The spatial epidemiology of Foot and Mouth Disease in Great Britain

Paul R. Bessell

Submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Royal (Dick) School of Veterinary Studies
University of Edinburgh
2008

Declaration

I declare that the research described within this thesis is my own work and that the thesis is my own composition.

Paul R. Bessell
Edinburgh, 2008

Which is the more appropriate name; foot and mouth disease or arse and elbow disease?

The Now Show. BBC Radio 4. 10 August 2007

Abstract

During 2007 the UK experienced outbreaks of three notifiable exotic livestock diseases; Foot and Mouth Disease (FMD), Highly Pathogenic Avian Influenza (HPAI) and bluetongue. Large epidemics of any of these diseases would have a serious impact on animal welfare, farming, food production and the economy. In light of this, understanding holdings which are most likely to acquire and spread infection and being able to identify areas at higher risk of an epidemic is valuable when preparing for and managing an epidemic. This thesis uses a spatial epidemiological framework and the detailed disease and demographic data from the 2001 Great Britain (GB) FMD epidemic to develop static models of the risk of FMD susceptibility and transmission. These models are used to develop maps of FMD risk. These methods are then applied to the outbreak of FMD in 2007.

The inputs for this analysis comprised a set of data relating to the farms diagnosed with FMD and farms culled as part of the disease control measures. The cleaning of these data is described and data which were estimated relating to dates of infection and putative sources of infection are evaluated. The distribution of farm holdings and animals is taken from the June 2000 GB agricultural census, off-fields of farms in the agricultural census are recorded in other datasets and these have been identified and linked to census holdings.

A model of holding level susceptibility is developed using both farm level variables and measures of animal numbers in the locality of the holding as well as the distance to the nearest farm infected before the ban on animal movements (seeds). The overall fit of the model was very good with an area under the Receiver Operator Characteristic (ROC) curve of 0.91. A further model was developed to describe the risk of FMD transmission. However, due to incompleteness of transmission data, this was a model of the risk of finding a subsequent Infected Premises (IP) within 3km of an IP. Risk factors were a combination of holding level variables and locality measures as well as data relevant to the infection, such as infectious period and the species initially infected. The area under the ROC curve for this model was 0.71, which is regarded as an acceptable fit. Geographical barriers to FMD transmission were investigated using a case-control methodology, linear barriers comprising rivers and railways had a significant protective effect with respect to disease transmission (odds ratio = 0.54, 95% CIs = 0.30,0.96, p=0.038).

Modelled values for the transmission and susceptibility models were transformed to a raster surface in ESRI ArcMap for both the disease as it was seeded in the 2001 epidemic and a non-specific background risk surface independent of the distribution of seeds. A risk map generated for the outbreak of FMD in Surrey in August 2007 suggested that there was little risk of a large outbreak in Surrey. Potential disease introductions through livestock movements from Surrey into Scotland were identified and these suggested that if the disease were introduced into Scotland there was great danger of substantial local spread.

These methods described in this thesis have been used to map risk of FMD and subsequently applied to inform the risk presented by a different outbreak of FMD. The study underlines the value of detailed data both disease and demographic, for epidemic management. Similar methods could and should be applied to other infectious diseases threats of livestock such as HPAI and bluetongue.

Acknowledgements

I am grateful to my supervisors; Mark Woolhouse, Darren Shaw, Nick Savill and William Mackanness for their patience and guidance during the last four years. The BBSRC and latterly the SEERAD EPIC project and the USDA for ensuring I could eat.

Many members of Epigroup, both former and current who have assisted during my PhD in particular Sarah Cleaveland, Ian Handel, Lisa Boden, Eric Fèvre and Margo Chase-Topping but also Darryn Knobel, Shaun Abeysinghe and Louise Matthews for their help along the way with various issues.

I should thank the many people who have the dubious privilege of sharing an office with me during the past three and a half years. These people include; Sonya Gowtage-Sequeira, John Kunda, Gabriel Shirima, Jo Halliday, Harriet Auty, Tiziana Lembo, Gerardo Acosta, Grainne Long, Suzanne St Rose, Natalie Nicholls, Eleanor Gaunt, Emily Courcier, Max Burton, Katie Atkins, Kate Mitchell, Claire Bourke, Francesca Scolamacchia, Victorye Volkova, Martin Miller and the very sadly missed Magai Kaare all of whom have had to put up with both my chat and my attempts to smash my computer into ten thousand pieces when it doesn't do what it is supposed to do.

I should very much like to thank my rowing coaches Alla Doubrovina and Nick Rankin without whose training schedules this thesis would have been finished six months sooner and I would have gone out of my mind and been sectioned. I would also like to thank Emma Coutts for her help along the way and for putting up with me as I stressed constantly and worked every hour getting this thesis finished.

Contents

List of Figures

List of Tables

List of Abbreviations

AIC	Akaike Information Criteria
AML	Arc Macro Language
AMLS	Animal Movement Licensing System
CAP	Common Agricultural Policy
CI	Credible Interval
CP	Contiguous premises
CPH	County/Parish/Holding
CTS	Cattle Tracing system
DC	Dangerous Contact
DCC	Disease Control Centre
DCS	Disease Control System
DEFRA	Department for the Environment Food and Rural Affairs
ELISA	Enzyme Linked Immunosorbant Assay
ESRI	Environmental Systems Research Institute
EU	European Union
FAO	Food and Agriculture Organisation
FMD	Foot and Mouth Disease
FMDV	Foot and Mouth Disease virus
GB	Great Britain
GLM	Generalised Linear Model
GLMM	Generalised Linear Mixed Model
HPAI	Highly Pathogenic Avian Influenza
IACS	Integrated Administration and Control System
IAH	Institute of Animal Health
IP	Infected premises
LDA	Linear Discriminant Analysis
LWDS	Livestock Welfare Disposal Scheme
MAFF	Ministry for Agriculture Fisheries and Food
NDVI	Normalised Difference Vegetation Index
NMB	National Movement Ban
NSP	Non-Structural Protein
OIE	Office International des Epizootes
OS	Ordnance Survey
PAF	Postal Address File
PZ	Protection zone
ROC	Receiver Operator Characteristic
RPA	Rural Payments Agency
SAMS	Scottish Animal Movement Licensing System
SAT	South African Territories
SIACS	Scottish Integrated Administration and Control System
SIR	Standardised Incidence Ratio
SOS	Slaughter on Suspicion
SZ	Surveillance Zone
UK	United Kingdom
VLA	Veterinary Laboratories Agency
VNT	Virus Neutralisation Test

Chapter 1

Introduction

1.1 The biology of Foot and Mouth Disease

1.1.1 Foot and Mouth Disease viruses

The family of viruses called the Foot and Mouth Disease viruses (FMDV) are the pathogens responsible for Foot and Mouth Disease (FMD). FMD affects all cloven hooved mammals including cattle, sheep, goats, pigs, buffalo, camelids and cervids. FMDV belongs to the apthovirus genus of the *Picornaviridae* family, genetically it is composed of a single RNA strand of around 8,500 positively polarised neucleotides (Ferrer-Orta and Fita, 2004). In keeping with other *Picornaviridae*, FMDV are 22-30nm wide taking an icosahedral form and does not have a lipid envelope (Ferrer-Orta and Fita, 2004). The naked RNA strand is surrounded by a protein shell (capsid) which is composed of 60 protometers each of which is made up of four capsid proteins: VPs1:3 which comprise the surface and the smaller VP4 which is hidden beneath the surface (Ferrer-Orta and Fita, 2004). As with other RNA viruses the process of RNA genome replication is highly prone to error so FMDV populations consist of genetically related but not identical genomes (Domingo et al., 2002).

There are 7 distinct forms or serotypes of FMD virus each of which produces a distinct set of immune responses and a distinct set of antibodies in the host (Anderson, 2002; Davies, 2002b; Domingo et al., 2002; Grubman and Baxt, 2004). These serotypes show geographic patterns with Europe, North and Central America, Greenland, Australasia and Oceania largely FMD-free with the exception of occasional

outbreaks (Knowles and Samuel, 2003). Within the Endemic areas of Africa, Asia, the Middle East and South America the seven serotypes were broadly geographically clustered as such (Knowles and Samuel, 2003):

1. **Types O and A** are the most widely distributed serotypes found throughout Africa, southern Asia, the Far East (not type A) and South America. Type O is the most common of all the FMDV serotypes.

2. **Type C** is largely restricted to the Indian subcontinent.

3. **SAT (South African Territories) 1, 2 and 3** is largely restricted to Sub-Saharan Africa. SAT1 and SAT2 have been slowly moving north through Africa and there have been outbreaks in the Middle East. All three serotypes have a reservoir in wild buffalo, which makes control difficult (Knowles and Samuel, 2003; Vosloo et al., 2002).

4. **Asia 1** is circulating in Asia.

The relationships between the seven FMD serotypes are shown in Figure 1.1. This shows that there can be some very large differences within an individual serotype (particularly the SAT serotypes). This is the result of the same serotype circulating within relatively isolated populations for long periods of time.

The serotype branches in Figure 1.1 can be further broken down into distinct topotypes. The topotypes have evolutionary lineages and are broadly geographically clustered. An example of how serotype O breaks down into eight topotypes is shown by Figure 1.2. Individual virus isolates are usually identified by the country of origin, for example the isolate O\UKG\12\2001 is an isolate from the 2001 UK epidemic. Furthermore sequencing isolates can be used to trace the course of an epidemic (Cottam et al., 2006). Antibodies generated during the immune response are serotype specific, therefore a single animal can be infected by several serotypes.

1.1.2 FMDV transmission and infection

The virus enters the host through the respiratory tract, through the skin via an abrasion or via ingestion. The virus then multiplies at the site of entry, either in

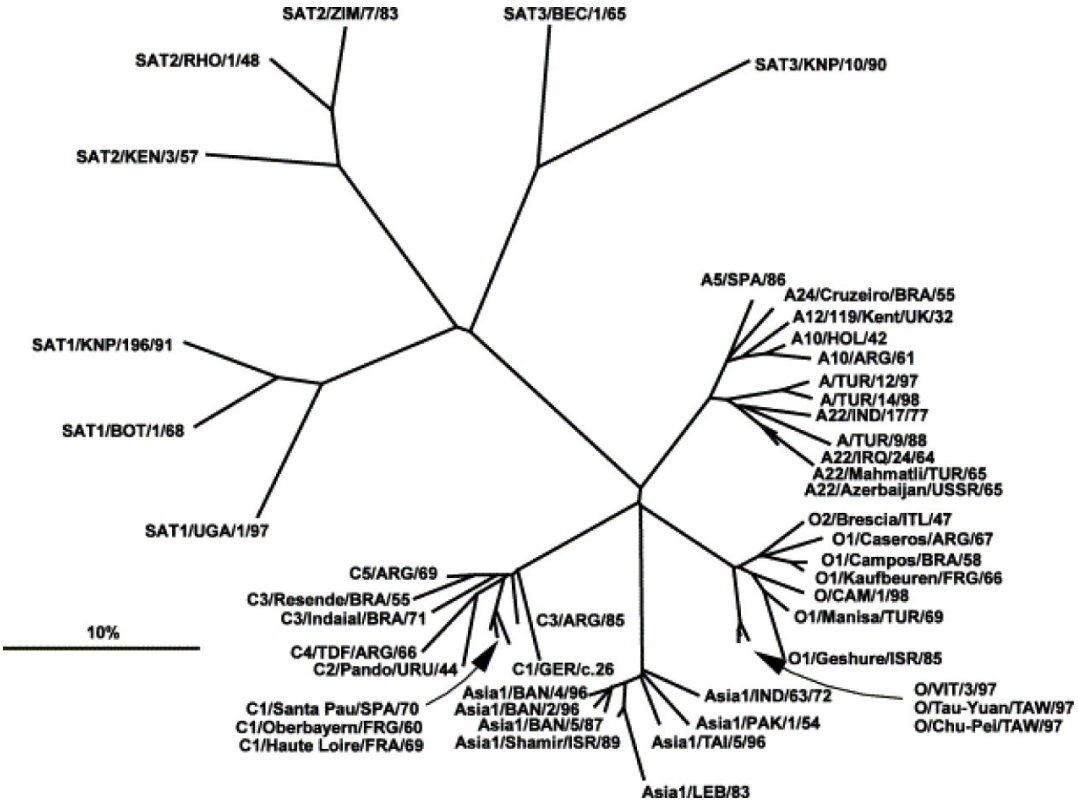


Figure 1.1: *An unrooted Neighbour-joining tree showing the relationship between outer-capsid polypeptides (VP1,2,3) of representatives of each of the seven FMD serotypes* (Knowles and Samuel, 2003, p.66). The sequences are based upon animo acid sequences. The first letter in the identifier for each isolate is the virus serotype. The Figure is from Knowles and Samuel (2003).

the abrasion tissue, or mucosa in the respiratory tract (Kitching and Hughes, 2002), before the virus spreads throughout the body via the blood stream or lymph system. Virus replication is concentrated in the epithelium in mature adults, in particular in the hooves and mouth. In juvenile animals replication is concentrated in heart muscle.

