Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

The use of habitat analysis in the control of wildlife tuberculosis in New Zealand

A thesis presented in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy at Massey University

Joanna McKenzie

1999

Printed by Massey University Printery, 1999

To obtain a copy of this thesis write to:

EpiCentre
Wool Building
Massey University
Palmerston North
New Zealand

Ph: +64 (0)6 350 5270
Fx: +64 (0)6 350 5716

E-mail: J.S.McKenzie@massey.ac.nz

Abstract

The brushtail possum (*Trichosurus vulpecula*) acts as a wildlife reservoir of *Mycobacterium bovis* infection for farmed livestock and other wildlife species in areas covering an estimated 23 percent of New Zealand. This reservoir of infection is severely hampering efforts to eradicate TB from farmed animals in New Zealand. New strategies need to be developed that focus on identifying areas of residual TB infection in possums and eradicating them, while maintaining low broad-scale population levels in the most cost-effective way. This thesis comprises a number of studies that apply multivariate modelling techniques and spatial analytical procedures to evaluate spatial patterns of possum habitat and of TB occurrence in cattle herds, and hence to identify high-risk localities for possum-derived TB and target control strategies of varying intensity where each will be most cost-effective.

An initial case-control study of 132 possum capture sites was used to identify habitat and topographic factors at the scale of the possum home range that were predictors of the probability of capturing tuberculous possums. Data on the following factors were collected from a 50 meter square area at each capture site: abundance of plant species and structural features of vegetation, the abundance and quality of available den sites, and topographic factors such as slope and aspect. Data were analysed using unconditional logistic regression modelling. Slope of the land and factors associated with the quality of dens were associated with the probability of capturing multiple TB possums. Slope of the land, height of trees, percent cover of vegetation in the height range 0.3 – 2 metres and percent of the ground covered in vascular vegetation were associated with the probability of capturing single TB possums.

A subsequent case control study was conducted at the larger scale of farms to identify possum habitat, geographic and farm factors associated with the probability of tuberculous possums being present on a farm, using random effects logistic regression modelling. A geographic information system (GIS) was used to combine and analyse digital geographic data sets to produce variables for the statistical analysis. Geographic data used for the study comprised farm boundaries, rivers, slope, and vegetation coverage which was derived from satellite imagery. Fragstats, a habitat pattern analysis programme, was used to produce variables representing patterns of possum habitat on farms. Variables were included in the model to account for spatial autocorrelation patterns in the data and a variogram analysis was conducted on residuals to test for any significant spatial autocorrelation that was not

accounted for in the model. Distance from the coastal forest, total area of the farm, area of pine forest, area of manuka/pasture and proportion of the farm covered with podocarp/broadleaved species were the significant variables in the final model.

An automated classification procedure was used to generate a vegetation map from a SPOT3 multi-spectral satellite image for this study. The resulting raster map comprised 8 vegetation classes, including 3 tree classes, 4 scrub classes and a pasture class. This map had a high degree of spatial accuracy and more detailed vegetation classes than existing digital data available in New Zealand. However, some categories contained a heterogeneous range of plant species which limited the discriminatory power of models using this vegetation data.

Spatial clustering of percent TB-positive years, five-year cumulative incidence and space-time clustering of annual cumulative incidence of TB in cattle were analysed for beef breeding farms in the area in which the farm-level case control study was conducted. Analyses used the spatial and space-time scan statistics which are available in SaTScan, a cluster analysis software package. Mapping of farms included in the resulting clusters enabled us to develop hypotheses on the association of different habitat patterns with the resulting disease patterns in cattle. This study also highlighted the value of these cluster analysis techniques as tools in surveillance for wildlife TB, by using farm-level patterns of TB as an indicator of the possible presence of wildlife TB.

