1]+(1-Sphh)*Spvhc[2]

rhop<-Sehh*(Sevhc[2]-Sevh)/sqrt(Sehh*Sevh*(1-Sehh)*(1-Sevh))      #hh-vh correlations

rhon<-(1-Sphh)*(Spvh-Spvhc[2])/sqrt(Sphh*Spvh*(1-Sphh)*(1-Spvh))

#ROC and AUC based on 10 samples:

xpos~dnorm(mu[2],tau[2])        # random draw from mE|infected

logit(Ppos)<-xpos

Xp~dbin(Ppos,10)

xneg~dnorm(mu[1],tau[1])        # random draw from lnE|uninfected

logit(Pneg)<-xneg

Xn~dbin(Pneg,10)

for (k in 1:11)    {

c[k]<- k-1                     # grid of points on mE scale

se[k]<-step(Xp-c[k])             # 1 if lnEp > cutoff, mean gives se

fp[k]<-step(Xn-c[k])

}

AUC<-step(Xp-Xn-0.1)            # 1 if Xp > Xn, mean gives AUC

#Priors

pi ~ dbeta(13.322, 6.281) ## Mode=0.70, 95% sure > 0.50
```

```
Sehh ~ dbeta(42.008,28.339) ## Mode=0.60, 95% sure > 0.50
Sphh ~ dbeta(9.628, 3.876) ## Mode=0.75, 95% sure > 0.50
Sevhc[1] ~ dbeta(20.996, 11.767) ## Mode=0.65, 95% sure > 0.50
Spvhc[1] ~ dbeta(9.628, 3.876) ## Mode=0.90, 95% sure > 0.50
Sevhc[2] ~ dbeta(20.996, 11.767) ## Mode=0.65, 95% sure > 0.50
Spvhc[2] ~ dbeta(9.628, 3.876) ## Mode=0.90, 95% sure > 0.50
mu[1] ~ dunif(-6,-1)
mu[2] ~ dunif(mu[1],0)
sig[1] ~ dunif(0,2)
sig[2] ~ dunif(0,2)
}
```

10 Bibliography

Abila, R., Foreman, S., 2006. Control of Foot and Mouth Disease in Southeast Asia, in: Proceedings of the 11th International Symposium on Veterinary Epidemiology and Economics.

Alexandersen, S., Brotherhood, I., Donaldson, A.I., 2002. Natural aerosol transmission of foot-and-mouth disease virus to pigs: minimal infectious dose for strain O-1 Lausanne. Epidemiol. Infect. 128, 301–312. https://doi.org/10.1017/s095026880100646x

Alexandersen, S., Donaldson, a I., 2002. Further studies to quantify the dose of natural aerosols of foot-and-mouth disease virus for pigs. Epidemiol. Infect. 128, 313–323. https://doi.org/10.1016/S0034-5288(02)90037-8

Alexandersen, S., Kitching, R.P., Mansley, L.M., Donaldson, A.I., 2003a. Clinical and laboratory investigations of five outbreaks of foot-and-mouth disease during the 2001 epidemic in the United Kingdom. Vet. Rec. https://doi.org/10.1136/vr.152.16.489

Alexandersen, S., Zhang, Z., Donaldson, A.I., Garland, A.J.M., 2003b. The pathogenesis and diagnosis of foot-and-mouth disease. J. Comp. Pathol. https://doi.org/10.1016/S0021-9975(03)00041-0

Anonymous, 2017. Terrestrial Animal Health Code. Foot and mouth disease. Office International des Epizooties (OIE).

Anonymous, 2013a. World Organisation for Animal Health (OIE) Terrestrial Animal Health Code [WWW Document]. URL http://www.oie.int/international-standard-setting/terrestrial-code/access-online/ (accessed 2.4.16).

Anonymous, 2013b. Eurostat: Statistics Explained [WWW Document]. URL http://ec.europa.eu/eurostat/statistics-explained/index.php/Glossary:LSU

Astudillo, V.M., Cane, B.G., Geymonat, D., Sathler, A.B., Garay Roman, S., Stutmoller, P., Gimeno, E.J., 2016. Risk assessment and risk regionalisation, based on the surveillance system for foot and mouth disease in South America. Rev. Sci. Tech. l'OIE. https://doi.org/10.20506/rst.16.3.1066

Backer, J.A., Engel, B., Dekker, A., van Roermund, H.J.W., 2012. Vaccination against foot-and-mouth disease II: Regaining FMD-free status. Prev. Vet. Med. 107, 41–50. https://doi.org/10.1016/j.prevetmed.2012.05.013

Backer, J.A., Hagenaars, T.J., Nodelijk, G., van Roermund, H.J.W., 2012. Vaccination against foot-and-mouth disease I: Epidemiological consequences. Prev. Vet. Med. 107, 27–40. https://doi.org/http://dx.doi.org/10.1016/j.prevetmed.2012.05.012

Barnett, P. V., Cox, S.J., 1999. The role of small ruminants in the epidemiology and transmission of foot-and-mouth disease. Vet. J. https://doi.org/10.1053/tvjl.1998.0338

Barteling, S.J., Sutmoller, P., 2002. Culling versus vaccination: Challenging a dogma in veterinary (FMD) science. Rep. Res. Gr. Standing Tech. Comm. Eur. Comm. Control Foot-and-Mouth Dis. 17–20.

Bates, T.W., Thurmond, M.C., Carpenter, T.E., 2003. Description of an epidemic simulation model for use in evaluating strategies to control an outbreak of foot-and-mouth disease. Am. J. Vet. Res. 64, 195–204.

Bates, T.W., Thurmond, M.C., Hietala, S.K., Venkateswaran, K.S., Wilson, T.M., Colston, B., Trebes, J.E., Milanovich, F.P., 2003. Surveillance for detection of foot-and-mouth disease. J. Am. Vet. Med. Assoc. 223, 609–616.

Bellet, C., Vergne, T., Grosbois, V., Holl, D., Roger, F., Goutard, F., 2012a. Evaluating the efficiency of participatory epidemiology to estimate the incidence and impacts of foot-and-mouth disease among livestock owners in Cambodia. Acta Trop. 123, 31–38. https://doi.org/10.1016/j.actatropica.2012.03.010

Bellet, C., Vergne, T., Grosbois, V., Holl, D., Roger, F., Goutard, F., 2012b. Evaluating the efficiency of participatory epidemiology to estimate the incidence and impacts of foot-and-mouth disease among livestock owners in Cambodia. Acta Trop. 123, 31–38. https://doi.org/10.1016/j.actatropica.2012.03.010

Bertram, M.R., Delgado, A., Pauszek, S.J., Smoliga, G.R., Brito, B., Stenfeldt, C., Hartwig, E.J., Jumbo, S.D., Abdoulmoumini, M., Oliva Marie, A.A., Salhine, R., Rodriguez, L.L., Garabed, R., Arzt, J., 2018a. Effect of vaccination on cattle subclinically infected with foot-and-mouth disease virus in Cameroon. Prev. Vet. Med. 155, 1–10. https://doi.org/10.1016/j.prevetmed.2018.04.003

Bertram, M.R., Vu, L.T., Pauszek, S.J., Brito, B.P., Hartwig, E.J., Smoliga, G.R., Hoang, B.H., Phuong, N.T., Stenfeldt, C., Fish, I.H., Hung, V. V., Delgado, A., VanderWaal, K., Rodriguez, L.L., Long, N.T., Dung, D.H., Arzt, J., 2018b. Lack of Transmission of Foot-and-Mouth Disease Virus From Persistently Infected Cattle to Naïve Cattle Under Field Conditions in Vietnam. Front. Vet. Sci. 5, 1–13. https://doi.org/10.3389/fvets.2018.00174

Bessell, P.R., Shaw, D.J., Savill, N.J., Woolhouse, M.E.J., 2010. Estimating risk factors for farm-level transmission of disease: Foot and mouth disease during the 2001 epidemic in Great Britain. Epidemics 2, 109–115. https://doi.org/http://dx.doi.org/10.1016/j.epidem.2010.06.002

Bessell, P.R., Shaw, D.J., Savill, N.J., Woolhouse, M.E.J., 2010. Statistical modeling of holding level susceptibility to infection during the 2001 foot and mouth disease epidemic in Great Britain. Int. J. Infect. Dis. 14, e210–e215. https://doi.org/10.1016/j.ijid.2009.05.003

Bivand, R., Rudel, C., Pebesma, E., Hufthammer, K.O., 2014a. rgeos: Interface to Geometry Engine - Open Source (GEOS).

Bivand, R., Rundel, C., Pebesma, E., Hufthammer, K.O., 2014b. rgdal: Bindings for the Geospatial Data Abstraction Library.

Blake, A., Sinclair, M.T., Sugiyarto, G., 2003. Quantifying the impact of foot and mouth disease on tourism and the UK economy. Tour. Econ. 9, 449–465.

Bouma, A., Elbers, A.R.W., Dekker, A., De Koeijer, A., Bartels, C., Vellema, P., Van Der Wal, P., Van Rooij, E.M.A., Pluimers, F.H., De Jong, M.C.M., 2003. The foot-and-mouth disease epidemic in The Netherlands in 2001. Prev. Vet. Med. https://doi.org/10.1016/S0167-5877(02)00217-9

Bradbury, N. V., Probert, W.J.M., Shea, K., Runge, M.C., Fonnesbeck, C.J., Keeling, M.J., Ferrari, M.J., Tildesley, M.J., 2017. Quantifying the Value of Perfect Information in Emergency Vaccination Campaigns. PLoS Comput. Biol. 13, 1–15. https://doi.org/10.1371/journal.pcbi.1005318

Bradhurst, R.A., Roche, S.E., East, I.J., Kwan, P., Garner, M.G., 2015. A hybrid modeling approach to simulating foot-and-mouth disease outbreaks in Australian livestock. Front. Envornmental Sci. 3, 1–20. https://doi.org/10.3389/fenvs.2015.00017