The infected animal excretes the virus in excretions (mainly faeces) and secretions, however, the greatest amount of virus is released through exhalation and in secretions from ruptured vesicles (Haydon et al., 2004). The volume of virus excreted in the breath varies by two orders of magnitude by species with pigs excreting the greatest volume of virus and sheep the least (Donaldson et al., 2001). Furthermore the stage of infection varies between species, with sheep becoming maximally infectious 1-2 days before the development of clinical signs, compared to cattle and pigs which are maximally infectious when they show early, acute signs of disease (Hughes et al.,

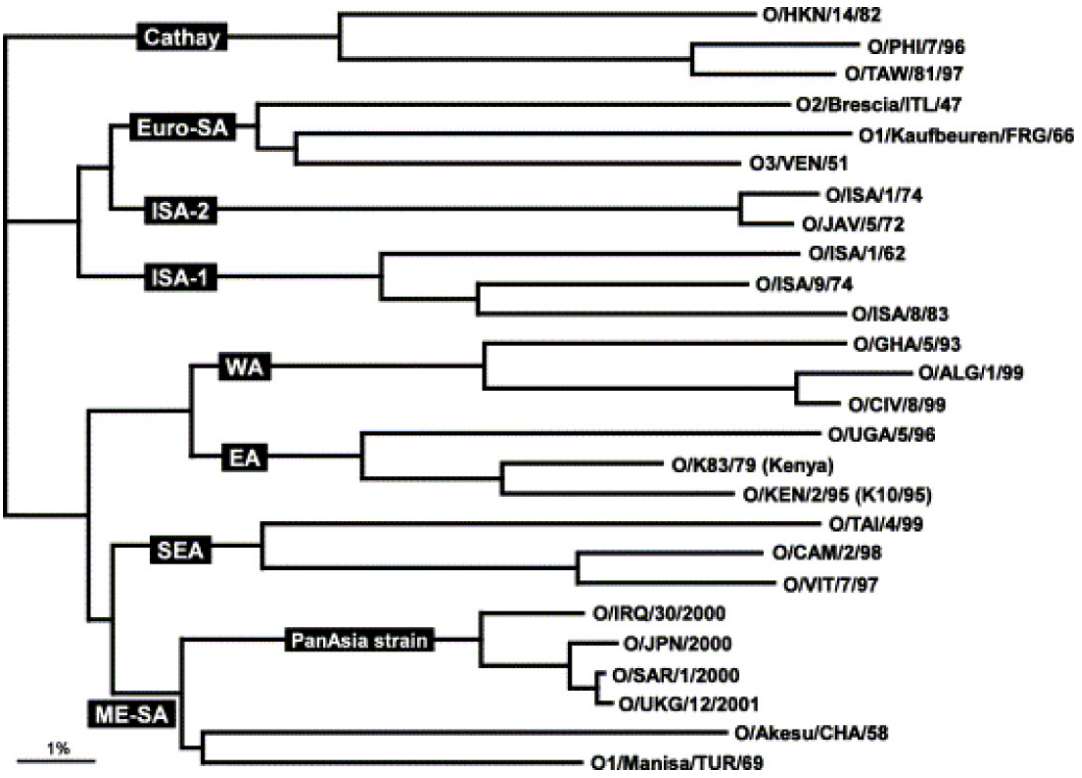


Figure 1.2: *Unrooted Neighbour-joining tree showing the relationship between selected type O FMDVs. Based on complete VP1 gene sequences* (Knowles and Samuel, 2003, p.68)). The identifiers in the black boxes (with the exception of *PanAsia strain* represent the virus topotypes. The Figure is from Knowles and Samuel (2003).

2002).

The infected animal is infectious for only a matter of a few days before the antibody response neutralises the virus. In most cases the virus is completely removed from the host, unless the animal becomes a carrier (Davies, 2002b). However due to the volume of virus produced by infected animals they can be very infectious during the period of infection. The virus can spread via a variety of means between animals across different distances. These mechanisms by which animals can become infected are:

1. **Inhalation.** Susceptible animals in close contact with infected animals principally acquire infection through inhalation of infected droplets and saliva, or through infection by virus in excretions and secretions (Aggarwal et al., 2002; Bouma et al., 2004; Donaldson and Alexandersen, 2002).

2. **Maternal transmission.** Transmission from mother to weaning infant through

infected milk (Donaldson and Alexandersen, 2002).

3. **Fomites.** Given the correct conditions the virus can survive outside a suitable host for up to 60 days (Davies, 2002b) and can be transported on inanimate objects known as fomites such as straw, farm machinery and clothing (Bartley et al., 2002). By these means the virus can be transmitted between animals on a single holding and between animals on separate holdings, potentially over long distances. Such mechanisms of spread are likely to be responsible for the outbreak in Surrey in 2007 (DEFRA, 2007d).

4. **Viral plumes.** Infected animals exhale virus, as a result the virus can be transmitted through the air. The distance over which this is possible is dependent upon the numbers and species of infected animals and climatic conditions. Pigs excrete an order of magnitude more virus than cattle, which in turn produce an order of magnitude more virus than sheep (Donaldson et al., 2001). Under the correct conditions with large numbers of infected pigs, the virus has been shown to have infected animals over large areas and long distances (Gloster et al., 2003; Sørensen et al., 2000; Tinline, 1970). The susceptibility of animals to windborne spread is dependent upon the quantity of virus to which the animal is exposed and the species being challenged (Donaldson et al., 2001) with pigs requiring the greatest infectious dose and cattle the smallest (Donaldson and Alexandersen, 2002, 2001).

In addition to these mechanisms of spread, the movement of infected animals can disseminate the virus over large geographical areas. Moving infected animals can bring FMD to the farms receiving the infected animals. Additionally, the vehicles in which the animals were transported may harbour the virus if the vehicle is not disinfected. By this mechanism the vehicle can act as a fomite and can cause disease in subsequent batches of animals.

Some days after the onset of clinical signs the animal seroconverts and the antigen is cleared. Lesions heal by serofibrinous in filling which takes around four to 5 days for mouth and feet lesions. After around 7 days the epithelium has recovered and there is no further clinical evidence of infection (Donaldson, 2004) and in most cases,

after clinical symptoms have cleared no antigen remains in the body. Cattle and buffalo can be the exception to this rule, the virus can persist in the animal for several months in the oropharynx (Alexandersen et al., 2002). Known as the carrier state these animals are defined as any animal from which live virus can be recovered 28 or more days after infection (Davies, 2002b). There is, however little evidence for transmission by carrier animals (Alexandersen et al., 2002).

1.1.3 FMD Diagnosis

Virus replication in the epithelium results in the separation of the epithelium from the connective tissue. This cavity then fills with vesicular fluid producing the vesicles in the feet and mouth of infected animals which earns the disease its name (Kitching and Alexandersen, 2002). The appearance and subsequent bursting of these vesicles to form lesions form the principal symptoms of FMD and can appear anywhere from 2 to 14 days post infection (Garland and Donaldson, 1990). Further symptoms include lameness (usually the first sign of infection in sheep), reduced productivity particularly in milk yield, lethargy - particularly in pigs, and discomfort - particularly in cattle (Kitching, 2002; Kitching and Alexandersen, 2002; Kitching and Hughes, 2002). Furthermore, infected animals often suffer pyrexia as part of the immune response. Although sheep generally show the fewest clinical signs; high mortality among lambs due to the acute myocarditis or "tiger heart" condition Kitching and Hughes (2002) can serve as a sign of infection within the flock.

In addition to clinical diagnosis, infections can be identified or confirmed in the laboratory by testing for antigen or antibodies. Wherever possible animals are normally tested for antigen by taking samples of infected tissue from lesions. However this is not always possible if there are no or few clinical signs in the animal being sampled. In these cases blood (serum) samples are taken. The speed with which these samples reach the laboratory is a key factor in the accurate diagnosis of FMD.

Different diagnostic tests are carried out depending on whether antigen or antibodies are being tested and on the nature and quality of sample of blood or tissue submitted. Figure 1.3 outlines the diagnostic testing procedure implemented during the 2001 UK FMD epidemic. If a large sample of epithelium is submitted the presence

or absence of antigen can be established immediately using a Direct Sandwich Enzyme Linked Immunosorbant Assay (DS ELISA) (Alexandersen et al., 2003b). When there is insufficient epithelium, or a sample of blood is submitted the sample may be inoculated in bovine thyroid tissue for three days to produce more antigen and the sample is studied to look for a cytopathic effect. Presence of antibody may also be detected in blood samples using a liquid phase blocking (lpb) ELISA (Hamblin et al., 1986a,b). Blood samples submitted for antibody testing as part of a serological survey are tested using a combination of solid phase competitive (spc) ELISA (Paiba et al., 2004) and non-structural protein (nsp) ELISA (Mackay et al., 2001) which typically have sensitivities and specificities of over 99%. Samples are then genetically sequenced to establish the serotype, strain and isolate. The virus neutralisation test (VNT) may be used to test for virus if the ELISA is inconclusive. The VNT is considered the gold standard antibody test but takes several days to yield results (OIE, 2000; Paiba et al., 2004).

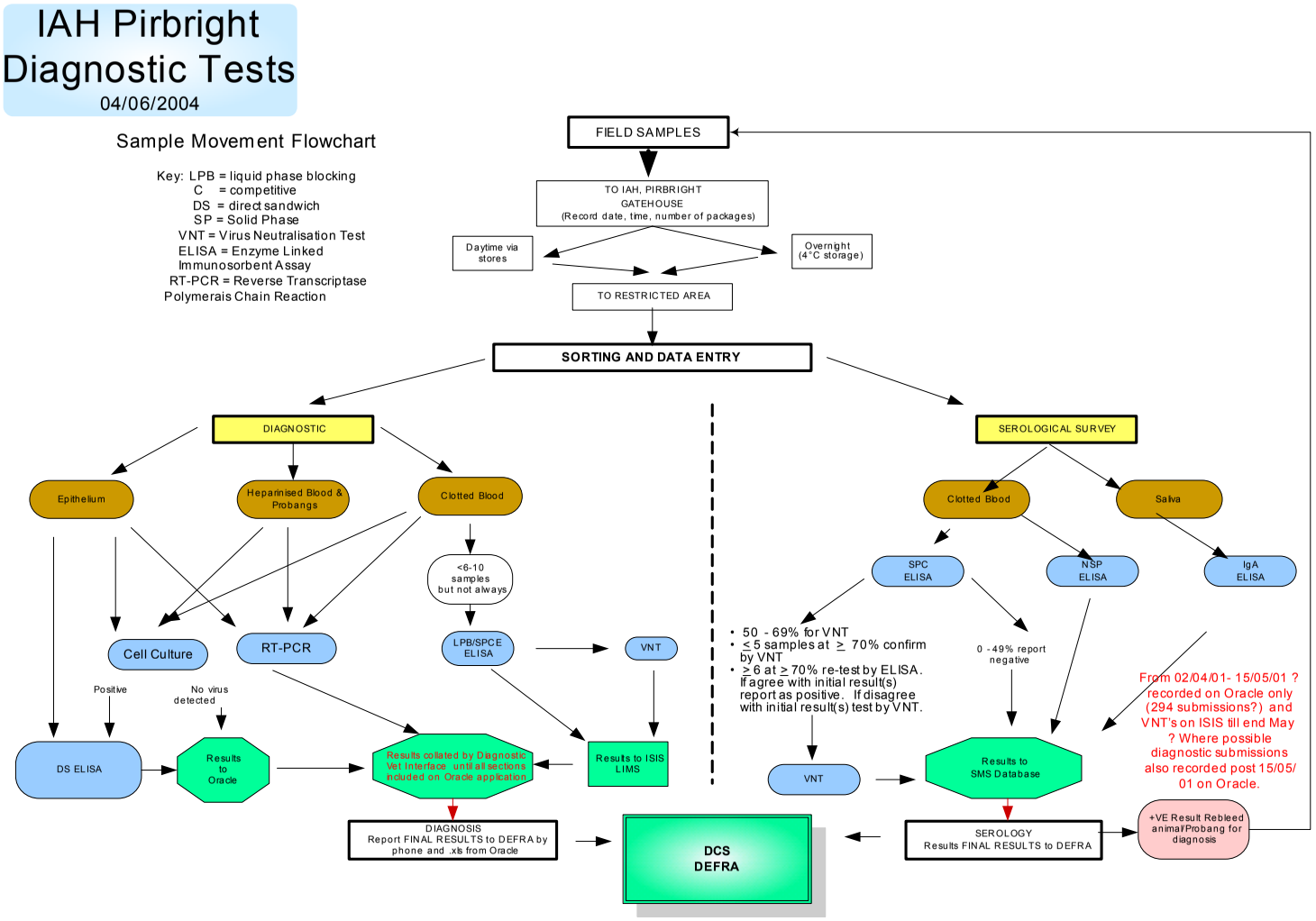


Figure 1.3: Flow diagram of FMD diagnostic procedures implemented by the IAH Pirbright during the 2001 FMD epidemic. Diagram courtesy of DR. N.Ferris (IAH).

