

Spatial survival analysis of infectious animal diseases

submitted by

Trevelyan John McKinley BSc.

to the University of Exeter

as a thesis for the degree of

Doctor of Philosophy in Mathematics.

March 2007

This thesis is available for Library use on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

I certify that all material in this thesis which is not my own work has been identified and that no material has previously been submitted and approved for the award of a degree by this or any other University.

Signature

Abstract

This thesis investigates the feasibility of using spatial survival modelling techniques to develop dynamic space-time predictive models of risk for infectious animal disease epidemics. Examples of diseases with potentially vast socioeconomic impacts include avian influenza, bovine tuberculosis and foot-and-mouth disease (FMD), all of which have received wide coverage in the recent media. The relatively sporadic occurrence of such large scale animal disease outbreaks makes determination of optimal control policies difficult, and policy makers must balance the relative impacts of different response strategies based on little prior information. It is in this situation that the use of mathematical and statistical modelling techniques can provide powerful insights into the future course of an infectious epidemic.

The motivating example for this thesis is the outbreak of FMD in Devon in 2001, however we are interested in developing more general techniques that can be applied to other animal diseases. Many of the models fitted to the 2001 UK FMD data set have focussed on modelling the global spread of the disease across the entire country and then using these models to assess the effects of nationwide response strategies. However it has been shown that the dynamics of the disease are not uniform across the whole of the UK and can vary significantly across different spatial regions. Of interest here is exploring whether modelling at a smaller spatial scale can provide more useful measures of risk and guide the development of more efficient control policies.

We begin by introducing some of the main epidemiological issues and concepts involved

in modelling infectious animal diseases, from the microscopic through to the farm population level. We then discuss the various mathematical modelling techniques that have been applied previously and how they relate to various biological principles discussed in the earlier chapters. We then highlight some limitations with these approaches and offer potential ways in which survival analysis techniques could be used to overcome some of these problems.

To this end we formulate a spatial survival model and fit it to the Devon data set with some naive initial covariates that fail to capture the dynamics of the disease. Some work by colleagues at the Veterinary Laboratories Agency, Weybridge (Arnold 2005), produced estimates of viral excretion rates for infected herds of different species type over time, and these form the basis for the development of a dynamic space-time varying viral load covariate that quantifies the viral load acting at any spatial location at any point in time. The novel use of this covariate as a means of censoring the data set via exposure is then introduced, though the models still fail to explain the variation in the epidemic process.

Two potential reasons for this are identified - the possible presence of non-localised infections and/or premise varying susceptibility. We then explore ways in which the survival approach can be extended to model more than one epidemic process through the use of mixture and long-term survivor models. Some simple simulations suggest that resistance to infection is the most likely cause of the poor model fits, and a series of more complex simulation experiments show that both the mixture and long-term survivor models offer various advantages over the conventional approach when resistance is present in the data set. However key to their performance is the ability to correctly capture the mixing, although in the worst case scenario they still replicate the results from the conventional model.

We also use these simulations to explore potential ways in which space-time predictions of the hazard of infection can be used as a means of targeting control policies to areas of 'high-risk' of infection. This shows the importance of ensuring that the scale of the control order matches the scale of the epidemic, and suggests possible dangers when using global

level models to derive response strategies for situations where the dynamics of the disease change at smaller spatial scales. Finally we apply these techniques to the Devon data set and offer some conclusions and future work.

Dedication

To Mum and all my family, who have given me their full and unconditional support. I love you all very much.

To Michelle, without whom I would never have had the courage to even attempt a PhD, and whose love, beauty and humour over the past seven years has provided me with more happiness than I ever thought possible.

And to Dad, whose guidance and wisdom has always been, and will always be invaluable. It gives me some comfort to know that you saw me submit this thesis, even though you are not able to see me graduate. I know that your faith in me, like my faith in you, never wavered.

Acknowledgements

I am privileged to have been surrounded by so many people who have consistently provided me with support and guidance throughout the past three and a half years. Most notably my supervisor Trevor Bailey, for trusting in me, allowing me access to vast quantities of his time and expertise, and for keeping me focussed.