Brangenberg, N., van Andel, M., 2011. Exotic disease focus: clinical and epidemiological investigation to exclude foot and mouth disease in cattle. Surveill. 38, 4–9.

Branscum, A.J., Gardner, I.A., Johnson, W.O., 2005. Estimation of diagnostic-test sensitivity and specificity through Bayesian modeling. Prev. Vet. Med. 68, 145–163. https://doi.org/10.1016/j.prevetmed.2004.12.005

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32. https://doi.org/10.1023/A:1010933404324

Brier, G.W., 1950. Verification of forecasts expressed in terms of probability. Mon. Weather Rev. 78, 1–3. https://doi.org/10.1175/1520-0493(1950)078<0001:VOFEIT>2.0.CO;2

Brocchi, E., Bergmann, I.E., Dekker, A., Paton, D.J., Sammin, D.J., Greiner, M., Grazioli, S., De Simone, F., Yadin, H., Haas, B., Bulut, N., Malirat, V., Neitzert, E., Goris, N., Parida, S., Sørensen, K., De Clercq, K., 2006. Comparative evaluation of six ELISAs for the detection of antibodies to the non-structural proteins of foot-and-mouth disease virus. Vaccine 24, 6966–6979. https://doi.org/10.1016/j.vaccine.2006.04.050

Bronsvoort, B.M. deC, Toft, N., Bergmann, I.E., Sørensen, K.-J., Anderson, J., Malirat, V., Tanya, V.N., Morgan, K.L., 2006. Evaluation of three 3ABC ELISAs for foot-and-mouth disease non-structural antibodies using latent class analysis. BMC Vet. Res. 2, 1–10. https://doi.org/10.1186/1746-6148-2-30

Brownstein, J.S., Skelly, D.K., Holford, T.R., Fish, D., 2005. Forest fragmentation predicts local scale heterogeneity of Lyme disease risk. Oecologia 146, 469–475. https://doi.org/10.1007/s00442-005-0251-9

Bruckner, G., Vosloo, W., Cloete, M., Dungu, B., Du Plessis, B., 2004. Foot-and-mouth disease control using vaccination: South African experience. [Review] [30 refs]. Dev. Biol. (Basel).

Bruckner, G.K., Vosloo, W., du Plessis, B.J.A., Kloek, P.E.L.G., Connoway, F.J., Ekron, M.D., Weaver, D.B., Dickason, C.J., Schreuder, F.J., Marais, T., Mogajane, M.E., 2002. Foot and mouth disease: the experience of South Africa. Rev. Sci. Tech. l'OIE. https://doi.org/10.20506/rst.21.3.1368

Buhnerkempe, M.G., Tildesley, M.J., Lindström, T., Grear, D.A., Portacci, K., Miller, R.S., Lombard, J.E., Werkman, M., Keeling, M.J., Wennergren, U., Webb, C.T., Lindstrom, T., Grear, D.A., Portacci, K., Miller, R.S., Lombard, J.E., Werkman, M., Wennergren, W., 2014. The Impact of Movements and Animal Density on Continental Scale Cattle Disease Outbreaks in the United States. PLoS One 9, e91724. https://doi.org/10.1371/journal.pone.0091724

Burdett, C.L., Kraus, B.R., Garza, S.J., Miller, R.S., 2015. Simulating the Distribution of Individual Livestock Farms and Their Populations in the United States : An Example Using Domestic Swine (Sus scrofa domesticus) Farms. PLoS One 10, 1–21. https://doi.org/10.5061/dryad.46pb3

Burgman, M., 2015. Use Experts Wisely. Nature 526, 317–318.

Calba, C., Antoine-Moussiaux, N., Charrier, F., Hendrikx, P., Saegerman, C., Peyre, M., Goutard, F.L., 2015. Applying participatory approaches in the evaluation of surveillance systems: a pilot study on swine fevers in Corsica. Prev. Vet. Med. 122, 389–398.

Caporale, V., Giovannini, A., di Francensco, C., Calistri, P., 2016a. Importance of the traceability of animals and animal products in epidemiology. Rev. Sci. Tech. l'OIE. https://doi.org/10.20506/rst.20.2.1279

Caporale, V., Giovannini, A., Zepeda, C., 2016b. Surveillance strategies for foot and mouth disease to prove absence of disease and absence of viral circulation. Rev. Sci. Tech. l'OIE. https://doi.org/10.20506/rst.31.3.2156

Carlin, B.P., Chib, S., 1995. Bayesian Model Choice via Markov Chain Monte Carlo Methods. J. R. Stat. Soc. Ser. B 57, 473–484. https://doi.org/10.2307/2346151

Carpenter, T.E., O'Brien, J.M., Hagerman, A.D., McCarl, B. a, 2011. Epidemic and Economic Impacts of Delayed Detection of Foot-And-Mouth Disease: A Case Study of a Simulated Outbreak in California. J. Vet. Diagnostic Investig. 23, 26–33. https://doi.org/10.1177/104063871102300104

Carpenter, T.E., Thurmond, M.C., Bates, T.W., 2004. A simulation model of intraherd transmission of foot and mouth disease with reference to disease spread before and after clinical diagnosis. J. Vet. Diagn. Invest. 16, 11–16. https://doi.org/10.1177/104063870401600103

Catley, A., Alders, R.G., Wood, J.L.N., 2012. Participatory epidemiology: Approaches, methods, experiences. Vet. J. https://doi.org/10.1016/j.tvjl.2011.03.010

Cheneau, Y., El Idrissi, A.H., Ward, D., 2004. An assessment of the strengths and weaknesses of current veterinary systems in the developing world. Rev. Sci. Tech.

Chib, S., Greenberg, E., 1995. Understanding the Metropolis-Hastings Algorithm. Am. Stat. 49, 327–335. https://doi.org/10.2307/2684568

Cleland, P.C., Chamnanpood, P., Baldock, F.C., Gleeson, L.J., 1995. An investigation of 11 outbreaks of foot-and-mouth disease in villages in northern Thailand. Prev. Vet. Med. https://doi.org/10.1016/0167-5877(94)00416-G

Cleland, P.C., Chamnanpood, P., Baldock, F.C., Gleeson, L.J., 1995. Questionnaire survey of foot and mouth disease (FMD) and of FMD control by vaccination in villages in northern Thailand. Rev. Sci. Tech. https://doi.org/10.2460/ajvr.69.2.240

Cochran, W.G., 1954. Some Methods for Strengthening the Common χ 2 Tests. Biometrics. https://doi.org/10.2307/3001616

Cocks, P., Abila, R., Bouchot, A., Benigno, C., Morzaria, S., Inthavong, P., 2009. FAO ADB and OIE

SEAFMD Study on Cross - Border movement and market chains of large ruminants and pigs in the Greater Mekong Sub - Region Table of Contents.

Cohen, J., 1960. A Coefficient of Agreement for Nominal Scales. Educ. Psychol. Meas. 20, 37–46.

Cox, L.A., Popken, D.A., VanSickle, J.J., Sahu, R., 2005. Optimal tracking and testing of U.S. and Canadian herds for BSE: A value-of-information (VOI) approach. Risk Anal. 25, 827–840. https://doi.org/10.1111/j.1539-6924.2005.00648.x

Cutler, D.R., 2007. Random Forests for Classification in Ecology. Ecology 88, 2783–2792. https://doi.org/10.1890/07-0539.1

DasGupta, A., Cai, T.T., Brown, L.D., 2001. Interval Estimation for a Binomial Proportion. Stat. Sci. https://doi.org/10.1214/ss/1009213286

Davies, G., 2002. The foot and mouth disease (FMD) epidemic in the United Kingdom 2001. Comp. Immunol. Microbiol. Infect. Dis. https://doi.org/10.1016/S0147-9571(02)00030-9

De Gourville, E., Duintjer Tebbens, R.J., Sangrujee, N., Pallansch, M.A., Thompson, K.M., 2006. Global surveillance and the value of information: The case of the global polio laboratory network. Risk Anal. 26, 1557–1569. https://doi.org/10.1111/j.1539-6924.2006.00845.x

de Rueda, C.B., de Jong, M.C., Eble, P.L., Dekker, A., 2014. Estimation of the transmission of foot-and-mouth disease virus from infected sheep to cattle. Vet Res 45, 58. https://doi.org/10.1186/1297-9716-45-58

Déglise, C., Suggs, S.L., Odermatt, P., 2012. SMS for disease control in developing countries: A systematic review of mobile health applications. J. Telemed. Telecare. https://doi.org/10.1258/jtt.2012.110810

Diggle, P., Ribeiro, P.J., 1998. Model-based Geostatistics. J. R. Stat. Soc. Ser. C (Applied Stat. 47, 299–350. https://doi.org/10.1007/978-0-387-48536-2

Disney, W.T., Peters, M.A., 2003. Simulation modeling to derive the value-of-information for risky animal disease-import decisions. Prev. Vet. Med. 61, 171–184. https://doi.org/10.1016/j.prevetmed.2003.07.001

Doel, T.R., 2016. Natural and vaccine-induced immunity to foot and mouth disease: the prospects for improved vaccines. Rev. Sci. Tech. l’OIE. https://doi.org/10.20506/rst.15.3.955