1.1.4 FMD Control

FMD is a notifiable disease under the Office International Epizooties (OIE) International Animal Health Code. The OIE recognises two statuses for FMD free countries: FMD free and FMD free with vaccination. FMD free countries are entitled to export animals providing there has not been a case of FMD for at least three months and providing at least three months have passed since the slaughter of the last vaccinated animal (OIE, 2000).

These trading restrictions dictate that the disease presents very different challenges in those countries where it is endemic compared to those which are disease free. Endemic countries typically concentrate on monitoring and containing the disease through mass vaccination, however, infected animals are usually allowed to recover and go through their full life (de C Bronsvoort et al., 2004, 2003). Following introduction of the virus, countries which are normally disease free focus on the swift eradication of the virus in order to restore full trading rights (Anderson, 2002). However, the swift eradication of the virus is balanced against the numbers of livestock slaughtered.

As a result of the speed with which the virus can spread through populations disease control in non-endemic countries is normally conducted at the herd or farm level. At the basic level the virus is contained through farm biosecurity which is often increased during epidemics. During the 2001 UK epidemic the Ministry of Agriculture Fisheries and Food (MAFF) identified eight precautional methods against FMD which included keeping livestock separate, cleaning and disinfecting, avoiding other farms and keeping unnecessary vehicles away (Anderson, 2002). However, these measures are dependent upon the compliance of the farmer.

However, if disease eradication is the goal, enhanced biosecurity alone may be insufficient to contain the disease. Control may be achieved through the slaughter of infected and at risk stock (Howard and Donnelly, 2000; Matthews et al., 2003). A further possibility for controlling FMD is vaccination (Barnett et al., 2002; Keeling et al., 2003; Toma et al., 2002) which is used extensively in FMD endemic countries (Perez et al., 2004). Although vaccines can confer protection to most animals around 4 days post inoculation (Barnett and Carabin, 2002) they have the following

limitations which have limited their use in FMD epidemic countries:

1. Vaccines are strain-specific although multiple vaccines have been developed.
2. The vaccine does not confer 100% immunity.
3. The vaccine stimulates production of antibodies but not memory cells and as a result the antibody titre decays with time. After approximately six months the antibodies are no longer present (Kitching et al., 2007).
4. If an animal has been infected with FMD before vaccination the animal will still develop the disease, although will not generally develop clinical signs.
5. Animals which have been vaccinated and been infected can still carry virus in the carrier state. Such animals must be identified and slaughtered.
6. Most ELISA antibody tests can not differentiate between antibody generated as a result of infection and the vaccine. However, non-structural protein ELISA tests have recently been developed which can differentiate between virus and vaccine (Lu et al., 2007).
7. If vaccination is in response to an epidemic, resources may be very limited and vaccinating all at risk animals may be difficult (Brownlie, 2001; Tildesley et al., 2006).

As a result of these limitations, in countries normally disease-free, vaccines are only used as a means of containing the virus. Vaccines can be used in three ways; either vaccination of an area, ring vaccination to protect stock surrounding an infected premises (IP), or selective vaccination of particular stock (Barnett et al., 2002; Davies, 2002b; Keeling et al., 2003; Toma et al., 2002). OIE restrictions dictate that countries which employ 'vaccination to live' must wait 12 months after the last case to regain disease-free status, or 24 months after the last case when routine prophylactic vaccination is applied to all animals (Anderson, 2002)

1.1.5 Monitoring the epidemic - R

Monitoring the status of the epidemic and the efficacy of control strategies is achieved by analysing the case reproduction ratio (R). An epidemic can be considered to be

spreading 'out of control' while R is greater than 1 (Tompkins et al., 2002). R is a statistic referring to the number of secondary cases which are likely to be generated by a primary case given a fully susceptible population. Therefore, if R is greater than 1 the epidemic will continue to increase in size, but if it is less than 1 the epidemic will die out. Calculation of R is used to indicate the state of the epidemic and to evaluate the effectiveness of control strategies.

1.2 FMD in the UK

In the years prior to 1967 there were regular outbreaks of FMD in the UK. However, in this period cases rarely exceeded 200 and the outbreaks were contained rapidly through a stamping out policy (Northumberland, 1968). An introduction of FMDV in October 1967 resulted in an epidemic lasting 212 days which included 2364 infected premises in 16 counties, principally Cheshire and Staffordshire (Haydon et al., 1997). Following that outbreak the UK managed to eradicate small outbreaks and between 1968 and 2001 there was one case of FMD in the UK, this was on the Isle of Wight in 1981 (Sørensen et al., 2000).

The UK and much of Western Europe remained largely disease free between 1981 and 2001. However, in 2001 there was an epidemic lasting 221 days comprising 2030 outbreaks in the UK (Alexandersen et al., 2003a; Davies, 2002a; Gibbens et al., 2001; Mansley et al., 2003). Additionally there were a number of secondary outbreaks in Ireland, France and The Netherlands (Bouma et al., 2003; Chmitelin and Moutou, 2002; Costelloe et al., 2002; Pluimers et al., 2002). This epidemic will be described in more detail with particular reference to the epidemic in Great Britain (GB) which comprised 2026 of the UK cases, the remaining four were in Northern Ireland.

The UK remained FMD free for 6 years until on August 3rd 2007 FMD was confirmed on a cattle farm in Surrey and a second IP was confirmed on a nearby cattle farm on the 6th August. The resulting epidemic comprised 8 IPs, with 1,578 animals (principally beef cattle) slaughtered on IPs, of these 278 animals were found to be infected with the type O strain of the virus (Ryan et al., 2008). The virus was introduced from a vaccine production plant at the IAH Pirbright (Cottam et al.,

2008). The outbreak comprised two groups of cases; two cases around the first IP and a secondary cluster of 6 IPs (one on two locations) 15km further north in Surrey as a result of introduction from IPs 1 and 2 onto IP5, the estimated date of infection is the 6th to the 18th August 2007 (Figure 1.4, DEFRA (2007c)). However, IP3 was the first holding to be declared an IP in this cluster. This was confirmed on the 12th September and is likely to have been infected as a result of local spread from IP5 (Cottam et al., 2008).

1.2.1 Epidemiology of the 2001 UK FMD epidemic

Origin and early dissemination

The likely source of the 2001 epidemic was a waste food feeding pig farm in Heddon-on-the-wall in Northumberland called Burnside Farm, where pigs may have been illegally fed uncooked swill infected with FMDV (Alexandersen et al., 2003a). The disease was identified on the 19th February 2001 in pigs sent from Burnside Farm to an abattoir in Essex. The time between introduction of the virus and the detection of the first case was possibly as long as three weeks (Alexandersen et al., 2003a).

The second oldest lesions were found on a mixed sheep and beef cattle farm in Ponteland close to Burnside Farm (Alexandersen et al., 2003a). These animals are likely to have been infected by the large airborne viral plume from Burnside Farm produced by the large number of infected pigs (Gloster et al., 2003). Infected sheep from the Ponteland farm were transported to a market in Hexham in Northumberland (Figure 1.5). Sub-clinically infected stock were unknowingly purchased and taken from Hexham to the Longtown market in north Cumbria, where a large number of animals became infected (Mansley et al., 2003). These infected animals were sold off in several lots and seeded infections locally in several locations in Cumbria and Dumfries and Galloway. The problem was exacerbated as this was a time of year with large numbers of sheep being moved (Anderson, 2002), one dealer in particular transported infected animals south to seed infection in multiple locations in the Welsh Borders area and Devon (Gibbens et al., 2001; Gibbens and Wilesmith, 2002; Mansley et al., 2003). Infection was also spread to Northern Ireland, and indirectly to The Netherlands (Figure 1.6).

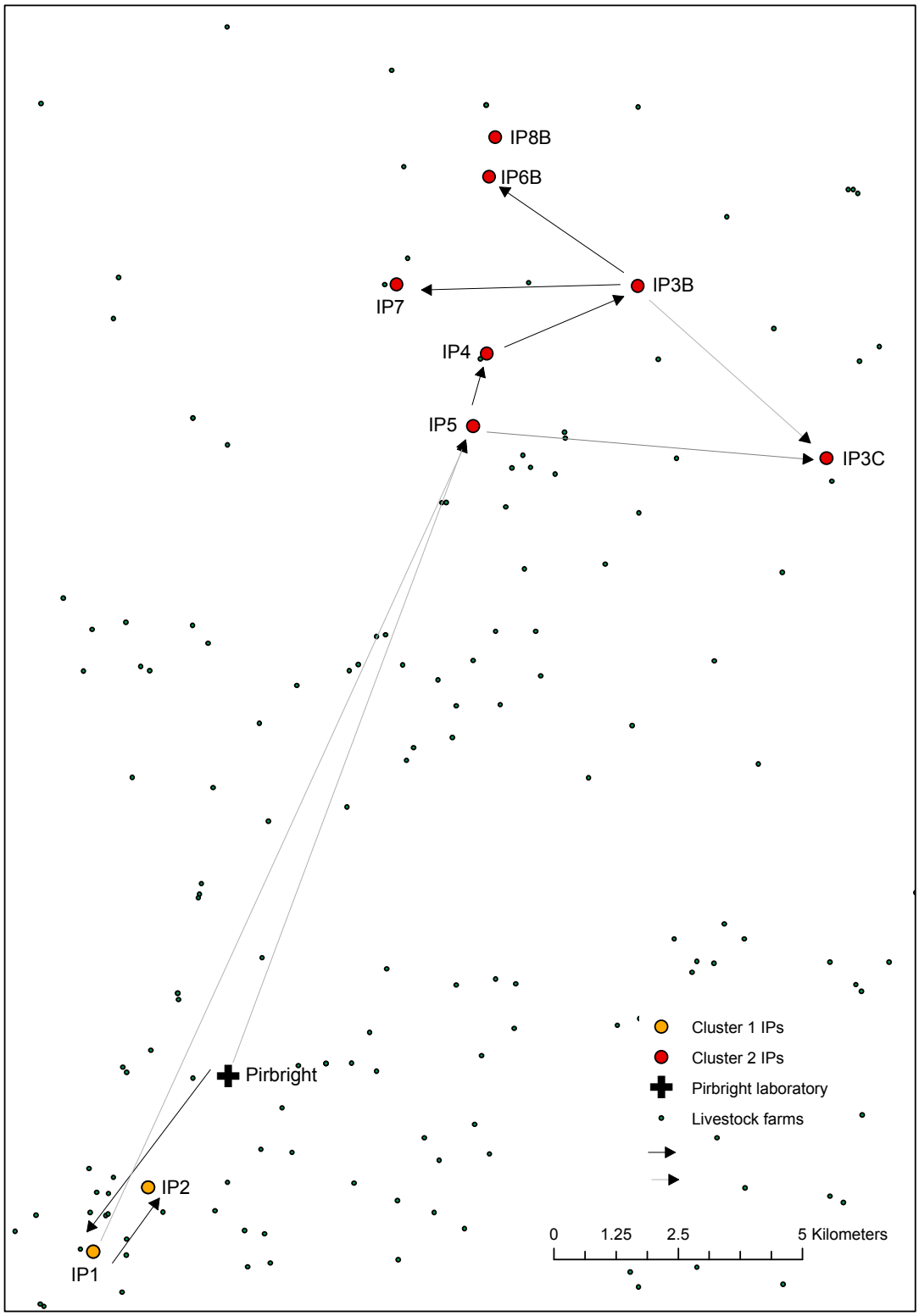


Figure 1.4: Distribution of IPs in the 2 clusters in Surrey against the farm population. The links are links identified by tracing according to DEFRA (2007c); no link had been suggested for IP8.