To Peter Durr and Mark Arnold at VLA for providing the data and modelling work that formed the basis of our viral load; all the staff at SECaM for making my time at Exeter so enjoyable; and to EPSRC/VLA (project grant reference CASE 0305) for providing the financial support necessary to complete this work. Also to everyone at CIDC, particularly James Wood, for trusting that I would *eventually* finish my thesis.

In addition I would like to thank the many people that have been complicit in keeping me (at least partially) sane over the years. In particular Sam and Emma, and also Ralph, Tim and Dave, for indulging my morbid sense of humour and helping me to put the world to rights (aided in no small part by the erudite surroundings of the Ram). To Raj for always reminding me that there is more to life than work - though making it easy for me to forget that at least some of it has to be; and also to Idayu Mahat, who though often busy with her own work never failed to find the time to aid me in mine.

I am indebted to Graham, Luke and Bruce for their computing support, and finally to Jez, without whose expertise James' trust would have probably been short-lived.

Contents

1	Introduction	1
2	Infectious disease and FMD	7
2.1	General issues	8
2.2	Foot-and-Mouth Disease	12
3	Mathematical modelling of infectious diseases	17
3.1	Compartmental models	18
3.2	Cellular automata	23
3.3	Statistical modelling approaches	24
3.3.1	Survival analysis	29
3.4	Mathematical modelling of FMD	31
3.5	Conclusions	37
4	Survival Modelling	39

4.1	Basic ideas	41
4.2	Censoring	43
4.3	Modelling techniques	44
4.3.1	Non-parametric methods for estimating the survivor function	44
4.3.2	Semi-parametric models	46
4.3.3	Parametric methods for modelling survival data	49
4.3.4	Time-dependent covariates	52
4.4	Bayesian model fitting	53
4.4.1	Metropolis-Hastings algorithm	55
4.4.2	Gibbs sampling	56
4.4.3	Identifiability	57
4.5	Extensions to conventional survival models	57
4.5.1	Long-term survivor or cure rate models	57
4.5.2	Mixture models	58
4.5.3	Competing risks models	59
4.5.4	Multi-state models	59
4.5.5	Change point models	60
4.5.6	Random effects (frailties)	61
4.5.7	Extensions to multivariate response variables	62

4.6	Conclusions	62
5	Preliminary survival modelling of FMD	64
5.1	The 2001 Devon data set	65
5.2	Specification of basic model	69
5.3	Method of prediction	75
5.4	Preliminary covariates	78
5.5	Initial model results	78
5.6	Viral load (VL)	83
5.6.1	Infectivity functions	84
5.6.2	Viral load at a premise	87
5.7	Results of model fitted with AV as covariate	89
5.8	Exposure and censoring	92
5.9	Results for AV model fitted to data censored via exposure	96
5.10	Susceptibility to infection	99
5.11	Conclusions	102
6	Modelling resistance to infection	105
6.1	Considerations in FMD and other animal diseases	106
6.2	Candidate models	107

6.2.1	Mixture models	108
6.2.2	Long-term survivor models	113
6.3	Simple simulation study	117
6.4	Predicted survival times	121
6.5	Conclusions	124
7	Applications of mixture modelling to infectious animal disease epidemics	127
7.1	Spatial simulation study	128
7.1.1	Details of the simulation	128
7.1.2	The simulated epidemics	131
7.1.3	Model formulations and prediction	132
7.1.4	Prediction	136
7.1.5	Comparative results	138
7.2	Spatial hazard maps and simulated work	143
7.2.1	Hazard maps for simulated epidemic	147
7.2.2	Predictive uses	151
7.2.3	Targeting control policies	154
7.3	Application to real data set	158
7.4	Conclusions	161

8	Conclusions and further considerations	164
A	Fitting non-standard likelihoods in WinBUGS	178
B	Initial value generation	181
	Bibliography	183