The final stage of the project described in this thesis was the development of EpiMAN-TB, a spatial decision support system for the management of wildlife-related TB in New Zealand. Development of this tool was undertaken to assist the application of the research results described in this thesis at the field level. The system combined a relational database, map display tools, spatial analysis tools, simple expert systems and simulation models of the spread of TB in possums at the scale of individual farms and administrative regions. The software comprises five modules set within a user-friendly interface. The modules are: possum TB hot spot prediction, farm TB risk prediction, evaluation of possum control programmes at the farm and the regional level, and an administration module.

Acknowledgements

"*Humans need adventure – projects with an uncertain outcome. We have a drive to examine, explore, push the limits of whatever we find ourselves involved in, be it music, physics, art, sport, mathematics or mountaineering.*"

Graeme Dingle (NZ adventurer), New Zealand Geographic, January-March 1996

Starting out on a PhD journey at the same time as starting the journey of "mother of two" was certainly an adventure with an uncertain outcome. Indeed if we knew the outcome of many of the things that we undertook in life we wouldn't begin them at all, depriving ourselves of an opportunity to grow beyond boundaries we didn't know we had. This has been some journey and it has stretched my boundaries in all ways from the intellectual to the physical to the emotional and the spiritual. I drew strength from and received support from many many people through the course of the past eight years, and I would like to acknowledge these people here.

It is customary to leave the acknowledgement of the support of one's partner in life and/or family to the end. However, in recognition of the enormity of the support involved in helping a mother of young children through a PhD I would like to begin with my husband, Pin. I believe that a good relationship allows each partner to achieve their potential and Pin has done this for me with my PhD. My deepest and most sincere thanks go to Pin for his continued support and for providing an enriching home environment for our kids when I wasn't there. Our two children, Tom and Bridget have only known Mum as doing a PhD (or P.D. as Bridget once called it!). I thank them for being tolerant (most of the time) and for the joy that they show in my finishing. A special thanks to my parents, Don and Ali, who have been very supportive particularly in providing a second home to our kids during the school holidays. Special thanks to my sister and brother-in-law, Fi and Dean, who provided me with a peaceful retreat in the country where I wrote a large part of my thesis. Thanks to all our friends and extended family who have supported us through this long and arduous period, really wondering "why's she doing this?!"

Roger Morris, my chief supervisor, and Dirk Pfeiffer, my secondary supervisor, combine a range of skills and experience, vision and commitment that are unequalled anywhere in the world. I am particularly grateful to Roger for providing me with the opportunity to do a PhD, for broadening the breadth and depth of my understanding of epidemiology, for his personal support, and for teaching me that I do have a limit to the number of things that I can commit myself to. To Dirk, I am particularly grateful for his personal support and friendship and for his creative insight into analytical epidemiology. Thanks also to my third supervisor Terry Ryan for his understanding of what it's like doing a PhD.

I'd like to make a special thanks to Andrea who helped me with my field work. Andrea's sense of fun and commitment to her work kept me going through many a tedious day, when the days were cold and wet, the trees dripping. But we both shared in the beauty of lunch on a hillside watching Mt Ruapehu erupting, and discussed plans for the future. Thanks to Chris Tutty who brought EpiMAN-TB into being and has helped enlighten me in what 'computer programmers do'. Thanks to Robyn O'Connor for her companionship and support during the early days. To Fiona whose 'telephone voice' I've always enjoyed and whose artistic talent helped in the production of this thesis. To Deb whose bright cheeriness has picked me up many times over recent months.

I have received tremendous support and companionship from the EpiCentre group as a whole and particularly those with whom I've shared an office over the years – Dirk, Robert, Ron, Tata, Barb, Leigh, Katharina, Carola and Naomi. Thank you.

I would like to acknowledge the considerable help I have received from several members of AgriQuality NZ (formerly MAF Quality Management) based in Masterton, Palmerston North, Taumarunui and Turangi who provided me with data and field information, in particular Gillian Atkinson and Garth Pannett. I have also received considerable help from the pest control staff in the Wellington Regional Council.

Finally, I would like to acknowledge the support of the AHB who funded this project.