Dohoo, I., Martin, W., Stryhn, H., 2009. Veterinary Epidemiologic Research. VER Inc., Charlottetown, Prince Edward Island, Canada.

Donaldson, A.I., Alexandersen, S., Sorensen, J.H., Mikkelsen, T., 2001. Relative risks of the uncotrollable (airborne) spread of FMD by different species. Vet. Rec. 148, 602–604.

Dubé, C., Ribble, C., Kelton, D., McNab, B., 2008. Comparing Network Analysis Measures to Determine Potential Epidemic Size of Highly Contagious Exotic Diseases in Fragmented Monthly Networks of Dairy Cattle Movements in Ontario, Canada. Transbound. Emerg. Dis. 55, 382–392.

Dubé, C., Stevenson, M.A., Garner, M.G., Sanson, R.L., Corso, B.A., Harvey, N., Griffin, J., Wilesmith, J.W., Estrada, C., 2007. A comparison of predictions made by three simulation models of foot-and-mouth disease. N. Z. Vet. J. 55, 280–288. https://doi.org/10.1080/00480169.2007.36782 ER

Eames, K.T.D., Keeling, M.J., 2003. Contact tracing and disease control. Procedings R. Soc. 270, 2565–2571. https://doi.org/10.1098/rspb.2003.2554

Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Townsend Peterson, A., J. Phillips, S., Richardson, K., Scachetti-Pereira, R., E. Schapire, R., Soberón, J., Williams, S., S. Wisz, M., E. Zimmermann, N., 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography (Cop.). 29, 129–151. https://doi.org/10.1111/j.2006.0906-7590.04596.x

Enright, J., Kao, R.R., 2016. A descriptive analysis of the growth of unrecorded interactions amongst cattle-raising premises in Scotland and their implications for disease spread. BMC Vet. Res. https://doi.org/10.1186/s12917-016-0652-5

ESRI, 2018. ArcGIS Desktop: Release 10.6.

Ferguson, N.M., Donnelly, C.A., Anderson, R.M., 2001. Transmission intensity and impact of control policies on the foot and mouth epidemic in Great Britain. Nature 413, 542–548.

Fleiss, J.L., 1971. Measuring nominal scale agreement among many raters. Psychol. Bull. 76, 378–382. https://doi.org/10.1037/h0031619

Flood, J.S., Porphyre, T., Tildesley, M.J., Woolhouse, M.E.J., 2013. The performance of approximations of farm contiguity compared to contiguity defined using detailed geographical information in two sample areas in Scotland: implications for foot-and-mouth disease modelling. BMC Vet. Res. 9, 198. https://doi.org/10.1186/1746-6148-9-198

Flory, G.A., Peer, R.W., Clark, R.A., Baccar, M.N., Le, T.-T., Mbarek, A. Ben, Farsi, S., 2017. Aboveground burial for managing catastrophic losses of livestock. Int. J. One Heal. 3, 50–56. https://doi.org/10.14202/IJOH.2017.50-56

Forbes, R., van Halderen, A., 2014. Foot-and-Mouth Disease Economic Impact Assessment: What it means for New Zealand. Wellington, New Zealand.

Forman, S., Planté, C., Murray, G., Rey, B., Belton, D., Evans, B., 2012. Position paper : Improving governance for effective Veterinary Services in developing countries – a priority for donor funding Why invest in good governance of Veterinary Services ? Rev. Sci. Tech.

Forsberg, O.I., Guttormsen, A.G., 2006. the Value of Information in Salmon Farming. Harvesting the Right Fish At the Right Time. Aquac. Econ. Manag. 7305. https://doi.org/10.1080/13657300600985215

Frössling, J., Nusinovici, S., Nöremark, M., Widgren, S., Lindberg, A., 2014. A novel method to identify herds with an increased probability of disease introduction due to animal trade. Prev. Vet. Med. Spec. Issue SVEPM 2014 - Support. Decis. Mak. Anim. Heal. through Adv. Multidiscip. Methodol. 2014 Soc. Vet. Epidemiol. Prev. Med. Conf. 117, 367–374. https://doi.org/http://dx.doi.org/10.1016/j.prevetmed.2014.07.013

Frössling, J., Ohlson, A., Björkman, C., Håkansson, N., Nöremark, M., 2012. Application of network analysis parameters in risk-based surveillance - Examples based on cattle trade data and bovine infections in Sweden. Prev. Vet. Med. https://doi.org/10.1016/j.prevetmed.2011.12.011

Fukai, K., Morioka, K., Onozato, H., Yoshida, K., Sakamoto, K., 2013. Comparative evaluation of three commercial ELISA kits for detection of antibodies to a nonstructural protein of foot-and-mouth disease virus. J. Vet. Med. Sci. https://doi.org/DN/JST.JSTAGE/jvms/12-0430 [pii]

Garner, M.G., East, I.J., Kompas, T., Ha, P. V., Roche, S.E., Nguyen, H.T.M., 2016. Comparison of alternatives to passive surveillance to detect foot and mouth disease incursions in Victoria, Australia. Prev. Vet. Med. https://doi.org/10.1016/j.prevetmed.2016.04.009

Garner, M.G., East, I.J., Stevenson, M.A., Sanson, R.L., Rawdon, T.G., Bradhurst, R.A., Roche, S.E., Van Ha, P., Kompas, T., 2016. Early Decision Indicators for Foot-and-Mouth Disease Outbreaks in Non-Endemic Countries. Front. Vet. Sci. 3. https://doi.org/10.3389/fvets.2016.00109

Gates, M.C., Humphry, R.W., Gunn, G.J., Woolhouse, M.E.J., 2014. Not all cows are epidemiologically equal: quantifying the risks of bovine viral diarrhoea virus (BVDV) transmission through cattle movements. Vet. Res. 45, 110. https://doi.org/10.1186/s13567-014-0110-y

Gates, M.C., Volkova, V. V., Woolhouse, M.E.J., 2013a. Impact of changes in cattle movement regulations on the risks of bovine tuberculosis for Scottish farms. Prev. Vet. Med. https://doi.org/10.1016/j.prevetmed.2012.07.016

Gates, M.C., Volkova, V. V., Woolhouse, M.E.J., 2013b. Risk factors for bovine tuberculosis in low incidence regions related to the movements of cattle. BMC Vet. Res. https://doi.org/10.1186/1746-6148-9-225

Gates, M.C., Volkova, V. V, Woolhouse, M.E.J., 2013. Impact of changes in cattle movement regulations on the risks of bovine tuberculosis for Scottish farms. Prev. Vet. Med. 108, 125–136. https://doi.org/http://dx.doi.org/10.1016/j.prevetmed.2012.07.016

Gates, M.C., Woolhouse, M.E.J., 2015. Controlling infectious disease through the targeted manipulation of contact network structure. Epidemics. https://doi.org/10.1016/j.epidem.2015.02.008

Gates, M.C., Woolhouse, M.E.J., Gunn, G.J., Humphry, R.W., 2013c. Relative associations of cattle movements, local spread, and biosecurity with bovine viral diarrhoea virus (BVDV) seropositivity in beef and dairy herds. Prev. Vet. Med. https://doi.org/10.1016/j.prevetmed.2013.07.017

Gates, M.C., Woolhouse, M.E.J., Gunn, G.J., Humphry, R.W., 2013d. Relative associations of cattle movements, local spread, and biosecurity with bovine viral diarrhoea virus (BVDV) seropositivity in beef and dairy herds. Prev. Vet. Med. 112, 285–295. https://doi.org/10.1016/j.prevetmed.2013.07.017

Geering, W. a, Lubroth, J., 2002. Prepartation of Foot-and-mouth disease contingency plans, FAO Animal Health Manual. FAO, Rome.

Gibbens, J.C., Sharpe, C.E., Wilesmith, J.W., Mansley, L.M., Michalopoulou, E., Ryan, J.B.M., Hudson, M., 2001. Descriptive epidemiology of the 2001 foot-and-mouth disease epidemic in Great Britain : the first five months. Vet. Rec. 149, 729–743.

Gleeson, L.J., 2002a. A review of the status of foot and mouth disease in South-East Asia and approaches to control and eradication. Rev. sci. tech. Off. int. Epiz. 21, 465–475. https://doi.org/10.2460/ajvr.69.2.240

Gleeson, L.J., 2002b. A review of the status of Foot-and-mouth disease in South-East Asia and approaches to control and eradication of Foot-and-mouth disease in South-East Asia. Rev. Sci. Tech. https://doi.org/10.2460/ajvr.69.2.240

Gopal, R., Goodchild, A., Hewinson, G., de la Rua Domenech, R., Clifton-Hadley, R., 2006. Introduction of bovine tuberculosis to north-east England by bought-in cattle. Vet. Rec. https://doi.org/10.1136/vr.159.9.265