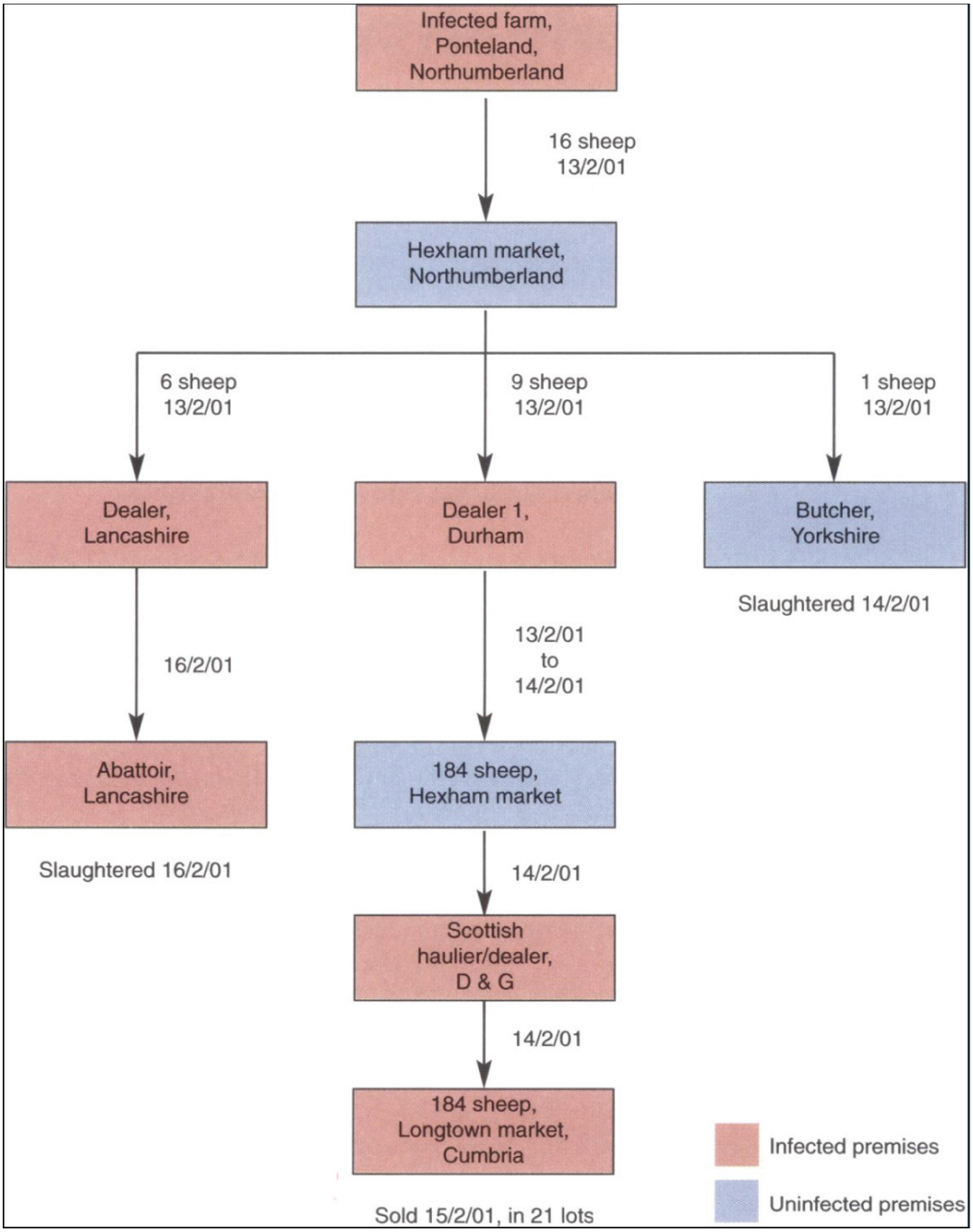


Figure 1.5: Path of virus dissemination from The Ponteland IP infected to Longtown market through sheep movements. Diagram from Mansley et al. (2003).

Following confirmation of the first case, a local animal movement ban was imposed in Essex on the 21st February and epidemiological investigations began. When the case in Northumberland was identified on the 23rd February a national movement ban (NMB) on susceptible species was imposed commencing at 5pm, although animals already in transit were allowed to complete their journey. Following imposition of the

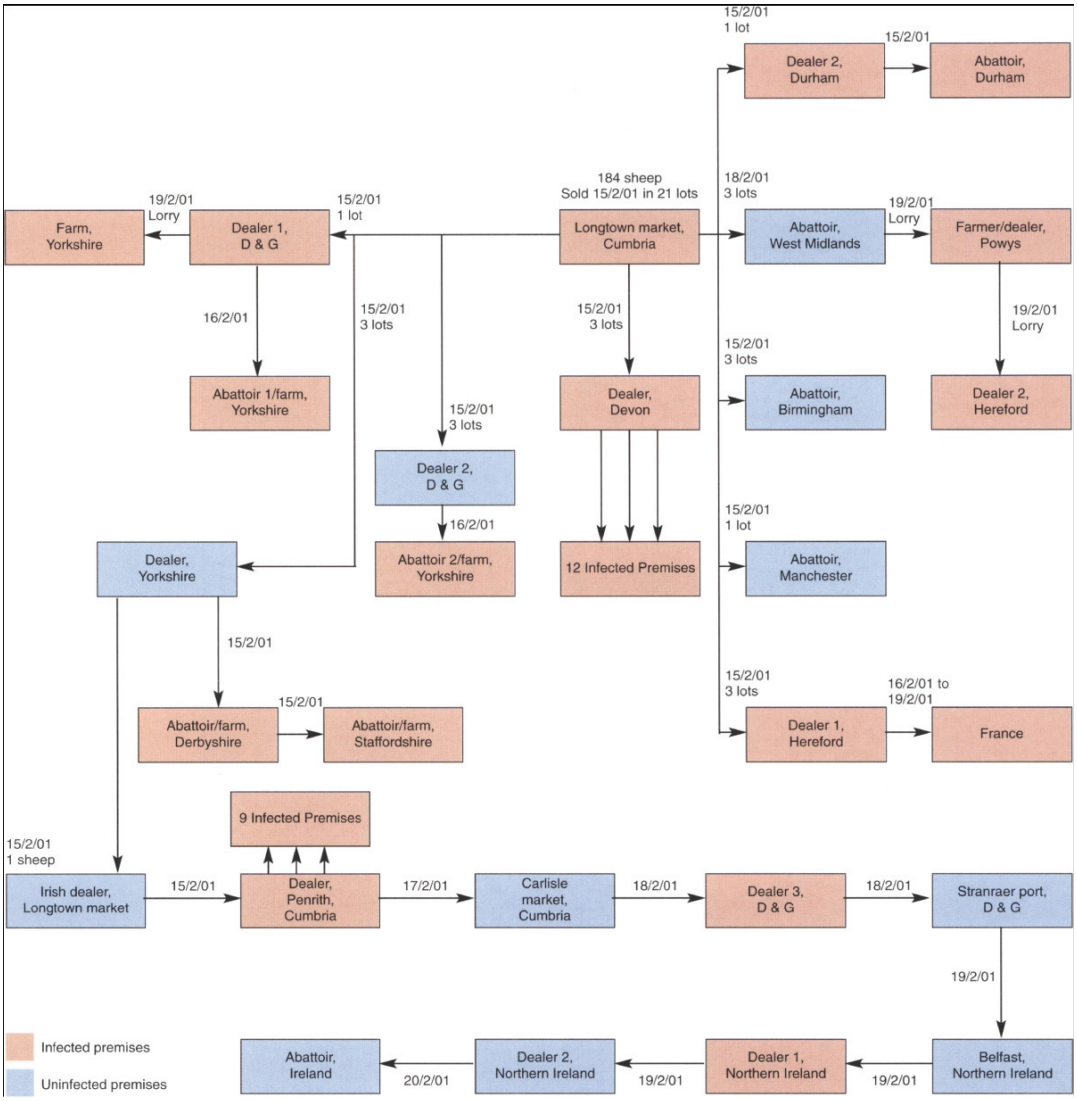


Figure 1.6: Path of virus dissemination from Longtown market throughout the country via sheep movements. Diagram from Mansley et al. (2003).

NMB virus spread changed from being long range through animal movements to being local through contact between animals, fomites and occasionally wind. Thus the virus was restricted largely to areas in which it had been seeded initially (Keeling et al., 2001). It is estimated that between 70 and 119 farms may have been infected before the NMB (Figure 1.7) (Gibbens and Wilesmith, 2002; Haydon et al., 2003; Mansley et al., 2003). Had the NMB been imposed when the disease was first identified, then around 40 premises would have been infected prior to the NMB and the epidemic would have been between one-third to one half its final size (Haydon et al., 2003).

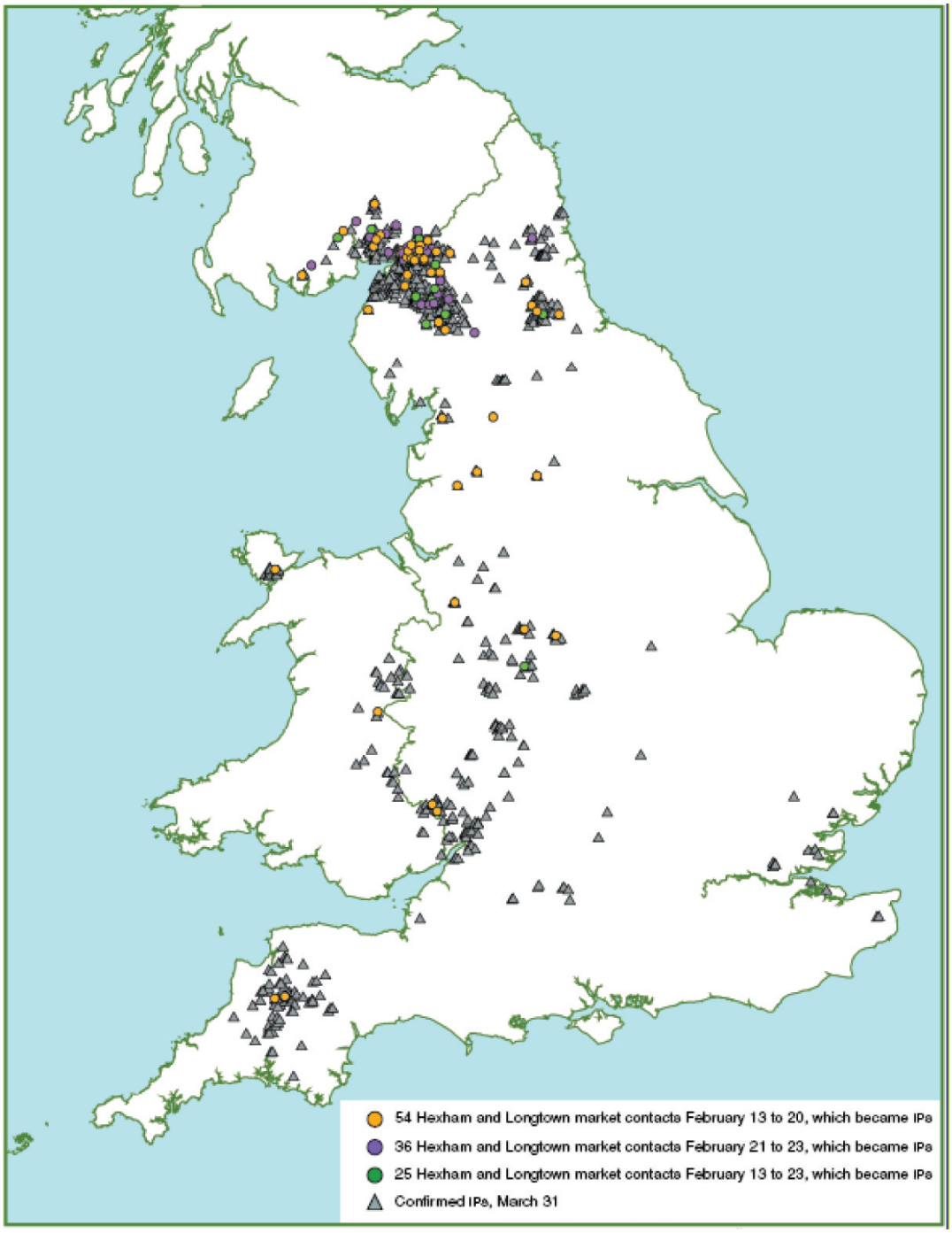


Figure 1.7: Distribution of IPs confirmed before the 31st March. Additionally IPs with a like to the Hexham or longtown markets are highlighted. Diagram from Mansley et al. (2003).

1.2.2 Animals affected

Unusually for the type O strain of FMD the main infected species were sheep (Kitching and Hughes, 2002), with spill over into the cattle population which frequently resulted in identification of infection (Gibbens et al., 2001). There were very few

cases in pigs after the index case. The heavy infection of sheep presented its own sets of problems, summarised by Davies (2001):

Infected sheep are often difficult to detect, particularly when the disease is in its early stages; it presents as lameness which, as veterinarians will know, is a very common condition at this time of year. The disease may not become evident until it is well established in the flock and the virus is being released in quantity (Davies, 2001, p.386).

Davies goes on to explain how this was facilitating disease spread locally. A large number of cattle were infected during the epidemic, however, cattle were not considered the main problem because cattle present distinct clinical symptoms early in the course of infection (Anon, 2001c; Davies, 2001). This was particularly the case during the early stages of the epidemic when it was feared that latently infected sheep were transmitting infection to cattle (Davies, 2001).

1.2.3 Progression of the epidemic

Following the NMB the disease continued to spread locally, principally in Cumbria, the Welsh borders, Devon, Dumfries and Galloway and North East England (Figure 1.7). Nationally the epidemic peaked in early April after which the number of new cases per day declined (Figure 1.8). This decline was retarded somewhat by a new disease cluster in the Settle area of North Yorkshire (Table 1.1). The virus was eradicated from Dumfries and Galloway and Devon fairly early in the tail (Table 1.1), but persisted in other areas until August or September.

Spread of the virus can be divided up into three broad phases which follow the typical pattern for an epidemic curve (Swinton et al., 2002) and corresponds to the phases in Figure 1.8:

1. **Phase 1 (Dark grey area)**: Initial rapid growth as the epidemic spreads through the susceptible population and susceptible holdings are removed, R is greater than 1.