Joanna McKenzie 21 May 1999
EpiCentre, Institute of Veterinary and Animal Biosciences,
Massey University, Palmerston North

Table of Contents

ABSTRACT I

ACKNOWLEDGEMENTS III

TABLE OF CONTENTS V

LIST OF FIGURES IX

LIST OF TABLES XI

INTRODUCTION 1

CHAPTER 1 SPATIAL ANALYSIS OF HEALTH DATA: A LITERATURE REVIEW 9

INTRODUCTION 11

GEOGRAPHIC INFORMATION SYSTEMS 12

SPATIAL DATA SETS 14

DATA VISUALISATION 16

STATISTICAL DATA ANALYSIS 18

Exploratory methods *20*

Statistical modelling *25*

GIS AS A SIMULATION MODELLING TOOL 30

USE OF GIS IN DECISION SUPPORT SYSTEMS 31

CONCLUSION 32

BIBLIOGRAPHY 33

CHAPTER 2 STATISTICAL MODELLING OF HABITAT ASSOCIATED WITH TUBERCULOUS POSSUMS IN NEW ZEALAND 41

ABSTRACT 43

INTRODUCTION 43

MATERIALS AND METHODS 49

Study population *49*

Independent variables *50*

Statistical analysis *55*

RESULTS 58

DISCUSSION 62

BIBLIOGRAPHY 70

CHAPTER 3 SPATIAL AND TEMPORAL PATTERNS OF VECTOR-BORNE TUBERCULOSIS INFECTION IN BEEF BREEDING CATTLE 77

ABSTRACT 79

INTRODUCTION 79

MATERIALS AND METHODS 81

Study area *81*

Cattle TB data *82*

Farm boundary map *82*

Study population *83*

Data availability and quality *83*

Measures of frequency of occurrence of cattle TB infection 85
Statistical analysis 86
RESULTS 89
DISCUSSION 93
BIBLIOGRAPHY 97
CHAPTER 4 AUTOMATIC CLASSIFICATION OF A MULTISPECTRAL SPOT3 IMAGE TO PRODUCE A VEGETATION MAP FOR MODELLING POSSUM HABITAT IN NEW ZEALAND 101
ABSTRACT 103
INTRODUCTION 103
EXISTING SOURCES OF VEGETATION DATA 105
Black and white aerial photographs 105
New Zealand Land Resource Inventory (NZLRI) 106
The vegetative cover map of New Zealand 106
Ecological survey of New Zealand indigenous forests 107
Protected Natural Areas Programme 107
New Zealand Land Cover Data Base 107
Satellite imagery 108
METHOD 109
Study area 109
Image analysis 109
RESULTS 112
Pine 113
Beech 113
Podocarp/broadleaf 113
Manuka/kanuka and Manuka/pasture 114
Shrubland 116
Manuka/gorse 116
DISCUSSION 117
ACKNOWLEDGEMENTS 120
BIBLIOGRAPHY 120
CHAPTER 5 APPLICATION OF GIS AND SPATIAL ANALYTICAL TECHNIQUES TO STATISTICAL MODELLING OF POSSUM-ASSOCIATED TB RISK ON FARMS 125
INTRODUCTION 127
MATERIALS AND METHODS 129
Study area 129
Geographical data 130
Cattle TB data 131
Study population 131
Outcome variable 132
Independent variables 132
Statistical analysis 136
Analysis of residuals 138

RESULTS 140

General farm-level model 140

Farm-level habitat pattern model 141

Vegetation class-level habitat pattern models 143

Summary vegetation class-level landscape model 146

Vegetation-slope model 147

Summary model 148

Analysis of residuals 150

DISCUSSION 150

BIBLIOGRAPHY 156

APPENDIX I 162

CHAPTER 6 EPIMAN-TB, A SPATIAL DECISION SUPPORT SYSTEM FOR THE MANAGEMENT OF WILDLIFE-RELATED TUBERCULOSIS IN CATTLE AND DEER IN NEW ZEALAND 165