Goutard, F.L., Binot, A., Duboz, R., Rasamoelina-Andriamanivo, H., Pedrono, M., Holl, D., Peyre, M.I., Cappelle, J., Chevalier, V., Figuié, M., Molia, S., Roger, F.L., 2015. How to reach the poor? Surveillance in low-income countries, lessons from experiences in Cambodia and Madagascar. Prev. Vet. Med. 120, 12–26. https://doi.org/10.1016/j.prevetmed.2015.02.014

Green, D.M., Kao, R.R., 2007. Data quality of the Cattle Tracing System in Great Britain. Vet. Rec. https://doi.org/10.1136/vr.161.13.439

Green, P.J., 1995. Reversible jump Markov chain monte carlo computation and Bayesian model determination. Biometrika 82, 711–732. https://doi.org/10.1093/biomet/82.4.711

Grubman, M., Baxt, B., 2004. Foot-and-mouth disease. Clin. Microbiol. Rev. 17, 465–493. https://doi.org/10.1128/CMR.17.2.465

Gwyther, C.L., Williams, A.P., Golyshin, P.N., Edwards-Jones, G., Jones, D.L., 2011. The environmental and biosecurity characteristics of livestock carcass disposal methods: A review. Waste Manag. 31, 767–778. https://doi.org/http://dx.doi.org/10.1016/j.wasman.2010.12.005

Halasa, T., Willeberg, P., Christiansen, L.E., Boklund, A., AlKhamis, M., Perez, a, Enøe, C., 2013. Decisions on control of foot-and-mouth disease informed using model predictions. Prev. Vet. Med. 112, 194–202. https://doi.org/http://dx.doi.org/10.1016/j.prevetmed.2013.09.003

Hammant, C.A., de Silva, E., Umehara, O., 1984. The effect of maternally derived antibodies on the response of calves to vaccination against foot and mouth disease. J. Hyg. (Lond). https://doi.org/10.1017/S0022172400064081

Hamoonga, R., Stevenson, M.A., Allepuz, A., Carpenter, T.E., Sinkala, Y., 2014. Risk factors for foot-and-mouth disease in Zambia, 1981-2012. Prev. Vet. Med. 114, 64–71. https://doi.org/10.1016/j.prevetmed.2014.01.014

Hansen, G.J. a., Jones, M.L., 2008. The Value of Information in Fishery Management. Fisheries 33, 340–348. https://doi.org/10.1577/1548-8446-33.7.340

Harvey, N., Reeves, A., Schoenbaum, M.A., Zagmutt-Vergara, F.J., Dub??, C., Hill, A.E., Corso, B.A., McNab, W.B., Cartwright, C.I., Salman, M.D., 2007. The North American Animal Disease Spread Model: A simulation model to assist decision making in evaluating animal disease incursions. Prev. Vet. Med. 82, 176–197. https://doi.org/10.1016/j.prevetmed.2007.05.019

Hayama, Y., Muroga, N., Nishida, T., Kobayashi, S., Tsutsui, T., 2012. Risk factors for local spread of foot-and-mouth disease, 2010 epidemic in Japan. Res. Vet. Sci. https://doi.org/10.1016/j.rvsc.2011.09.001

Hayama, Y., Osada, Y., Oshiki, D., Tsutsui, T., 2017. An economic assessment of foot and mouth disease in Japan. Rev. Sci. Tech. l'OIE. https://doi.org/10.20506/rst.36.1.2622

Hayama, Y., Yamamoto, T., Kobayashi, S., Muroga, N., Tsutsui, T., 2016. Potential impact of species and livestock density on the epidemic size and effectiveness of control measures for foot-and-mouth disease in Japan. J. Vet. Med. Sci. 78. https://doi.org/10.1292/jvms.15-0224

Hayama, Y., Yamamoto, T., Kobayashi, S., Muroga, N., Tsutsui, T., 2013. Mathematical model of the 2010 foot-and-mouth disease epidemic in Japan and evaluation of control measures. Prev. Vet. Med. 112. https://doi.org/10.1016/j.prevetmed.2013.08.010

Haydon, D.T., Chase-Topping, M., Shaw, D.J., Matthews, L., Friar, J.K., Wilesmith, J., Woolhouse, M.E.J., 2003. The construction and analysis of epidemic trees with reference to the 2001 UK foot-and-mouth outbreak. Proc. Biol. Sci. 270, 121–127. https://doi.org/10.1098/rspb.2002.2191

Hess, G.D., Garner, M.G., Yang, X., 2008. A Sensitivity Analysis of an Integrated Modelling Approach to Assess the Risk of Wind-borne Spread of Foot-and-mouth Disease Virus from Infected Premises. Environ. Model. Assess. 13, 209–220. https://doi.org/10.1007/s10666-007-9097-3

Hibi, J., Kurosawa, A., Watanabe, T., Kadowaki, H., Watari, M., Makita, K., 2015. Post-traumatic stress disorder in participants of foot-and-mouth disease epidemic control in Miyazaki, Japan, in 2010. J. Vet. Med. Sci. https://doi.org/10.1292/jvms.14-0512

Hidano, A., Enticott, G., Christley, R.M., Gates, M.C., 2018. Modeling Dynamic Human Behavioral Changes in Animal Disease Models: Challenges and Opportunities for Addressing Bias. Front. Vet. Sci. https://doi.org/10.3389/fvets.2018.00137

Hidano, A., Gates, M.C., 2019. Why sold, not culled? Analysing farm and animal characteristics associated with livestock selling practices. Prev. Vet. Med.

https://doi.org/10.1016/j.prevetmed.2019.03.005

Hijmans, R.J., 2015. raster: Geographic Data Analysis and Modeling R.

Hodges, J.L., Lehmann, E.L., 2007. Estimates of Location Based on Rank Tests. Ann. Math. Stat. https://doi.org/10.1214/aoms/1177704172

Holliman, A., 2005. Differential diagnosis of diseases causing oral lesions in cattle. In Pract. https://doi.org/10.1136/inpract.27.1.2

Hollings, T., Burgman, M., van Andel, M., Gilbert, M., Robinson, T., Robinson, A., 2018. How do you find the green sheep? A critical review of the use of remotely sensed imagery to detect and count animals. Methods Ecol. Evol. https://doi.org/10.1111/2041-210X.12973

Hollings, T., Robinson, A., van Andel, M., Jewell, C., Burgman, M., 2017. Species distribution models: A comparison of statistical approaches for livestock and disease epidemics. PLoS One 12, e0183626. https://doi.org/10.1371/journal.pone.0183626

Honhold, N., Taylor, N.M., 2006. Data quality assessment: comparison of recorded and contemporary data for farm premises and stock numbers in Cumbria, 2001, Society for Veterinary Epidemiology and Preventive Medicine. Proceedings of a meeting held at Exeter, UK, 29-31 March 2006.

Honhold, N., Taylor, N.M., Wingfield, A., Einshoj, P., Middlemiss, C., Eppink, L., Wroth, R., Mansley, L.M., 2004. Evaluation of the application of veterinary judgement in the pre-emptive cull of contiguous premises during the epidemic of foot-and-mouth disease in Cumbria in 2001. Vet. Rec. https://doi.org/10.1136/vr.155.12.349

Hosmer, D.W., Lemeshow, S., May, S., 2008. Applied Survival Analysis. Regression Modeling of Time-to-Event Data, Technometrics. https://doi.org/10.2307/1270580

Hseu, Z.Y., Chen, Z.S., 2017. Experiences of mass pig carcass disposal related to groundwater quality monitoring in Taiwan. Sustain. https://doi.org/10.3390/su9010046

Hugh-Jones, M.E., 1972. Epidemiological studies on the 1967-1968 Foot and Mouth Epidemic: Attack Rates and Cattle Density. Res. Vet. Sci. 13, 411–417.

Hui, S.L., Walter, S.D., 1980. Estimating the Error Rates of Diagnostic Tests. Biometrics. https://doi.org/10.2307/2530508

Jamal, S.M., Belsham, G.J., 2013. Foot-and-mouth disease: Past, present and future. Vet. Res.

https://doi.org/10.1186/1297-9716-44-116

Jewell, C., Brown, R., 2014. Bayesian data assimilation provides rapid decision support for vector borne diseases Supporting Information Preparation of NAIT movement network 1–6.

Jewell, C.P., 2016a. Bayesian ZIP model with covariates on the zero probability v1. https://doi.org/http://fhm-chicas-code.lancs.ac.uk/jewell/zipBayes

Jewell, C.P., 2016b. Variogram v1. https://doi.org/http://fhm-chicas-code.lancs.ac.uk/jewell/GPUVariogram.git

Jewell, C.P., van Andel, M., Vink, W.D., Mcfadden, A.M.J., 2015. Compatibility between livestock databases used for quantitative biosecurity response in New Zealand. N. Z. Vet. J. 0169. https://doi.org/10.1080/00480169.2015.1117955

Joung, H.K., Han, S.H., Park, S.-J., Jheong, W.-H., Ahn, T.S., Lee, J.-B., Jeong, Y.-S., Jang, K.L., Lee, G.-C., Rhee, O.-J., Park, J.-W., Paik, S.Y., 2013. Nationwide Surveillance for Pathogenic Microorganisms in Groundwater near Carcass Burials Constructed in South Korea in 2010. Int. J. Environ. Res. Public Health 10, 7126–7143. https://doi.org/10.3390/ijerph10127126