2. **Phase 2 (Mid-grey area)**: Epidemic under control. The numbers of cases declines steeply as the susceptible population is reduced, R is less than 1.

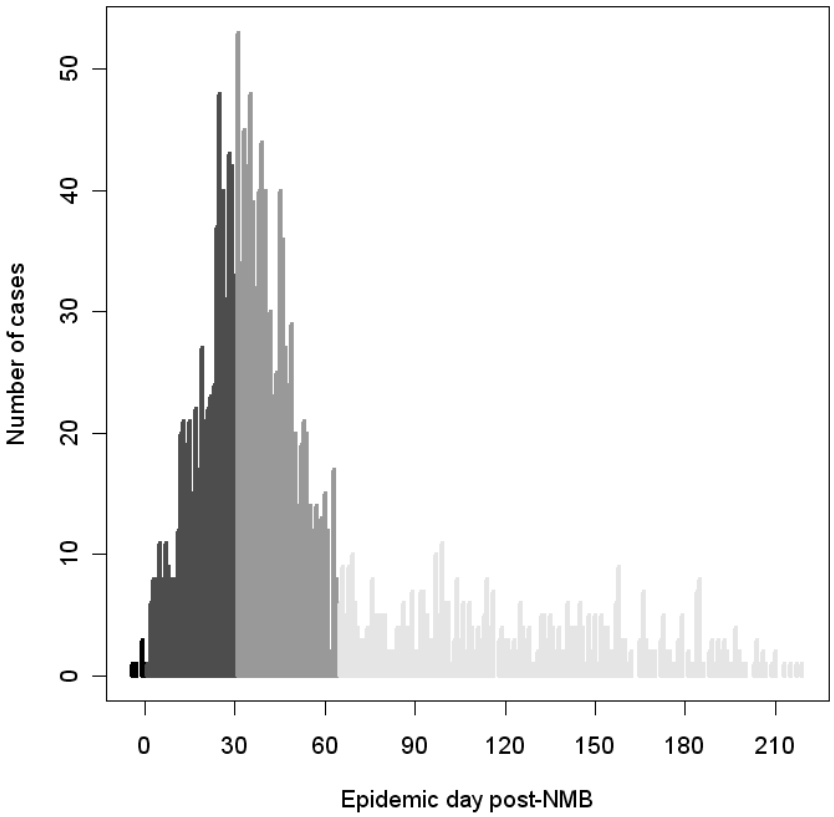


Figure 1.8: Number of cases per day for the GB FMD epidemic. The dark grey corresponds to the initial rapid growth of the epidemic, the intermediate grey to the epidemic under control and in decline and the light grey to the epidemic tail.

3. **Phase 3 (Light grey area)**: The long tail. This typically corresponds with recovered individuals reentering the susceptible population (Swinton et al., 2002), however in this epidemic it corresponds with the virus entering new areas, R fluctuates either side of 1.

1.2.4 Control measures

The goal of disease control was to eradicate the virus and to regain the UK's 'FMD free without vaccination' status (Keeling, 2005). This required the rapid containment of the epidemic and broadly comprised two components:

1. Enhanced biosecurity. These measures and restrictions broadly comprised three levels:

   (a) Biosecurity on susceptible holdings. Such measures include footpath clo-

Region	Total IPs	Report date		Duration (Days)
		First IP	Final IP	
Cumbria	892	28 Feb	30 Sep	214
Welsh Borders	253	26 Feb	11 Aug	166
North East England	190	22 Feb	28 Sep	218
Dumfries and Galloway	176	28 Feb	23 May	84
Devon	172	2 March	17 June	107
Settle	102	10 May	16 Aug	98
Rest of England	214	19 Feb	21 Aug	193
Rest of Wales	16	25 Feb	25 April	59
Scottish Borders	11	28 March	30 May	63

Table 1.1: Duration of the epidemic in the major FMD regions. The regions described in this table are illustrated in Figure 1.9.

sures and the disinfection of vehicles and individuals entering the holding.

(b) Biosecurity on holdings at elevated risk. This was at two levels (National Audit Office, 2002, p.57):

i. Upon confirmation of the IP (either clinically or on laboratory grounds) an infected area is declared around the IP with a minimum radius of 10km. Inside the infected area all stock movements are banned, livestock vehicles must be cleansed and disinfected and milk can only be fed to animals on the same premises.

ii. A 3km Protection Zone (PZ) is imposed around the IP, these restrictions can also placed on farms believed to have some link with an IP. The restrictions include the isolation of all animals and movement restrictions on owners of susceptible stock. Additionally all visitors must disinfect thoroughly.

(c) Biosecurity on IPs prior to slaughter. Further restrictions and enhanced disinfection. Personnel that had worked on an IP were declared 'dirty' and were not allowed on another holding for a further 72 hours.

2. The rapid identification and slaughter of IPs. This was aided through patrol visits to premises within the 3km PZ and 10km Surveillance Zone (SZ). Furthermore, declaration of an IP was accompanied by the swift identification and slaughter of DCs.

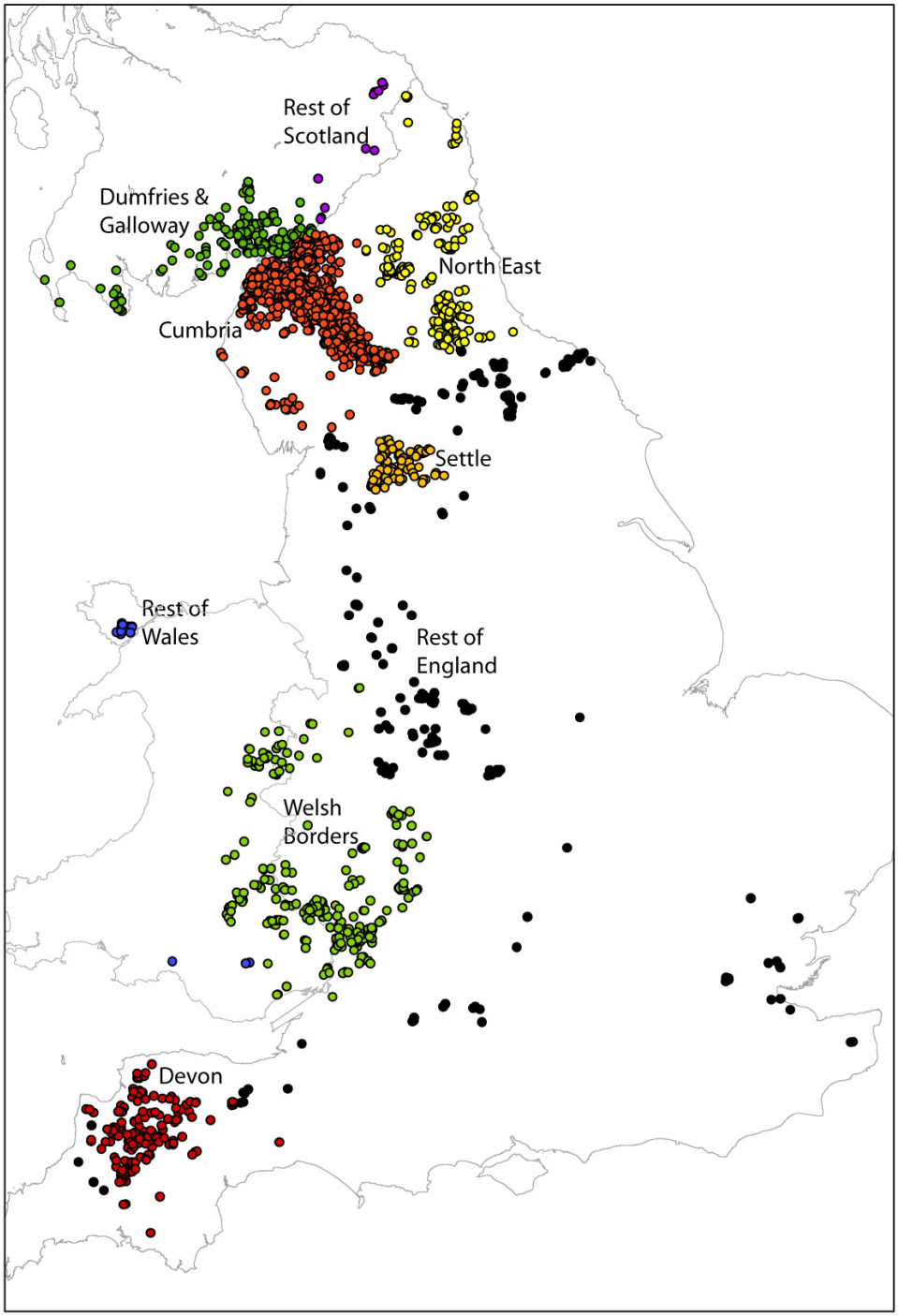


Figure 1.9: Distribution of IPs broken down by the regions described in Table 1.1. The colour of data points corresponds to the labels.

However, during the early stages of the epidemic reporting to slaughter times were in many cases substantially longer than during the last major FMD outbreak

in 1967/1968 (Haydon et al., 2004). This was partly due to a large decrease in veterinary resources since the 1967/68 epidemic and the more localised spatial scale of the 1967/68 outbreak (Anderson, 2002). Further logistical problems and public concerns frequently lead to a long delay before disposal of carcasses (Gloster et al., 2001; Jones et al., 2004; Lowles et al., 2002; Scudamore et al., 2002).

Towards late March evidence suggested (Anderson, 2002; Ferguson et al., 2001a; Keeling et al., 2001; Woolhouse et al., 2001) that the case reproduction number was above 1 and that the epidemic was out of control. It was asserted that the epidemic would remain out of control unless more intensive control policies were adopted (Anderson, 2002). Various alternative strategies for epidemic control were considered (Anderson, 2002) and subsequently the 24/48 policy came about. The 24/48 policy required that all stock on IPs were to be culled within 24 hours of confirmation, and all stock on Dangerous Contact premises (DCs) or premises contiguous to IPs (CPs) were to be culled within 48 hours of confirmation. Other local policies were adopted including a cull of sheep and pigs within 3km of any IP in Dumfries and Galloway and Cumbria, amidst concern that the sheep population may be harbouring widespread sub-clinical infection. A similar 'Local' cull was implemented in Anglesey and north west Wales (Anderson, 2002). Also, towards the end of March the delay in receiving laboratory confirmation on suspect IPs was identified as a bottle neck (Anderson, 2002). To overcome this the Slaughter on Suspicion (SOS) category was implemented whereby if a veterinarian could not diagnose FMD, but at the same time could not be certain that FMD was not present, the animals would be slaughtered as a precaution without waiting for clinical diagnosis (Anderson, 2002). These policies lead to a substantial increase in the ratio of IPs to premises culled showing no clinical signs of FMD (non-IPs) (Figure 1.10).

Various vaccination strategies were considered (Brownlie, 2001) and resources placed on standby for vaccination of cattle housed indoors in Cumbria during April. Cattle housed indoors would be easier to vaccinate and it was feared that once the cattle were released they would contract the disease from latently infected sheep causing a resurgence of the epidemic (Anon, 2001d). This vaccination was never implemented due to strong opposition by farmers groups and a shortage of the necessary

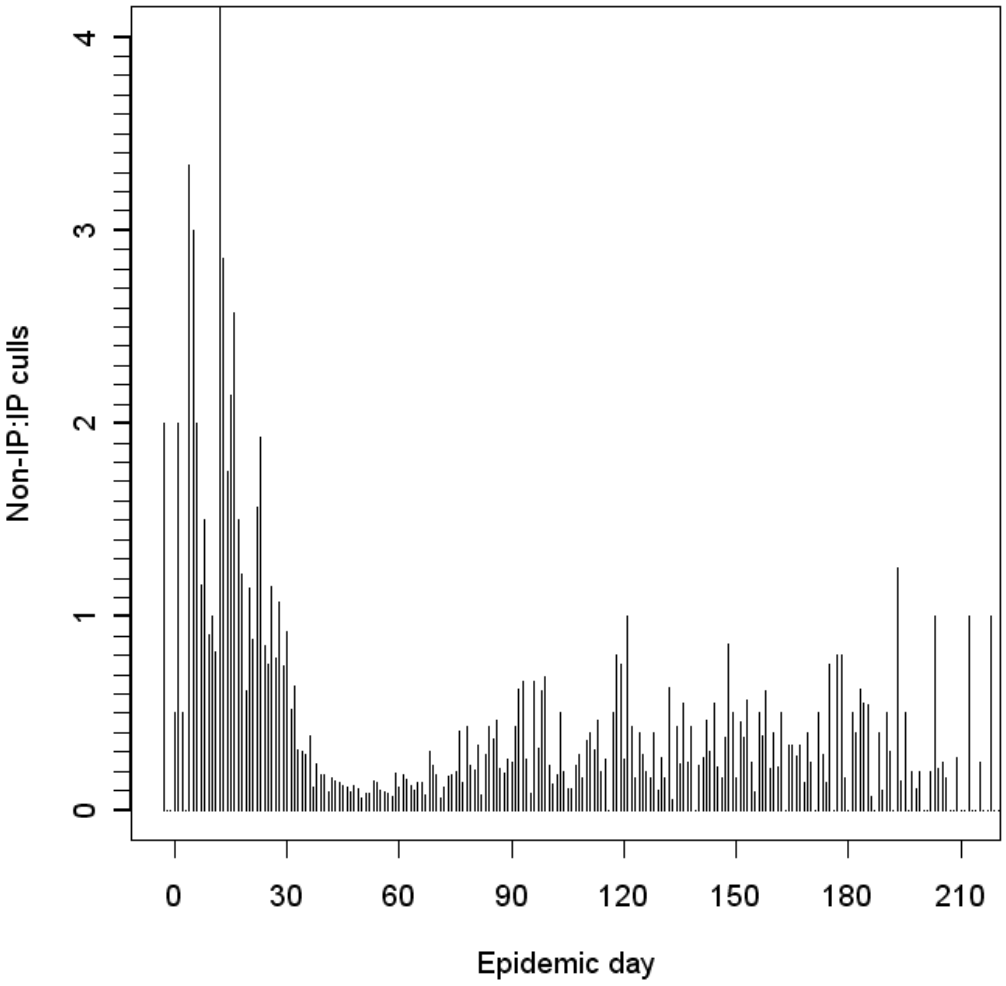


Figure 1.10: Daily nonIP:IP culling ratios by the day of the epidemic on which animals were slaughtered, day 0 is the day the NMB was imposed. A ratio of 2 means that 2 non-IPs were culled for each IP on that day.

resources (Anon, 2001e).