ABSTRACT 167

INTRODUCTION 167

DECISION SUPPORT SYSTEMS 170

OVERVIEW OF EPIMAN-TB 172

DESCRIPTION OF THE SYSTEM 173

Spatial data 175

Non-spatial data 176

FUNCTIONS 176

TB hot spot analysis module 177

On-farm possum control module 182

Regional TB spread module 184

Farm risk predictor module 190

Administration module 190

BIBLIOGRAPHY 192

GENERAL DISCUSSION 197

HABITAT PREDICTORS OF POSSUM TB RISK AT THE POSSUM HOME RANGE LEVEL 199

APPLICATION OF GIS TO TB MANAGEMENT AND RESEARCH 200

Vegetation coverage 200

Farm boundary data 201

HABITAT AND GEOGRAPHICAL PREDICTORS OF POSSUM TB RISK AT THE FARM LEVEL 202

SPATIAL PATTERNS OF TB AT THE FARM LEVEL 203

EPIMAN(TB) 204

POSSUM DENSITY 206

IMPLICATIONS FOR CONTROL 206

BIBLIOGRAPHY 207

APPENDIX 209

APPENDIX 1: FORMS FOR RECORDING POSSUM HABITAT DATA 211

List of Figures

Figure 1. Map of the North Island of New Zealand showing general locations of the study sites. 54

Figure 2. Histograms showing the percent of sites within 5-degree slope categories for each TB status group. 60

Figure 3. Histograms of the percent sites within 5-metre mean top height categories for each TB status group. 60

Figure 4. Histograms of the percent sites within 10 percent abundance categories of vegetation cover in tier 5 for each TB status group. 60

Figure 5 . Histograms showing the percent of sites within categories for the abundance score of quality 1 dens. 61

Figure 6. Histogram showing the percent of total dens of each quality at each location 62

Figure 7. ROC curves for the "multiple TB-positive" logistic regression model that included the abundance of dens and the "single TB-positive" model, showing the sensitivity and specificity with which each model predicts the TB status of a site using different probability values as cut-off points 63

Figure 8. Map of the North Island of New Zealand showing the location of the study area. 80

Figure 9. Distribution of the 129 beef breeding study farms. 84

Figure 10. Distribution of point data used to represent the location of the study farms. Circle indicates an area where smaller farm size leads to a higher density of points 84

Figure 11. Distribution of the five-year cumulative incidence of TB-positive cattle (1986-90). 88

Figure 12. Location of the most likely high and low-rate clusters for five-year CI (1986-90), together with the relative risk of TB-positive cattle on farms within versus without each cluster 88

Figure 13. Distribution of percent TB-positive years. 90

Figure 14. Location of the most likely high and low-rate clusters for percent TB-positive years. 90

Figure 15. Location of all significant space-time clusters in the analysis with a maximum time window of 16 years (90%) and a maximum spatial window of 65 farms (50%), together with the time period during which there was significant clustering. 92

Figure 16. Location of multiple-farm space-time clusters for the analysis with a maximum time window of 2 years, together with the time period during which there was significant clustering. Clusters extending across 2 years are circled. 92

Figure 17. Regional map of the North Island of New Zealand showing the location of the study area in the northern part of the Wellington Region 110

Figure 18. Map showing the location of the PNA survey sites used for checking the vegetation classification 112

Figure 19. Vegetation map produced by automatic classification of a SPOT3 multispectral image. 114

Figure 20. Regional map of the North Island of New Zealand showing the location of the study area. 128

Figure 21. Spatial distribution of percent TB-positive years for the 129 study farms. 139

Figure 22. Spatial distribution of predicted percent TB-positive years for the 129 study farms. 139

Figure 23. Spatial distribution of residual values from the final random effects logistic regression model 149

Figure 24. Results of the variogram analysis to test for autocorrelation amongst residuals. 149

Figure 25. Vegetation map of the study area with insets showing farms with a high degree of heterogeneity (top right) and a high degree of homogeneity (bottom left) 152

Figure 26. An area of heterogeneous cover with a high degree of patchiness and the interspersion of many vegetation classes. 153