Kahn, S., Geale, D.W., Kitching, P.R., Bouffard, A., Allard, D.G., Duncan, J.R., 2002. Vaccination against foot-and-mouth disease: The implications for Canada. Can. Vet. J. 43, 349–354.

Keeling, M.J., Woolhouse, M.E.J., Shaw, D.J., Matthews, L., Chase-Topping, M., Haydon, D.T., Cornell, S.J., Kappey, J., Wilesmith, J., Grenfell, B.T., 2001. Dynamics of the 2001 UK Foot and Mouth Epidemic: Stochastic Dispersal in a Heterogeneous Landscape. Science (80-.). 294, 813–817. https://doi.org/10.1126/science.1065973

Keeling, M.J., Woolhouse, M.E.J., Shaw, D.J., Matthews, L., Haydon, D.T., Cornell, S.J., Kappey, J., Wilesmith, J., Bryan, T., Chase-topping, M., Grenfell, B.T., 2009. Landscape in a Heterogeneous Dynamics of and Mouth Epidemic : Stochastic Dispersal Landscape.

Keisler, J.M., Collier, Z.A., Chu, E., Sinatra, N., Linkov, I., 2014. Value of information analysis: The state of application. Environ. Syst. Decis. 34, 3–23. https://doi.org/10.1007/s10669-013-9439-4

Kim, H. su, Kim, K., 2012. Microbial and chemical contamination of groundwater around livestock mortality burial sites in Korea - a review. Geosci. J. https://doi.org/10.1007/s12303-012-0036-1

Kiss IZ Kao RR, G.D.M., 2006. The network of sheep movements within Great Britian: network properties and their implications for infectious disease spread. J. R. Soc. 3, 669–677.

Kitching, R.P., 2002. Clinical variation in foot and mouth disease: cattle. Rev. Sci. Tech. L Off. Int. Des Epizoot. 21, 513–518.

Kitching, R.P., Hughes, G.J., 2016. Clinical variation in foot and mouth disease: sheep and goats. Rev. Sci. Tech. l'OIE. https://doi.org/10.20506/rst.21.3.1342

Kitching, R.P., Thrusfield, M., Taylor, N.M., 2006. Use and abuse of mathematical models : an illustration from the 2001 foot and mouth. Rev. Sci. Tech. 25, 293–311.

Klain, T.J., Lusk, J.L., Tonsor, G.T., Schroeder, T.C., 2014. An experimental approach to valuing information. Agric. Econ. (United Kingdom) 45, 635–648. https://doi.org/10.1111/agec.12112

Knight-Jones, T.J.D., Gubbins, S., Bulut, A.N., Stärk, K.D.C., Pfeiffer, D.U., Sumption, K.J., Paton, D.J., 2016. Mass vaccination, immunity and coverage: Modelling population protection against foot-and-mouth disease in Turkish cattle. Sci. Rep. https://doi.org/10.1038/srep22121

Knight-jones, T.J.D., Rushton, J., 2013. The economic impacts of foot and mouth disease – What are they , how big are they and where do they occur ? Prev. Vet. Med. 112, 161–173. https://doi.org/10.1016/j.prevetmed.2013.07.013

Kompas, T., Ha, P. Van, Nguyen, H.T.M., East, I., Roche, S., Garner, G., 2017. Optimal surveillance against foot-and-mouth disease: the case of bulk milk testing in Australia. Aust. J. Agric. Resour. Econ. 61, 515–538. https://doi.org/10.1111/1467-8489.12224

Kostoulas, P., Nielsen, S.S., Branscum, A.J., Johnson, W.O., Dendukuri, N., Dhand, N.K., Toft, N., Gardner, I.A., 2017. STARD-BLCM: Standards for the Reporting of Diagnostic accuracy studies that use Bayesian Latent Class Models. Prev. Vet. Med. https://doi.org/10.1016/j.prevetmed.2017.01.006

Kyaw, A.T., 2014. Socio-economic impacts of Foot and Mouth Disease among cattle farmers in Sagaing and Mandalay areas, Myanmar.

Kyaw, N.O., 2010. Epidemiological study to support the establishment of a progressive zoning approach for the control of Foot and Mouth Disease in Myanmar. J. Chem. Inf. Model. Murdoch University. https://doi.org/10.1017/CBO9781107415324.004

Lambert, D., 1992. Zero-Inflated Poisson Regression, With an Application To Defects in Manufacturing. Technometrics 34, 1–14.

Landis, J., Koch, G., 1977. The Measurement of Observer Agreement for Categorical Data. Biometrics

33, 159–174.

Langford, I.H., Leyland, A.H., Rasbash, J., Goldstein, H., 2009. Multilevel Modelling of the Geographical Distributions of Diseases. Appl. Stat. 48, 253–268.

Lee, M., Wagenmakers, E., 2010. A course in Bayesian graphical modeling for cognitive science. … manuscript. Retrieved from … 1–188.

Leslie, D., Black, L., 2006. Tourism and the impact of the foot and mouth epidemic in the UK: Reactions, responses and realities with particular reference to Scotland. J. Travel Tour. Mark. https://doi.org/10.1300/J073v19n02_04

Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest. R News 2, 18–22.

Longjam, N., Deb, R., Sarmah, A.K., Tayo, T., Awachat, V.B., Saxena, V.K., 2011. A brief review on diagnosis of foot-and-mouth disease of livestock: Conventional to molecular tools. Vet. Med. Int. https://doi.org/10.4061/2011/905768

Lunn, D.J., Thomas, A., Best, N., Spiegelhalter, D., 2000. WinBUGS - A Bayesian modelling framework: Concepts, structure, and extensibility. Stat. Comput. https://doi.org/10.1023/A:1008929526011

Madder, M., Walker, J.G., Van Rooyen, J., Knobel, D., Vandamme, E., Berkvens, D., Vanwambeke, S.O., De Clercq, E.M., 2012. E-Surveillance in Animal Health: Use and evaluation of mobile tools. Parasitology. https://doi.org/10.1017/S0031182012000571

Madin, B., 2011. An evaluation of Foot-and-Mouth Disease outbreak reporting in mainland South-East Asia from 2000 to 2010. Prev. Vet. Med. https://doi.org/10.1016/j.prevetmed.2011.07.010

Mansley, L.M., 2004. The challenge of FMD control in the 2001 UK FMD epidemic, in: Madec, F. (Ed.), Animal Production in Europe : The Way Forward in a Changing World “in-between” Congress of the International Society for Animal Hygiene; Saint-Malo, France . pp. 345–350.

Mansley, L.M., Donaldson, A.I., Thrusfield, M. V., Honhold, N., 2016. Destructive tension: mathematics versus experience – the progress and control of the 2001 foot and mouth disease epidemic in Great Britain. Rev. Sci. Tech. l’OIE. https://doi.org/10.20506/rst.30.2.2054

Mardones, F.O., zu Donha, H., Thunes, C., Velez, V., Carpenter, T.E., 2013. The value of animal movement tracing: A case study simulating the spread and control of foot-and-mouth disease in California. Prev. Vet. Med. 110, 133–138.

https://doi.org/http://dx.doi.org/10.1016/j.prevetmed.2012.11.024

Mariner, J.C., Hendrickx, S., Pfeiffer, D.U., Costard, S., Knopf, L., Okuthe, S., Chibeu, D., Parmley, J., Musenero, M., Pisang, C., Zingeser, J., Jones, B. a, Ali, S.N., Bett, B., McLaws, M., Unger, F., Araba, a, Mehta, P., Jost, C.C., 2011. Integration of participatory approaches into surveillance systems. Rev. Sci. Tech. Int. Off. Epizoot. 30, 653–659.

McDonald, W.L., Rawdon, T.G., Fitzmaurice, J., Bolotovski, I., Voges, H., Humphrey, S., Fernando, K., Canagasebey, Y., Thornton, R.N., McIntyre, L., 2009. Survey of bulk tank milk in new zealand for mycoplasma bovis, using species-specific nested PCR and culture. N. Z. Vet. J. https://doi.org/10.1080/00480169.2009.36867

McFadden, D., 1974. Conditional logit analysis of qualitative choice behavior, in: In:Zarembka, P. (Ed.), Frontiers in Econometrics. https://doi.org/10.1108/eb028592

Mcleod, R., 2010. Realised and Potential Economic Benefits of the Southeast Asia Foot and Mouth Disease Campaign.

Mitchell, P.D., 2003. Value of imperfect input information in agricultural production. Agric. Syst. 75, 277–294. https://doi.org/10.1016/S0308-521X(02)00070-7

Moore, J.L., Runge, M.C., 2012. Combining Structured Decision Making and Value-of-Information Analyses to Identify Robust Management Strategies. Conserv. Biol. 26, 810–820. https://doi.org/10.1111/j.1523-1739.2012.01907.x

Morgan, K.L., Handel, I.G., Tanya, V.N., Hamman, S.M., Nfon, C., Bergman, I.E., Malirat, V., Sorensen, K.J., Bronsvoort, B.M. d. C., 2014. Accuracy of herdsmen reporting versus serologic testing for estimating foot-and-mouth disease prevalence. Emerg. Infect. Dis. https://doi.org/10.3201/eid2012.140931