In order to demonstrate disease-free status and resume trading a program of post-epidemic serosurveillance must be undertaken. This involved sampling animals from all small-ruminant holdings within the PZ such that a prevalence of 5% could be detected with 95% confidence (Thrusfield et al., 2005a). Within the SZ serosurveillance must detect one infected holding given an estimated flock prevalence of 2% (Thrusfield et al., 2005a). These samples were analysed using a competitive ELISA with a VNT to confirm positive results.

1.2.5 Animal welfare

Disease control measures, in particular enhanced animal movement restrictions within the PZ and SZ were having an increasingly adverse effect on the health of many animals (Anderson, 2002). Many farms had pregnant animals, or animals being reared which during a 'normal' year would have been moved on during the spring. The movement ban meant that many farms had more stock than the available resources could cater for, such resources were principally feed, pasture and shelter (Anon, 2001a). To alleviate these problems licensed movements of stock were allowed in disease-free areas. The Livestock Welfare (Disposal) Scheme (LWDS) was introduced in infected areas whereby adversely affected stock could be slaughtered in order to reduce livestock numbers on the farm (Anon, 2001b,a). A total of 2,293,000 animals, principally sheep (1,821,000), pigs (366,000) and cattle (166,000) were slaughtered under this scheme (Anderson, 2002). Sheep were principally affected as a result of lambing in the Spring which lead to overstocking in many areas.

1.3 Analysis of the UK 2001 epidemic

The disease dataset generated during the 2001 epidemic has resulted in a large amount of analysis of the epidemic. This has principally comprised mathematical and statistical modelling and reconstruction of the pattern of transmission.

1.3.1 Mathematical Modelling

Three mathematical models of FMD spread were used during the FMD epidemic and the findings of the models shown to the government to advise on control policy (Ferguson et al., 2001b; Keeling et al., 2001; Morris et al., 2001; Anderson, 2002) and a further model developed since the epidemic (Diggle, 2006). The models were all distinctly different, one model being deterministic and initially not spatially explicit (Ferguson et al., 2001b) but with a spatial component added at a later stage (Ferguson et al., 2001a), whilst the others were spatially explicit stochastic models of spread (Keeling et al., 2001; Morris et al., 2001). The models of Morris et al. (2001) and Keeling et al. (2001) differ in that Morris et al. (2001) requires many more parame-

ters than the transmission kernel and species level transmissibility and susceptibility required by the model of Keeling et al. (2001). The three models were all populated with data from the June 2000 agricultural census and all models produced similar results during the epidemic which possibly lead to a great deal of belief in the reliability of their results (Anderson, 2002). However, there are potentially very limiting assumptions in these models which need to be further investigated (Haydon et al., 2004; Kao, 2002; Keeling et al., 2001), these include (not all limitations apply to all models):

1. The farm data used to seed the model is accurate both in its spatial and demographic components. There were discrepancies between the farm demographic data used and the distribution of livestock which was observed during the epidemic (Haydon et al., 2004; Keeling et al., 2001).

2. Euclidean distance is an accurate measure of disease transmission across space. Models assume that there are no geographic features which may facilitate, or hinder the spread of the virus. This has been investigated at a coarse scale by Savill et al. (2006) who demonstrate that when holdings are separated by an estuary road distance may be a more accurate predictor of risk than Euclidean distance.

3. The transmission kernel of Keeling et al. (2001) is based upon contact tracing data which tended to overestimate the importance of local spread (Haydon et al., 2004).

4. That susceptibility and transmission can be modelled as a linear function of the number of animals on the farm (Diggle, 2006; Keeling et al., 2001).

Further investigations of these assumptions are required in order to validate the models' structure and to make improvements to the models.

Further mathematical modelling involved the deconstruction of the landscape to a hexagonal lattice and treating each hexagon as a cell (Kao, 2003). The effects of different control strategies within the hexagons are explored, including culling based upon expert knowledge and it was found that expert knowledge would not have

alleviated the epidemic. Further empirical analyses (Honhold et al., 2004a; Thrusfield et al., 2005a) looked at the effects of control strategies in different regions of the UK. The studies underline the importance of minimising reporting to slaughter time and some results suggest that the control effort may not have been entirely effective. However, the results ignore spatial and species level differences and assumes that R will decline in a linear manner during the epidemic which calls these analyses into question.

1.3.2 FMD transmission analysis

Although data on likely routes of infection were gathered during the epidemic these data were only partially complete and the accuracy is uncertain. To overcome the incompleteness, Haydon et al. (2003) use the contact tracing data to reconstruct the pattern of spread as epidemic trees, using the data to fill in missing branches. Where contact tracing data was missing a source farm is randomly assigned based on proximity and infectious window. Further analysis of the trees was carried out by 'pruning' certain branches. For instance by implementing the NMB on the 20th February rather than the 23rd February the average epidemic size was 793 farms, and by reducing time between reporting and slaughter from 1.23 to 1 day the average epidemic was 1093 farms. This simple model included several generalisations, in particular that only one of several possible sources was assigned as the route of infection, and that once a farm had been 'pruned' that it would not have become infected by some other source. It has subsequently been demonstrated that accurate epidemic trees can be constructed using genome sequencing (Cottam et al., 2006). This was applied to 23 FMD virus isolates from 21 farms in the UK and Ireland, the authors found an average of 1.5 nucleotide substitutions per farm infection, suggesting a rapid rate of mutation, although these analyses need to be applied to more farms.

1.3.3 Overview of FMD analysis

The 2001 FMD epidemic generated the most complete dataset of any epidemic of animals (Haydon et al., 2003), and it resulted in a large amount of analysis. Much of this retrospective analysis has focused upon the control of the epidemic (Honhold

et al., 2004a,b; Keeling et al., 2003; Taylor et al., 2004). Whilst this is valuable research for informing future epidemics the mechanisms by which the virus is moving between holdings and factors which put holdings at risk of becoming infected transmitting infection have been largely ignored. By understanding such processes in their spatial and temporal settings, tools can be developed to identify areas and holdings at greatest risk of infection and the information can be used for targeting resources. This requires detailed spatial epidemiological analysis of the epidemic in order to develop methodologies which can be applied prospectively to future epidemics.

1.4 Spatial epidemiology

1.4.1 Background

"Spatial epidemiology is the study of spatial variation in disease risk or incidence" (Ostfeld et al., 2005, p.328). Although spatial epidemiology undoubtedly goes back further, it is often traced back to the work of the physician John Snow and his analysis of cholera in London in August and September of 1854 (Snow, 1965). Snow mapped the locations of some 616 cholera deaths in Soho in London (Figure 1.11) during a particularly serious outbreak. Mapping cases in relation to the location of a potential pollution source, in this case water pumps, showed that cases were clustered around a pump on Broad Street (Figure 1.11). Disabling this pump bought an end to the epidemic and demonstrated for the first time an underlying biological cause for cholera, namely waterborne bacteria.

Snow's work is a very simple example of how analysing spatial patterns of disease can reveal underlying processes. The spatial aspects of diseases were explored in depth by Pavlovsky (1966) who coined the term landscape epidemiology. Pavlovsky (1966) identifies three fundamental aspects of diseases which underpin landscape epidemiology:

1. That disease tends to be limited geographically. Many diseases are constrained in their spatial extent.

2. These spatial variations arises from variations in the physical and/or biological conditions that support the pathogen and/or its vectors;

Figure 1.11: John Snow's Broad street Cholera map: the black lines represent deaths from cholera, locations of water pumps are also marked. Source: University of Illinois at Chicago (2008).

3. By mapping these biotic and abiotic conditions current and future risk can be predicted.

The two questions which arise are (Ostfeld et al., 2005):

1. How to understand and quantify the conditions which result in a disease distribution.

2. How to use these parameters to map risk and make the predictions.

Subsequently, research into spatial epidemiology can be divided into two broad strands:

1. **Disease mapping and cluster analysis.** At the simplest level such studies involve the simple mapping of case data to explore patterns in a disease. In human diseases in particular, the mapping of cases is related to the population at risk to look for excess incidences (Lawson et al., 1999, 2003; Openshaw, 1996),

or potentially areas of reduced incidences. In relation to FMD, clustering of cases has been investigated in two ways:

(a) Spatial transmission kernels (Ferguson et al., 2001a; Keeling et al., 2001; Diggle, 2006) describe the probability of an individual farm becoming infected given its Euclidean distance to an infectious source. These transmission kernels help to explain the development of national clusters around initial point sources (acting similarly to Snows pumps), in this case the point sources currently infected IPs.

(b) Spatio-temporal cluster analysis. Analysis of spatial and temporal windows over which transmission occurred and exploration of these values to understand the processes operating (Picado et al., 2007; Wilesmith et al., 2003).

These analyses of FMD were all carried out using point data as both the numerator (case) data and the demographic data are available as point datasets. However in many studies, particularly of human diseases (Lawson et al., 1999, 2003; Openshaw, 1996) but some livestock diseases (Stevenson et al., 2000; Perez et al., 2002; Ward and Perez, 2004), these data are only available as counts or rates aggregated over some spatial unit, in these instances different techniques are used for analysis.

2. **Geographical correlation.** Following the identification of a pattern in disease incidences further analysis can be conducted to see if the pattern is the result of some environmental risk factors. Such factors may include a point source of pollution such as the pumps in Snow's cholera work, or some underlying environmental risk factor such as the Normalised Difference Vegetation Index (NDVI) or temperature. Such factors are commonly identified statistically based on the distance to point sources, or a measure of the background measures of infection risk. In the study by Odiit et al. (2006) on sleeping sickness this is implemented using background measures (such as NDVI) and by measuring the distance to the swampland which is the habitat for the tsetse vectors and can be considered as point sources. Once these predictors have

been identified their distribution can be used to generate risk maps or to infer the likely distribution of a disease in area in which the disease has not been sampled (Clements et al., 2007, 2002; Hay and Snow, 2006; Tatem et al., 2007).

An additional area of analysis is analysis of the role of landscape features in disease transmission (Ostfeld et al., 2005). This requires fine scale infection data on an infectious disease which is not always available. Such analyses have been applied to the effects of fragmented habitat patches on lyme disease transmission (Allan et al., 2003) and on the effects of rivers on rabies transmission (Smith et al., 2002).

1.4.2 Tools and data sources for spatial epidemiology

The previous twenty years have seen a dramatic increase in the volume of spatial epidemiological research. These increases can be attributed directly to the introduction of ESRIs ArcInfo Geographical Information System (GIS) in the 1980s and in the 1990s the more user-friendly ArcView which made mapping a tool widely available to epidemiologists. Use of GIS has been fuelled by massive increases in the availability of digital spatial data. The development of relatively cheap hand held Global Positioning Systems (GPS) has provided a means for gathering very accurate (within 5m) case data in the field. Digital landscape data has also increased in volume and availability, in particular in the UK to a point where EDINA (www.edina.ac.uk) is licensed to distribute all Ordnance Survey (OS) map products free of charge to academic users. However, the distribution and spread of viral diseases is dependent on the distribution of information on the host population. In both animals and humans this is very dynamic and difficult to monitor remotely. Therefore, developing an accurate map of the host demography is essential for performing spatial analysis on a viral disease.