Figure 27. An area of homogeneous Manuka/kanuka cover. 153

Figure 28. An overview of the structure of EpiMAN-TB. 174

Figure 29. Opening screen of EpiMAN-TB. 177

Figure 30. Functions available within the hot spot predictor module. 178

Figure 31. Screen for entering hot spot rules associated with each combination of vegetation and slope category. 179

Figure 32. Hot spot analysis screen which enables the user to view vegetation, slope and hot spot data simultaneously 180

Figure 33. Map viewer function in which all spatial data in EpiMAN-TB can be viewed. 181

Figure 34. Screen for entering the possum TB control strategies to be modelled in PossPOP. 182

Figure 35. Screen shot from PossPOP showing sub-areas defined by the user which have a different control programme compared to the surrounding area. 183

Figure 36. Screen for defining the vegetation and hot spot maps to run PossPOP. 184

Figure 37: Schematic diagram of TB-Spread. 186

Figure 38. Components of the TB-Spread module. 187

Figure 39. Screen for running TB-Spread once the initial parameters have been set. 189

List of Tables

Table 1. Details of the Braun-Blanquet scale used to score the abundance of plant species within each height tier. 51

Table 2. List of tree species located in tiers 2 and 3 that were classified as either 'good den trees' or 'poor den trees'. 52

Table 3. Description of the four categories used to score the quality of dens. 53

Table 4: Results of univariate chi-squared analyses for categorical variables. 55

Table 5. Summary statistics and results of univariate analyses using the Mann-Whitney U test for all continuous variables for which data was collected in the habitat study. 56

Table 6. Final logistic regression model for "multiple TB-positive" sites, using the group of variables in which dens were represented as the abundance of dens of each quality. 58

Table 7. Final logistic regression model for "multiple TB-positive" sites, using the group of variables in which dens were represented as the presence/absence of multiple dens. 59

Table 8. Final logistic regression model for "single TB-positive" sites. 59

Table 9: Time period, relative risk, p-value and number of farms for all significant space-time clusters with maximum space and time windows, shown in Figure 15 94

Table 10. Time period, relative risk and p-value for multiple-farm space-time clusters with maximum spatial window of 5% and time window of 12%. 94

Table 11. Description of the classes in a vegetation map that was produced by automatic classification of a SPOT3 multispectral image 115

Table 12. A cross-classification matrix showing the results of accuracy checking the classification of the seven non-pasture vegetation classes against data collected during a PNA survey at 127 locations 116

Table 13. The TB hot spot risk assigned to each vegetation-slope category. 134

Table 14. Possum density per hectare of each vegetation class. 135

Table 15. Descriptive statistics for the 129 beef breeding farms included in our study. 138

Table 16. List of variables entered into the general farm model. 140

Table 17. Significant variables in the general farm random effects logistic regression model with 141

Table 18. Variables entered into the farm-level habitat pattern model. 141

Table 19. Significant variables in the farm-level habitat pattern random effects logistic regression model with their odds ratio plus 95% confidence interval, without and with Forest Distance 143

Table 20. Variables entered into the logistic regression model for each vegetation class in the class-level habitat pattern models 144

Table 21. Significant variables in each of the class-level habitat pattern random effects logistic regression models, without and with the addition of Forest Distance, with their odds ratio plus 95% confidence limits. 145

Table 22. Significant variables in the vegetation class-level random effects logistic regression model that was built from the individual vegetation class models, without and with the addition of Forest Distance. 146

Table 23. List of variables entered into the vegetation-slope model. 147

Table 24. Significant variables in the vegetation-slope random effects logistic regression model with their odds ratio plus 95% confidence interval, without and with Forest Distance added. 148

Table 25. Significant variables in the summary random effects logistic regression model that was built from the significant variables for each of the individual models, with their odds ratio, 95% confidence limits and p-value ... 148

Table 26. Description of the vegetation patterns within each land unit vegetation category. ... 187

Table 27. Possum density per hectare of each vegetation class. ... 188