Morgan, K.L., Handel, I.G., Tanya, V.N., Hamman, S.M., Nfon, C., Bergman, I.E., Malirat, V., Sorensen, K.J., Bronsvoort, B.M.D.C., 2014. Accuracy of Herdsmen Reporting versus Serologic Testing for Estimating Foot-and-Mouth Disease Prevalence. Emerg. Infect. Dis. 20, 1–7.

Muellner, P., Muellner, U., Gates, M.C., Pearce, T., Ahlstrom, C., O'Neill, D., Brodbelt, D., Cave, N.J., 2016. Evidence in Practice – A Pilot Study Leveraging Companion Animal and Equine Health Data from Primary Care Veterinary Clinics in New Zealand. Front. Vet. Sci. 3. https://doi.org/10.3389/fvets.2016.00116

Mullahy, J., 1986. Specification and testing of some modified count data models. J. Econom. 33, 341–

365. https://doi.org/http://dx.doi.org/10.1016/0304-4076(86)90002-3

Muroga, N., Hayama, Y., Yamamoto, T., Kurogi, A., Tsuda, T., Tsutsui, T., 2012. The 2010 Foot-and-Mouth Disease Epidemic in Japan. J. Vet. Med. Sci. https://doi.org/10.1292/jvms.11-0271

Muroga, N., Kobayashi, S., Nishida, T., Hayama, Y., Kawano, T., Yamamoto, T., 2013. Risk factors for the transmission of foot-and-mouth disease during the 2010 outbreak in Japan : a case – control study. BMC Vet. Res. 9, 1–9.

Nampanya, S., Khounsy, S., Abila, R., Young, J.R., Bush, R.D., Windsor, P.A., 2016. Financial Impacts of Foot-and-Mouth Disease at Village and National Levels in Lao PDR. Transbound. Emerg. Dis. 63, e403–e411. https://doi.org/10.1111/tbed.12319

Nampanya, S., Khounsy, S., Phonvisay, A., Young, J.R., Bush, R.D., Windsor, P.A., 2015. Financial Impact of Foot and Mouth Disease on Large Ruminant Smallholder Farmers in the Greater Mekong Subregion. Transbound. Emerg. Dis. 62, 555–564. https://doi.org/10.1111/tbed.12183

Nampanya, S., Richards, J., Khounsy, S., Inthavong, P., Yang, M., Rast, L., Windsor, P.A., 2013. Investigation of foot and mouth disease hotspots in northern lao PDR. Transbound. Emerg. Dis. 60, 315–329. https://doi.org/10.1111/j.1865-1682.2012.01350.x

Nicolas, G., Robinson, T.P., Wint, G.R.W., Conchedda, G., Cinardi, G., Gilbert, M., 2016. Using Random Forest to improve the downscaling of global livestock census data. PLoS One 11, 1–16. https://doi.org/10.1371/journal.pone.0150424

Ntzoufras, I., 2011. Bayesian Modeling Using WinBUGS. Comput. Stat. 698, 592. https://doi.org/10.1002/9780470434567

O'Toole, R., Matthews, A., Mulvey, M., 2002. Impact of the 2001 Foot and Mouth Outbreak on the Irish Economy, The IMAGE Project Working Paper, Trinity College Dublin.

OIE, FAO, 2012. The Global Foot and Mouth Disease Control Strategy: Strengthening animal health sysyems through improved control of major diseases. https://doi.org/978-92-5-107273-8

OIE World Organization for Animal Health, 2016. A strategic framework to control, prevent and eradicate foot and mouth disease in South-East Asia and China.

Otte, M., Nugent, R., McLeod, A., 2004. Transboundary Animal Diseases : Assessment of socio-economic impacts, Fao.

Owen, K., Stevenson, M.A., Sanson, R.L., 2011. A sensitivity analysis of the New Zealand standard model of foot and mouth disease. Rev. Sci. Tech. - Off. Int. des Epizoot. 30, 513–526.

Paiba, G.A., Roberts, S.R., Houston, C.W., Williams, E.C., Smith, L.H., Gibbens, J.C., Holdship, S., Lysons, R., 2007. UK surveillance: Provision of quality assured information from combined datasets. Prev. Vet. Med. Lead Pap. Elev. Symp. Int. Soc. Vet. Epidemiol. Econ. (ISVEE), Cairns, Aust. 81, 117–134. https://doi.org/http://dx.doi.org/10.1016/j.prevetmed.2007.04.006

Paine, S., Mercer, G.N., Kelly, P.M., Bandaranayake, D., Baker, M.G., Huang, Q.S., Mackereth, G., 2010. Transmissibility of 2009 pandemic influenza A (H1N1) in New Zealand : effective reproduction number and influence of age , ethnicity and importations 1–9.

Park, J.H., Lee, K.N., Ko, Y.J., Kim, S.M., Lee, H.S., Shin, Y.K., Sohn, H.J., Park, J.Y., Yeh, J.Y., Lee, Y.H., Kim, M.J., Joo, Y.S., Yoon, H., Yoon, S.S., Cho, I.S., Kim, B., 2013. Control of foot-and-mouth disease during 2010-2011 epidemic, South Korea. Emerg. Infect. Dis. https://doi.org/10.3201/eid1904.121320

Paton, D.J., F??ssel, A.E., Vosloo, W., Dekker, A., De Clercq, K., 2014. The use of serosurveys following emergency vaccination, to recover the status of “foot-and-mouth disease free where vaccination is not practised.” Vaccine 32, 7050–7056. https://doi.org/10.1016/j.vaccine.2014.10.064

Paton, D.J., Sumption, K.J., Charleston, B., 2009. Options for control of foot-and-mouth disease: Knowledge, capability and policy. Philos. Trans. R. Soc. B Biol. Sci. 364, 2657–2667. https://doi.org/10.1098/rstb.2009.0100

Pebesma, E., Bivand, R., Rowlingson, B., Gomez-Rubio, V., Hijmans, R., Sumner, M., MacQueen, D., Lemon, J., O’Brien, J., 2014. sp: Classes and Methods for Spatial Data.

Perez, A.M., Ward, M.P., Carpenter, T.E., 2004. Control of a foot-and-mouth disease epidemic in Argentina. Prev. Vet. Med. https://doi.org/10.1016/j.prevetmed.2004.08.002

Perry, B., Randolph, T., McDermott, J., Thornton, KSones, P., 2002. Investing in animal health research to alleviate poverty, ILRI (International Livestock Research Institute), Nairobi, Kenya. https://doi.org/10.1017/CBO9781107415324.004

Perry, B.D., Kalpravidh, W., Coleman, P.G., Horst, H.S., McDermott, J.J., Randolph, T.F., Gleeson, L.J., 1999. The economic impact of foot and mouth disease and its control in South-East Asia: a preliminary assessment with special reference to Thailand. Rev Sci Tech 18, 478–497.

Pharo, H.J., 2002. Foot-and-mouth disease: an assessment of the risks facing New Zealand. N. Z. Vet.

J. 50, 46–55. https://doi.org/10.1080/00480169.2002.36250

Pharo, H.J., 2002. Foot -and-mouth disease: An assessment of the risks facing New Zealand. N. Z. Vet. J. https://doi.org/10.1080/00480169.2002.36250

Pink, B., 2011. New Zealand: An Urban/Rural Profile http://www.stats.govt.nz/browse_for_stats/Maps_and_geography/Geographic-areas/urban-rural-profile.aspx. Statistics New Zealand, Wellington, New Zealand.

Porphyre, T., Auty, H.K., Tildesley, M.J., Gunn, G.J., Woolhouse, M.E.J., 2013. Vaccination against Foot-And-Mouth Disease: Do Initial COnditions Affect Its Benefit? PLoS One 8.

Porphyre, T., Auty, H.K., Tildesley, M.J., Gunn, G.J., Woolhouse, M.E.J., 2013a. Vaccination against Foot-And-Mouth Disease: Do Initial Conditions Affect Its Benefit? PLoS One 8. https://doi.org/10.1371/journal.pone.0077616

Porphyre, T., Auty, H.K., Tildesley, M.J., Gunn, G.J., Woolhouse, M.E.J., 2013b. Vaccination against Foot-And-Mouth Disease: Do Initial Conditions Affect Its Benefit? PLoS One 8, e77616. https://doi.org/10.1371/journal.pone.0077616

Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer classification and regression tree techniques: Bagging and random forests for ecological prediction. Ecosystems 9, 181–199. https://doi.org/10.1007/s10021-005-0054-1

Probert, W.J.M., Shea, K., Fonnesbeck, C.J., Runge, M.C., Carpenter, T.E., Dürr, S., Garner, M.G., Harvey, N., Stevenson, M.A., Webb, C.T., Werkman, M., Tildesley, M.J., Ferrari, M.J., 2015. Decision-making for foot-and-mouth disease control: Objectives matter. Epidemics 15, 1–11. https://doi.org/10.1016/j.epidem.2015.11.002

R Core Team, 2014. R: A language and environment for statistical computing.

Rawdon, T., 2018. Epidemiology investigations: New Zealand Mycoplasma bovis response 2017-2018, in: Proceedings of the Food Safety, Animal Welfare & Biosecurity, Epidemiology & Animal Health Management, and Industry Branches of the NZVA, 2018 EPI Proceedings. p. pp 25-26.