1.5 Outline of this thesis

This thesis will utilise the spatial epidemiological methods described above and apply them to the 2001 FMD epidemic to develop spatial measures of FMD risk. Once these have been developed they will be applied to develop retrospective risk maps of the

2001 FMD epidemic and prospective risk maps applied to the 2007 FMD outbreak in Surrey (DEFRA, 2007c). These analyses will require both detailed spatial, demographic and epidemiological datasets on the 2001 FMD epidemic as well as spatial and demographic datasets on farms in the UK. These will be developed at the start of the thesis. Therefore, the thesis will contain the following chapters:

1. **Analysis and cleaning of FMD data.** Epidemic data gathered during the 2001 epidemic will be evaluated to identify potential inaccuracies or inconsistencies.

2. **Building, analysis and cleaning of demographic data.** Assembling an accurate inventory of all farms in the UK and accuracy assessments of both the spatial components and demographic (numbers of animals) components in terms of their compatibility with data from the 2001 FMD epidemic.

3. **Risk factors for holding level susceptibility.** Epidemiological and demographic risk factors for susceptibility will be identified and analysed using a case-control approach. This will be a spatial correlation analysis with predictors including point sources (locations of disease introduction) and local predictors (livestock densities).

4. **Risk factors for holding level transmission.** The pattern of transmission events during the 2001 epidemic will be reconstructed to assess which IPs infected other farms. Using this data a model of the likelihood of infecting other holdings will be developed using epidemiological and demographic risk factors.

5. **Barriers to FMD transmission.** Fine scale analysis of known transmission events in a case-control framework will identify potential landscape features which may influence disease transmission.

6. **Risk mapping of FMD.** The models developed in 3,4 and 5 will be combined using a GIS to develop risk maps of risk of FMD. Both retrospectively for the 2001 UK FMD epidemic and prospectively for the UK 2007 outbreak.

Chapter 2

FMD data

2.1 Introduction

Data on the 2001 FMD epidemic were recorded in two databases at separate institutions. Many studies (for instance (Ferguson et al., 2001b,a; Honhold et al., 2004a,b; Gloster et al., 2005; Haydon et al., 2003; Kao, 2003; Keeling et al., 2001, 2003; Lawson and Zhou, 2005; Matthews et al., 2003; Savill et al., 2006; Taylor et al., 2004; Tildesley et al., 2006; Wilesmith et al., 2003)) make reference to the use of FMD data gathered during the 2001 epidemic. However, few studies attempt to validate the quality of the data, or compare datasets where data are duplicated (Keeling et al., 2001; Savill et al., 2007a). Savill et al. (2007a) evaluate the impact of data quality on their attempts to model changes in transmission potential over the course of an IP's infection period. The authors quantify the effect of missed infections and incorrectly estimated infection dates on their analysis. However, there have been no other analyses of these data.

This chapter will evaluate the different sources of FMD data for the 2001 GB epidemic. Where data have been estimated the accuracy of these estimates will be evaluated and where possible corrected. Based upon this, these analyses will develop a single dataset comprising all culled farms (IPs and non-IPs) including the following fields for both IPs and non-IPs:

1. County-Parish-Holding (CPH) number.

2. A coordinate.

3. The reason for the holding being culled.

4. The date of slaughter.

5. The numbers of animals by species which were slaughtered.

Further fields to be identified for IPs only:

1. Estimated date of infection.

2. Date of reporting.

3. Most likely source of infection.

4. Results of laboratory diagnosis.

These data will be extracted from the DEFRA Disease Control System (DCS) database and the Veterinary Laboratories Agency (VLA) FMD epidemiology database and cleaned.

2.2 Data management during the 2001 FMD epidemic

Epidemic management and much of the data gathering was organised locally by Disease Control Centres (DCCs), of which 16 existed for some period of the epidemic. A DCC was established in an area when a case was identified in that area (National Audit Office, 2002) and as a result the locations of DCCs generally follow the locations of seed IPs. New cases are generally allocated to their nearest DCC, although national boundaries and to a lesser extent county boundaries can be seen to influence this (Figure 2.1). As a result IP clusters can generally be grouped according to DCC. However, when administrative boundaries (national or county boundaries) intervene such as is the case for the Carlisle/Ayr DCCs, natural clusters can be broken up (Figure 2.1).

The FMD dataset used comprises only data for the epidemic in GB and therefore the remainder of the thesis will be solely concerned with analysis of the GB epidemic. This data is an amalgamation of two databases (DEFRA, 2007a):

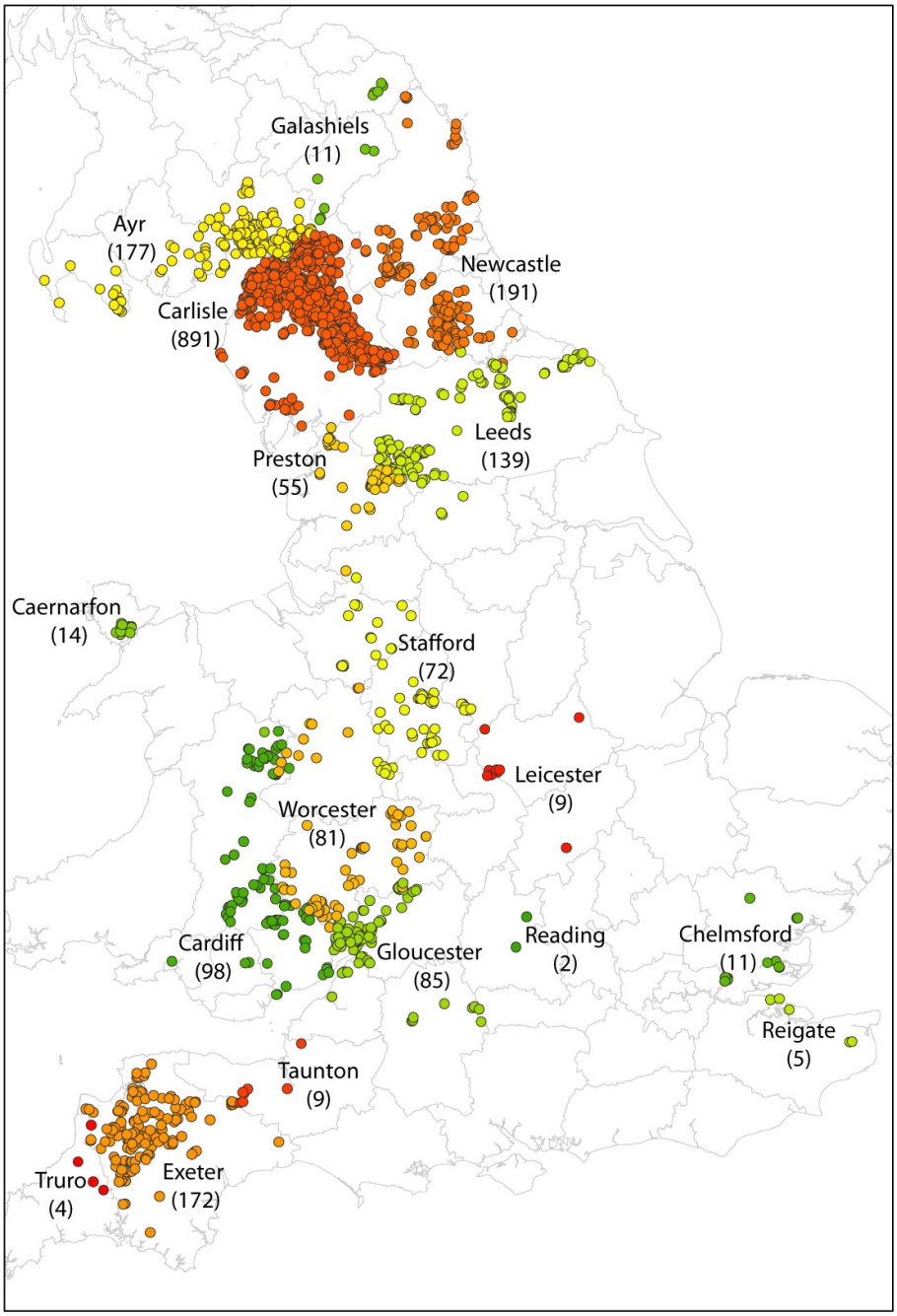


Figure 2.1: Distribution of IPs coloured by their respective DCCs. The names of the location of the DCCs are alongside the IPs and the number of IPs assigned to each DCC is in brackets below the name. IPs are overlayed on county boundaries.

1. **DEFRA Disease Control System (DCS) database:** Established to administer the day to day management of the epidemic. This contains records of all premises culled for Disease Control purposes (LWDS data were held elsewhere). The data includes:

(a) Holding details: CPH number, address, OS map reference of the farm, administering DCC.

(b) Details of slaughter: Numbers of animals by species[1], slaughter date.

(c) Details of reason for slaughter: In the cases of DCs, which IP it is linked to and why the holding is at elevated risk. If these tracings prove to be infected these DCs were reclassified as IPs.

2. **Veterinary Laboratories Agency (VLA) epidemiology database.** The database of all IPs. This includes:

   (a) Slaughter details: Numbers of animals culled by species.

   (b) Timeline of the IP: Dates of reporting, confirmation, slaughter and estimated date of infection.

   (c) Results of on-site epidemiology: Dates of oldest lesion (from which the infection date is estimated), the species initially infected, most likely source IP, most likely route of transmission. This was possible for the majority of holdings, however there are some missing data in particular for IPs identified by serosurveillance.

   (d) Results of Pirbright diagnostics: The results of tests for FMD antigen and antibody conducted by the IAH (Pirbright).

Some data fields for IPs are duplicated between the DCS and VLA databases. Where this is the case the DCS records will be used as this will ensure consistency between the IP and non-IP culls. Preliminary cleaning of the data was carried out by Dr DJ Shaw at the University of Edinburgh, further steps in cleaning the data will be described in this chapter.

The culled holdings in the DCS database can be broken down into three groups based upon reasons for which some or all of the animals of a holding were culled:

1. **Infected Premises (IP)**: (2026 farms) Premises on which animals have been identified based on clinical or laboratory diagnosis as having or having had FMD.

[1]This is a record of the number of animals for which compensation was paid. Only adult animals were compensated, the value of animals born since the start of the epidemic was included in the value of the mother and therefore were not recorded in these numbers (National Audit Office, 2002)

2. **Slaughter on Suspicion (SOS)**: (250 farms) Introduced on the 24th March to slaughter animals in cases where the veterinarian "felt unable to confirm a case on clinical grounds but was equally unable to be sure there was no infection" (Anderson, 2002, p.95). Prior to the 24th March no action would be taken in such cases, pending laboratory confirmation - under the SOS policy animals were slaughtered upon declaration of SOS status and the status of the farm changed to IP if the samples were positive - this was the case in 16% of SOS culls (National Audit Office, 2002). No preemptive culling of linked holdings would be carried out unless infection was confirmed (Thrusfield et al., 2005a).

3. **Dangerous contact**: (8570 farms) This category includes all holdings on which no evidence was found indicating the presence of FMD but it was felt that the holding was at elevated risk of harbouring pre-clinical infection. The reasons for a holding being declared a DC were recorded in a DCFLink column in the DCS database and can be divided into 5 categories:

   (a) **Traditional dangerous contact (DC)**: (1423 farms) Premises culled because they were linked in some way to an IP. For instance animals, people or vehicles moving between a farm with infected animals to one apparently uninfected.

   (b) **Contiguous Premises (CP)**: (3619 farms) Officially introduced on the 27th March to include premises with fields or buildings neighbouring an IP. The CP culling was relaxed from the 26th April by allowing the exemption of some cattle and rare breeds from culling and more local control in CP culling (Honhold et al., 2004b). This gave veterinary inspectors power to cull only parts of a holding where it was felt the entire holding had not been exposed (National Audit Office, 2002). The CP culling was never carried out fully, so in many cases not all contiguous parcels of land were culled (National Audit Office, 2002).

   (c) **3km cull**: (2980 farms) The cull of 700,000 sheep on 2,000 flocks in north Cumbria and south west Scotland was approved by MAFF on the 15th March and formally implemented on the 22nd March (National Audit Of-

fice, 2002). These holdings lay within 3km of an IP and were thought to be at elevated risk of already being infected from the initial seed at Longtown market (National Audit Office, 2002; Thrusfield et al., 2005b). The 3km cull ended in mid May and was never implemented fully in Cumbria (National Audit Office, 2002).

(d) **Local cull**: (280 farms) Principally in North West Wales where all farms which had purchased animals from the Welshpool market while it was infected (one of the early nodes from which infection spread) were culled.

(e) **Other or combination**: (268 farms) In most instances this is farms for which there was more than one reason for culling it. Usually it was a contiguous farm which had also been in contact with an IP via some other route such as personnel exchanges, which would qualify it as a CP and a DC. Additionally, this category includes 38 holdings whose animals were found positive for FMD antibodies in PZ and SZ serosurveillance but were not classified as IPs (Section 1.2.4).

The proportion of animals by species taken in each of the cull types varies greatly (Table 2.1). The two main species culled were cattle and sheep but within these groups there is variation in the reason for their culling. Over 50% of the cattle culled were culled on IPs compared to under 30% of sheep, the principal reason for this difference is the 3km cull of sheep (Table 2.1). This reflects the 3km cull being predominantly a cull of sheep and pigs.