List of Figures

4.1	Examples of exponential hazard and survivor functions	50
4.2	Examples of Weibull hazard and survivor functions	51
4.3	Possible graphical representation of competing risks model in multi-state framework	60
5.1	Spatial distribution of premises in Devon at the end of the 2001 epidemic .	70
5.2	Temporal distribution of infections in Devon during the 2001 epidemic . . .	70
5.3	Comparison map of predictions from initial models	82
5.4	Estimated scaled infectivity curves for different relative herd sizes in Devon by species type	85
5.5	Spatial maps of viral load over time with 3km effective bandwidth	90
5.6	Theoretical VL, cumulative and average cumulative VL plots over time . . .	91
5.7	Theoretical threshold and exposure based on VL	95
5.8	Spatial distribution of premises in Devon ‘censored via exposure’ at 50 days	99
6.1	Plot of failure times against infectious covariate for non-spatial simulation .	119

6.2	Plots of actual vs. predicted failure times for non-spatial simulation	125
6.3	Predictive posterior distribution comparison for arbitrary individual obtained from conventional model fitted to data set with no resistance and high censoring for non-spatial simulation	126
7.1	Spatial maps of simulated epidemics at 50 days	133
7.2	Epidemic plots for spatial simulations over time	134
7.3	Plots of actual vs. predicted failure times from spatial simulation with high levels of resistance	144
7.4	Comparative contour maps of hazards of infection in the next day, from conventional (left) and mixture (right) models fitted from day 14 of simulated epidemic	148
7.4	Comparative contour maps of hazards of infection in the next day, from conventional (left) and mixture (right) models fitted from day 14 of simulated epidemic (cont.)	149
7.4	Comparative contour maps of hazards of infection in the next day, from conventional (left) and mixture (right) models fitted from day 14 of simulated epidemic (cont.)	150
7.5	Estimated hazard over time for simulated epidemic (weeks 3-5)	151
7.6	Predictive risk maps of probability of infection in next week (left) and probability of belonging to top twenty future IPs (right)	153
7.7	Plot of simulated epidemics with varying control policies	157

List of Tables

5.1	Posterior parameter estimates from models (5.8) and (5.9) fitted to the Devon data set at 50 days	80
5.2	Predictive output over a 60 day window for models (5.8) and (5.9) fitted to Devon data set at 50 days	81
5.3	Parameter estimates for scaled infectivity functions	86
5.4	Infected herd size information for Devon in 2001	86
5.5	Posterior parameter estimates from model (5.9) with AV covariate fitted to Devon data set at 50 days	91
5.6	Predictive output over a 60 day window for model (5.9) with AV covariate fitted to Devon data set at 50 days	92
5.7	Posterior parameter estimates from model (5.9) with AV covariate fitted to Devon data set ‘censored via exposure’ at 50 days	97
5.8	Predictive output over a 60 day window for model (5.9) with AV covariate fitted to Devon data set ‘censored via exposure’ at 50 days	97
5.9	Posterior parameter estimates from models (5.9) with AV and additional susceptibility covariates fitted to the Devon data set censored via exposure at 50 days	101

5.10	Predictive output over a 60 day window for model (5.9) with AV and additional susceptibility covariates fitted to Devon data set at 50 days	101
6.1	Parameter estimates from models fitted to non-spatial simulation	122
6.2	Estimated numbers of individuals misspecified with regards resistance in non-spatial simulation	123
7.1	Summary values for simulated spatial epidemics at 50 days	132
7.2	Posterior parameter estimates from models fitted to simulated epidemics with varying resistance to infection at 43 days	140
7.2	Posterior parameter estimates from models fitted to simulated epidemics with varying resistance to infection at 43 days (cont.)	141
7.3	Posterior predicted summary values for models fitted to simulated epidemics at 43 days	142
7.4	Comparative numbers of infected and culled premises from simulated epidemics with varying control policies	156
7.5	Posterior parameter estimates from the mixture model (7.12) with regression covariate AV, and uninfected animal and species-specific densities in the mixing parameter - fitted to Devon data set ‘censored via exposure’ at 50 days	160
7.6	Predictive output over a 60 day window for mixture model (7.12) with regression covariate AV, and uninfected animal and species-specific densities in the mixing parameter - fitted to Devon data set ‘censored via exposure’ at 50 days	161