Reiczigel, J., Földi, J., Ózsvári, L., 2010. Exact confidence limits for prevalence of a disease with an imperfect diagnostic test. Epidemiol. Infect. https://doi.org/10.1017/S0950268810000385

Reid, S.M., Ferris, N.P., Brüning, A., Hutchings, G.H., Kowalska, Z., Åkerblom, L., 2001. Development of a rapid chromatographic strip test for the pen-side detection of foot-and-mouth disease virus antigen. J. Virol. Methods. https://doi.org/10.1016/S0166-0934(01)00334-2

Riley, S., 2010. Coping without Farm Location Data during a Foot-and-Mouth Outbreak. Proc. Natl. Acad. Sci. U. S. A. 107, 957–958. https://doi.org/10.2307/40535583

Rindskopf, D., Rindskopf, W., 1986. The value of latent class analysis in medical diagnosis. Stat Med.

Risk solutions, 2005. Cost Benefit Analysis of Foot and Mouth Disease Controls, Defra report.

Robertson, C., Sawford, K., Daniel, S.L.A., Nelson, T.A., Stephen, C., 2010. Mobile phone-based infectious disease surveillance system, Sri Lanka. Emerg. Infect. Dis. 16, 1524–1531. https://doi.org/10.3201/eid1610.100249

Robinson, T.P., Wint, G.R.W., Conchedda, G., Van Boeckel, T.P., Ercoli, V., Palamara, E., Cinardi, G., D'Aietti, L., Hay, S.I., Gilbert, M., 2014. Mapping the Global Distribution of Livestock. PLoS One 9, e96084.

Robinson, T.P., Wint, G.R.W., Conchedda, G., Van Boeckel, T.P., Ercoli, V., Palamara, E., Cinardi, G., D'Aietti, L., Hay, S.I., Gilbert, M., Beard, K.H., Elith, J., H. Graham, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., Li, J., G. Lohmann, L., A. Loiselle, B., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., McC. M. Overton, J., Townsend Peterson, A., J. Phillips, S., Richardson, K., Scachetti-Pereira, R., E. Schapire, R., Soberón, J., Williams, S., S. Wisz, M., E. Zimmermann, N., Thuiller, W., Elith, J., Leathwick, J.R., Hastie, T., Araújo, M.B., Guisan, A., 2006. Mapping the Global Distribution of Livestock. PLoS One 9, e96084. https://doi.org/10.1111/j.1365-2656.2008.01390.x

Roche, S.E., Garner, M.G., Sanson, R.L., Cook, C., Birch, C., Backer, J.A., Dubé, C., Patyk, K.A., Stevenson, M.A., Yu, Z.D., Rawdon, T.G., Gauntlett, A.F., 2015. Evaluating vaccination strategies to control foot-and-mouth disease: a model comparison study. Epidemiol. Infect 143, 1256–1275. https://doi.org/10.1017/S0950268814001927

Roche, S.E., Garner, M.G., Wicks, R.M., East, I.J., de Witte, K., 2014. How do resources influence control measures during a simulated outbreak of foot and mouth disease in Australia? Prev. Vet. Med. 113. https://doi.org/10.1016/j.prevetmed.2013.12.003

Rogan, W.J., Gladen, B., 1978. Estimating prevalence from the results of a screening test. Am. J. Epidemiol. https://doi.org/10.1093/oxfordjournals.aje.a112510

Rweyemamu, M., Roeder, P., MacKay, D., Sumption, K., Brownlie, J., Leforban, Y., 2008. Planning for the progressive control of foot-and-mouth disease worldwide. Transbound. Emerg. Dis.

https://doi.org/10.1111/j.1865-1682.2007.01016.x

Sanson, R., Pearson, A., 1997. Agribase - a national spatial farm database, in: Eighth IVSEE, Paris. p. 12.12.1-12.16.3.

Sanson, R., Rawdon, T., Owen, K., Hickey, K., Van Andel, M., Yu, Z., 2017. Evaluating the benefits of vaccination when used in combination with stamping-out measures against hypothetical introductions of foot-and- mouth disease into New Zealand: a simulation study. N. Z. Vet. J. 65, 124–133. https://doi.org/10.1080/00480169.2016.1263165

Sanson, R., Rawdon, T., Owen, K., Yu, Z., 2013. Evaluating the benefits of vaccination for foot-and-mouth disease as an adjunct to a standard stamping out policy.

Sanson, R.L., 2005. A survey to investigate movements off sheep and cattle farms in New Zealand, with reference to the potential transmission of foot-and-mouth disease. N. Z. Vet. J. 53, 223–233. https://doi.org/10.1080/00480169.2005.36550

Sanson, R.L., 1994. The epidemiology of foot-and-mouth disease: implications for New Zealand. N. Z. Vet. J. 42, 41–53. https://doi.org/10.1080/00480169.1994.35785 ER

Sanson, R.L., Harvey, N., Garner, M.G., Stevenson, M.A., 2011. Foot and mouth disease model verification and ‘relative validation.’ Rev. Sci. Tech. L`Office Int. Des Epizoot. 30, 527–540.

Sarandopoulos, J., 2015. Early Predictors of the Size and Duration of Foot-and-Mouth Disease Epidemics. The University of Melbourne.

Savill, N.J., Shaw, D.J., Deardon, R., Tildesley, M.J., Keeling, M.J., Woolhouse, M.E.J., Brooks, S.P., Grenfell, B.T., 2007. Effect of data quality on estimates of farm infectiousness trends in the UK 2001 foot-and-mouth disease epidemic. J. R. Soc. Interface 4, 235–241.

Sawford, K.E., 2011. Animal health surveillance for early detection of emerging infectious disease risks. ProQuest Diss. Theses.

Schärrer, S., Widgren, S., Schwermer, H., Lindberg, A., Vidondo, B., Zinsstag, J., Reist, M., 2015. Evaluation of farm-level parameters derived from animal movements for use in risk-based surveillance programmes of cattle in Switzerland. BMC Vet. Res. 11. https://doi.org/10.1186/s12917-015-0468-8

Sergeant, E., 2019. Epitools Epidemiological Calculators [WWW Document]. Ausvet Pty Ltd. https://doi.org/Ausvet Pty Ltd.

Shea, K., Tildesley, M.J., Runge, M.C., Fonnesbeck, C.J., Ferrari, M.J., 2014. Adaptive Management and the Value of Information: Learning Via Intervention in Epidemiology. PLoS Biol. 12, 9–12. https://doi.org/10.1371/journal.pbio.1001970

Smith, P., Luthi, N.B., Li, H., Kyaw, N.O., Phonvisay, A., Premashthira, S., Abila, R., Kukreja, K., Miller, C., Widders, P., 2012. Movement pathways and market chains of large ruminants in the Greater Mekong Sub-region.

Sorensen, J.H., Mackay, D.K., Jensen, C.O., Donaldson, A.I., 2000. An integrated model to predict the atmospheric spread of foot-and-mouth disease virus. Epidemiol Infect 124, 577–590.

Sørensen, K.J., Madsen, K.G., Madsen, E.S., Salt, J.S., Nqindi, J., Mackay, D.K.J., 1998. Differentiation of infection from vaccination in foot-and-mouth disease by the detection of antibodies to the non-structural proteins 3D, 3AB and 3ABC in ELISA using antigens expressed in baculovirus. Arch. Virol. 143, 1461–1476. https://doi.org/10.1007/s007050050390

Stage, A.R., 1976. An expression of the effects of aspect, slope, and habitat type on tree growth. For. Sci. 22, 457–460.

Stage, A.R., Salas, C., 2007. Interactions of elevation, aspect, and slope in models of forest species composition and productivity. For. Sci. 53, 486–492.

Stevenson, M.A., Sanson, R.L., Miranda, A.O., Lawrence, K.A., Morris, R.S., 2007. Decision support systems for monitoring and maintaining health in food animal populations. N. Z. Vet. J. 55, 264–272. https://doi.org/10.1080/00480169.2007.36780

Stevenson, M.A., Sanson, R.L., Stern, M.W., O'Leary, B.D., Sujau, M., Moles-Benfell, N., Morris, R.S., Sanson, A.L., Stern, M.W., O'Leary, B.D., Sujau, M., Moles-Benfell, M., Morris, R.S., 2013. InterSpread Plus: a spatial and stochastic simulation model of disease in animal populations. Prev. Vet. Med. 109, 10–24. https://doi.org/http://dx.doi.org/10.1016/j.prevetmed.2012.08.015

Stringer, L.A., Jones, G., Jewell, C.P., Noble, A.D., Heuer, C., Wilson, P.R., Johnson, W.O., 2013. Bayesian estimation of the sensitivity and specificity of individual fecal culture and Paralisa to detect Mycobacterium avium subspecies paratuberculosis infection in young farmed deer. J. Vet. Diagnostic Investig. 25, 759–764. https://doi.org/10.1177/1040638713505587

Strobl, C., Boulesteix, A.L., Zeileis, A., Hothorn, T., 2007. Bias in random forest variable importance measures: Illustrations, sources and a solution. BMC Bioinformatics 8.

https://doi.org/10.1186/1471-2105-8-25

Sumption, K., Rweyemamu, M., Wint, W., 2008. Incidence and distribution of foot-and-mouth disease in Asia, Africa and South America; Combining expert opinion, official disease information and livestock populations to assist risk assessment. Transbound. Emerg. Dis. https://doi.org/10.1111/j.1865-1682.2007.01017.x

Sumption, K.J., McLaws, M., Bartels, C.J.M., De Leeuw, P., Domenech, J., Lubroth, J., Ferrari, G., 2012. The Progressive Control Pathway for FMD (PCP-FMD): a tool for developing sustainable long-term national and regional FMD control., in: Paper Presented at the Proceedings of The FAO/OIE Global Conference on Foot and Mouth Disease Control.

Sutmoller, P., Barteling, S.S., Olascoaga, R.C., Sumption, K.J., 2003. Control and eradication of foot-and-mouth disease. Virus Res. 91, 101–144.

Tana, T., 2014. The MPI Animal General Surveillance programme.

Team, R.C., 2017. R: A language and environment for statistical computing.

Therneau, T., 2015. A Package for Survival Analysis in S.

Therneau, T., Atkinson, B., Ripley, B., 2015. rpart: Recursive Partitioning and Regression Trees.

Therneau, T., Grambsch, P.M., 2000. Modeling Survival Data: Extending the Cox Model. Springer, New York.

Thrusfield, M., 2013. Veterinary Epidemiology [electronic resource]. EBL.

Thrusfield, M., 2007. Veterinary Epidemiology, Third Edit. ed. Blackwell Science, Oxford, UK.

Thrusfield, M., Mansley, L.M., Dunlop, P., Taylor, J., Pawson, A., Stringer, L., 2005. The foot-and-mouth disease epidemic in Dumfries and Galloway, 2001. 1: Characteristics and control. Vet. Rec. 156, 229–252.

Tildesley, M.J., Deardon, R., Savill, N.J., Bessell, P.R., Brooks, S.P., Woolhouse, M.E.J., Grenfell, B.T., Keeling, M.J., 2008. Accuracy of models for the 2001 foot-and-mouth epidemic. Proc. Biol. Sci. 275, 1459–1468. https://doi.org/10.1098/rspb.2008.0006

Tildesley, M.J., House, T.A., Bruhn, M.C., Curry, R.J., O’Neil, M., Allpress, J.L.E., Smith, G., Keeling, M.J., 2010. Impact of spatial clustering on disease transmission and optimal control. Proc. Natl. Acad. Sci. 107, 1041–1046. https://doi.org/10.1073/pnas.0909047107

Tildesley, M.J., Ryan, S.J., 2012. Disease Prevention versus Data Privacy: Using Landcover Maps to Inform Spatial Epidemic Models. PLoS Comput. Biol. 8, 1–13. https://doi.org/10.1371/journal.pcbi.1002723

Tildesley, M.J., Smith, G., Keeling, M.J., 2012. Modeling the spread and control of foot-and-mouth disease in Pennsylvania following its discovery and options for control. Prev. Vet. Med. https://doi.org/10.1016/j.prevetmed.2011.11.007

Tomassen, F.H.M., de Koeijer, A., Mourits, M.C.M., Dekker, A., Bouma, A., Huirne, R.B.M., 2002. A decision-tree to optimise control measures during the early stage of a foot-and-mouth disease epidemic. Prev. Vet. Med. 54, 301–324. https://doi.org/http://dx.doi.org/10.1016/S0167-5877(02)00053-3

Truong, D.B., Goutard, F.L., Bertagnoli, S., Delabouglise, A., Grosbois, V., Peyre, M., 2018. Benefit-cost analysis of foot-and-mouth disease vaccination at the farm-level in South Vietnam. Front. Vet. Sci. 5, 1–11. https://doi.org/10.3389/fvets.2018.00026

Van Andel, M., Hollings, T., Bradhurst, R., Robinson, A., Burgman, M., Gates, M.C., Bingham, P., Carpenter, T., 2018. Does Size Matter to Models? Exploring the Effect of Herd Size on Outputs of a Herd-Level Disease Spread Simulator. Front. Vet. Sci. https://doi.org/10.3389/fvets.2018.00078

van Andel, M., Hollings, T., Bradhurst, R.A., Robinson, A., Burgman, M., Gates, M.C., Bingham, P., Carpenter, T.E., 2018. Does Size Matter to Models ? Exploring the Effect of Herd Size on Outputs of a Herd-Level Disease Spread Simulator. Front. Vet. Sci. 5. https://doi.org/10.3389/fvets.2018.00078

van Andel, M., Jewell, C., McKenzie, J., Hollings, T., Robinson, A., Burgman, M., Bingham, P., Carpenter, T., 2017. Predicting farm-level animal populations using environmental and socioeconomic variables. Prev. Vet. Med. 145, 121–132. https://doi.org/10.1016/j.prevetmed.2017.07.005

van Segbroeck, S., Santos, F.C., Pacheco, J.M., 2010. Adaptive contact networks change effective disease infectiousness and dynamics. PLoS Comput. Biol. https://doi.org/10.1371/journal.pcbi.1000895

Vergne, T., Grosbois, V., Durand, B., Goutard, F., Bellet, C., Holl, D., Roger, F., Dufour, B., 2012. A capture-recapture analysis in a challenging environment: Assessing the epidemiological situation of foot-and-mouth disease in Cambodia. Prev. Vet. Med. 105, 235–243.

https://doi.org/10.1016/j.prevetmed.2011.12.008

Vinet, L., Zhedanov, A., 2010. A “missing” family of classical orthogonal polynomials. Classif. Regres. Trees. https://doi.org/10.1088/1751-8113/44/8/085201

VoseSoftware, 2017. ModelRisk 10.3.

Wada, M., Carpenter, T.E., van Andel, M., 2016a. The importance of accurate spatial livestock data to inform decision making: a case study of foot-and-mouth disease control in New Zealand, in: Geovet. Valdivia, Chile.

Wada, M., Stevenson, M., Cogger, N., Carpenter, T., 2016b. Evaluation of the Control Strategy for the 2010 Foot-and-Mouth Disease Outbreak in Japan Using Disease Simulation. Transbound. Emerg. Dis. doi:10.111, 1–12. https://doi.org/10.1111/tbed.12467

Ward, M.P., Highfield, L.D., Garner, M.G., 2009. Use of Simulation Modelling to Estimate the Effective Reproductive number for outbreaks of Foot-and-Mouth disease. Unpublished 2–4.

Watson, P., 2004. Differential diagnosis of oral lesions and FMD in sheep. In Pract. https://doi.org/10.1136/inpract.26.4.182

Webb, C.T., Ferrari, M., Lindström, T., Carpenter, T., Dürr, S., Garner, G., Jewell, C., Stevenson, M., Ward, M.P., Werkman, M., Backer, J., Tildesley, M., 2017. Ensemble modelling and structured decision-making to support Emergency Disease Management. Prev. Vet. Med. 138, 124–133. https://doi.org/10.1016/j.prevetmed.2017.01.003

Werkman, M., Tildesley, M.J., Brooks-Pollock, E., Keeling, M.J., 2016. Preserving Privacy Whilst Maintaining Robust Epidemiological Predictions. Epidemics. https://doi.org/10.1016/j.epidem.2016.10.004

Williams, B.K., Eaton, M.J., Breininger, D.R., 2011. Adaptive resource management and the value of information. Ecol. Modell. 222, 3429–3436. https://doi.org/10.1016/j.ecolmodel.2011.07.003

Williams, C., Ferguson, M., 2005. Recovering from crisis. Strategic alternatives for leisure and tourism providers based within a rural economy. Int. J. Public Sect. Manag. https://doi.org/10.1108/09513550510599265

Wilson, P., Kinsella, L., 2004. The impact of foot and mouth disease on the price of beef. EuroChoices. https://doi.org/10.1111/j.1746-692X.2004.tb00031.x

Woolhouse, M.E.J., 2003. Foot-and-mouth disease in the UK: What should we do next time? Jounral

Appl. Microbiol. 94, 126–130.

Yang, P., W. and Sung, H.Y.P.C.R.C., 1999. Epidemiological characteristics and financial costs of the 1997 foot-and-mouth disease epidemic in Taiwan. Vet. Rec. 145, 731–734.

Yokota, F., Thompson, K.M., 2004. Value of information literature analysis: a review of applications in health risk management. Med Decis Mak. 24, 287–298. https://doi.org/10.1177/0272989x04263157

Young, J.R., Suon, S., Andrews, C.J., Henry, L.A., Windsor, P.A., 2013. Assessment of Financial Impact of Foot and Mouth Disease on Smallholder Cattle Farmers in Southern Cambodia. Transbound. Emerg. Dis. 60, 166–174. https://doi.org/10.1111/j.1865-1682.2012.01330.x

Zweig, M.H., Campbell, G., 1993. Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. Clin. Chem. 39, 561–577. https://doi.org/ROC; Receiver-Operating Characteristic; SDT; Signal Detection Theory