2.3 Designating an IP

As discussed above, a farm could be designated as an IP on clinical grounds or following laboratory analysis of samples. Farms were diagnosed on clinical grounds following inspection of animals by a veterinary inspector. Clinical signs of FMD were discussed with MAFF officials at the Page Street headquarters before the farm was officially declared as an IP. Veterinary inspectors would visit a holdings for three reasons (National Audit Office, 2002):

1. Reports of suspect disease from farmers or their private veterinarians.

Cull	Total Farms	Cattle (%)	Sheep (%)	Pigs (%)	Goats (%)	Total Animals
IP	2,026	294,716	974,785	20,475	865	1,290,841
	(18.7)	(50.7)	(28.0)	(14.0)	(33.6)	
SOS	250	12,659	109,967	2,334	295	125,255
	(2.3)	(2.2)	(3.2)	(1.6)	(11.4)	
DC	1,423	60,577	401,131	32,628	102	494,438
	(13.1)	(10.4)	(11.5)	(22.3)	(4.0)	
CP	3,619	192,387	1,001,272	54,105	695	1,248,459
	(33.4)	(33.1)	(28.7)	(37.0)	(27.0)	
3km	2,980	11,082	884,151	4,373	559	900,165
	(27.5)	(1.9)	(25.4)	(3.0)	(21.7)	
Local	280	1,532	54,302	20,145	14	75,993
	(2.6)	(0.3)	(1.6)	(13.8)	(0.5)	
Other	268	8,846	61,006	12,085	47	81,984
	(2.5)	(1.5)	(1.7)	(8.3)	(1.8)	
	10,846	581,799	3,486,614	146,145	2,577	4,217,135

Table 2.1: The number of animals removed in each cull category. The percentages are percentages of the column total.

2. Direct contact tracings from other IPs.

3. From patrol visits within the PZ of other IPs.

In most instances following declaration of an IP, samples of tissue and/or blood were taken from suspect animals for laboratory diagnosis at the IAH FMD World Reference laboratory at Pirbright. The testing strategy employed was described in the introduction (section 1.1.3). In addition to analysis of samples from clinical cases of FMD, samples were analysed from DCs and as a result of serosurveillance. In total there were three reasons by which samples of blood or tissue (blood in the case of animals with no or few clinical signs of infection) were taken:

1. Samples from suspect cases. Such holdings could follow two paths:

   (a) The holding was declared an IP on clinical grounds and samples taken to support this diagnosis.

   (b) Samples were taken due to uncertainty in diagnosis and would only be declared an IP upon laboratory confirmation. Such holdings were later culled under the SOS scheme prior to the analysis of samples.

2. Samples taken from holdings culled as Dangerous Contacts.

3. Samples taken from animals in serological surveillance to demonstrate disease freedom.

IPs were either diagnosed based on solely clinical diagnosis, solely laboratory diagnosis or both clinical and laboratory diagnosis. From the data it is not possible to identify IPs which were declared solely on laboratory diagnosis, but those diagnosed on just clinical diagnosis can be identified.

Samples were tested for antibody and/or antigen and the results of the tests conducted are presented in Table 2.2. From Table 2.2 it can be seen that a majority (over 63% of samples) were positive for antigen and all but 2 of these were not subsequently tested for antibody. Of the remaining IPs, 239 (11.8%) were negative for both antibodies and antigen, a further 145 of those which tested negative for the antigen have not been tested for antibodies to FMD. Furthermore 310 (15.3%) samples have not been tested for antigen or antibody. Of the negative samples, 177 (74.1%) were solely from sheep.

		Antibody			
		+	-	no test	Total
Antigen	+	1 (0.05)	1 (0.05)	1,271 (62.7)	1,273
	-	15 (0.7)	239 (11.8)	145 (7.2)	399
	no test	32 (1.5)	12 (0.5)	310 (15.3)	354
	Total	48	252	1,726	

Table 2.2: The results of antigen and antibody tests for FMD. Numbers in brackets are the number as a percentage of all 2,026 IPs.

Statistical analysis of laboratory diagnosed IPs has been conducted by McLaws et al. (2006). The authors look at relationships between whether the IP was laboratory negative compared to positive, the type of surveillance which detected the IP, the presence of other IPs within 3km, suspect species, age of lesion, time of reporting relative to the epidemic and local DCC. All variables were significant in univariate analysis and the authors conclusions from multivariable modelling were (data from McLaws et al. (2006)):

1. The odds of being laboratory positive was higher if the disease was suspected in cattle rather than in sheep (odds ratio (OR)=4.62, 95% CIs=3.23, 6.61).

2. The odds of being laboratory positive was lower when the disease was detected by active (instigated by disease control authorities) rather than passive (instigated by the farmer) surveillance (OR=0.22, 95% CIs=0.13, 0.37).

3. Both of the above relationships are decreased when the disease was detected in cattle by active surveillance (OR=0.41, 95% CIs=0.22,0.77).

4. The odds of being laboratory positive is higher if the lesions on the holding are less than 3 days old compared to holdings with lesions older than 3 days (OR=1.62, 95%CIs=1.21, 2.18).

5. The odds of an IP being laboratory positive are higher if the case was reported during the first month of the epidemic or after the end of the second month (defined as the tail) (first month OR=3.26, 95% CIs=2.09,5.08, tail OR=2.98, 95% CIs=1.51,5.86). This probability is further increased if the case was reported by active rather than passive surveillance (OR=2.60, 95% CIs=1.14,4.41)

McLaws et al. (2006) report significant variations between DCCs, however due to confidentiality issues they do not publish these results. Preliminary analysis by Dr D.J. Shaw (personal communication) of laboratory results at the county level shows that the percentage of laboratory positive IPs was lower in counties falling within the Gloucester, Worcester and Stafford DCCs (less than 40%) compared with around 70% for Cumbria (Carlisle DCC) and around 90% for North Yorkshire (Leeds DCC). McLaws et al. (2006) do not consider those IPs which were not tested in the laboratory, however further analysis could have been used to identify whether these holdings are more similar to negative or positive IPs.

The analysis of McLaws et al. (2006) shows that laboratory negative IPs were not occurring at random and were more prevalent among sheep farms where the disease is harder to detect. Furthermore, laboratory negative IPs occurred at the peak of the epidemic when resources were stretched and opportunities for validating diagnosis would have been fewer. The relationship with certain DCCs, particularly DCCs in the Welsh borders region (Figure 2.1) where IPs were more sparse suggests that there may have been some pseudo epidemics where it appeared that an epidemic was occurring but in reality very little virus was circulating.

2.4 IP data evaluation

The VLA epidemiology data contains three fields which are estimated values or best guesses. These fields are; the estimated date of infection, the most probable source of infection and the most likely route of transmission. These fields will be evaluated in turn with the exception of the route of transmission which is not used in this thesis and therefore will not be evaluated.

2.4.1 Estimated date of infection

In most instances estimating a date of infection was based upon the ageing of lesions. Veterinarians estimated the age of the lesions and a figure was added to this to represent the latent pre-clinical period. There are four implicit assumptions which must be met for this to generate an accurate date of infection.

1. There are lesions present on the IP. An IP could be confirmed in the laboratory upon positive antigen or antibody tests. The infected animals on such holdings may be sub-clinical or recovered in which case there will be no visible lesions.

2. Where lesions are visible, that all potentially infected animals were inspected to find the animal with the oldest lesions. This is further complicated by holdings on which the animal with the oldest infection has recovered.

3. The inspecting veterinarian has the knowledge and experience to accurately age a lesion.

4. The period between infection and the development of clinical signs is the same for all infections.

To generate an accurate estimate of the date of infection all four of these assumptions must be met. There is a small subset of IPs for which a precise date of infection can be found such as farms which became an IP upon the arrival of an infected batch of animals.

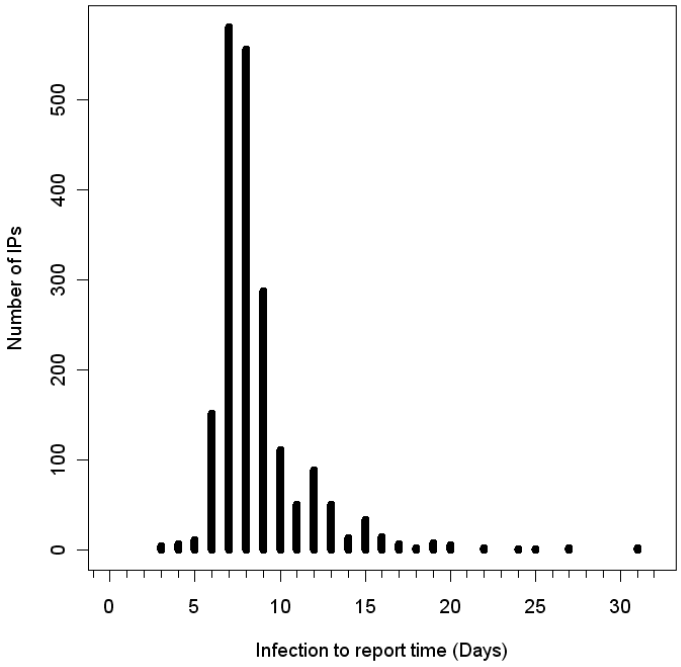
Dates of infection were estimated for all bar 37 IPs. The distribution of infectious periods and period from infection to reporting is shown by Figure 2.2. There is some evidence for a seven day effect in the infection to reporting data shown by the mode

at 7 days, one day before the median and 1.65 days before the mean value. This could suggest that infection dates were estimated as being a week previously, however the seven day peak in Plot 2.2(a) is not an outlier, which suggests that this is genuine rather than an artefact of estimation. The modal value for infection to slaughter time is 9 days, whilst the median is 10 and the mean 10.2 days. The distribution has a negative skew suggesting that the majority of infections are detected soon after the development of clinical symptoms whilst a minority persist for up to four weeks in near sub-clinical state before detection.

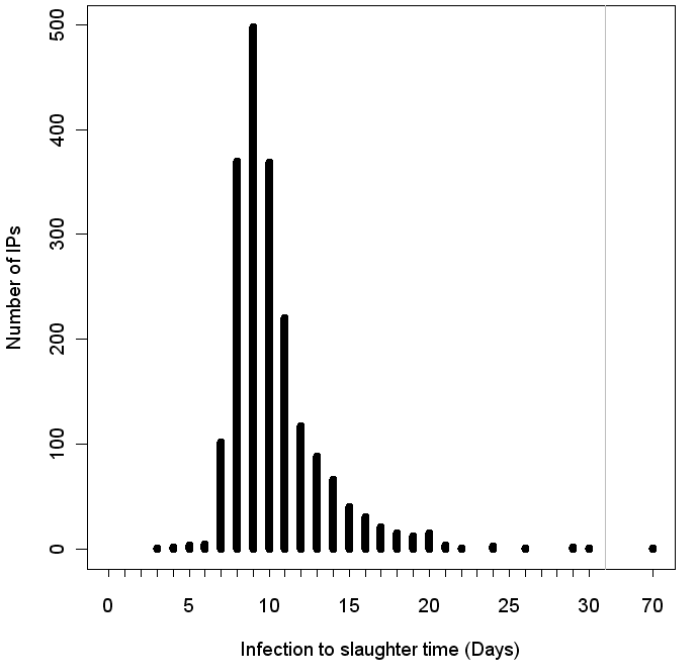
As a result of the weaker clinical signs of infection in sheep there is a delay in detection, and the time from infection to slaughter is significantly greater in sheep compared to cattle ($F_{1,1875}$=164.57, p<0.001, ANOVA test, only holdings on which cattle or sheep were the species initially infected and the infection occurred after the NMB are analysed). The median delay from infection to slaughter in cattle is 9 days compared to 10 in sheep, this difference is further exaggerated in the upper quartile of the distribution of values for sheep (Figure 2.3).

Plotting infection to slaughter times against the epidemic timepoint at which the holding is estimated to have been infected (Figure 2.4), shows that in the early stages of the epidemic infection to slaughter time was high and it gradually declined as the epidemic progressed.

Of the 37 IPs for which there is no infection date 36 are sheep infections and of the 37, 29 were positive upon serological screening for antibody, only 8 were checked for antigen. This suggests that the IPs without an infection date were identified during serological surveys and are unlikely to have had animals with lesions by which the infection can be aged. As a result it is likely that these holdings were infected for longer than the median period of 10 days as there must have been sufficient time for a number of animals to recover. However, it would be prudent to underestimate rather than overestimate the likely infectious period for holdings because the infectious period will be used to evaluate sources of infection and a longer infectious period increases the likelihood of misidentifying a source. Therefore for the IPs which have no infection date the median value for sheep infections of the slaughter date minus 10 days will be used. As these IPs were predominantly the result of serological screening



(a) Infection to reporting



(b) Infection to slaughter

Figure 2.2: Frequency distributions of infection to reporting and infection to slaughter time for the 1989 IPs with estimated dates of infection.

many were reported as IPs after they were slaughtered. Therefore using the report date to generate a date of infection may result in an infection date which is later than

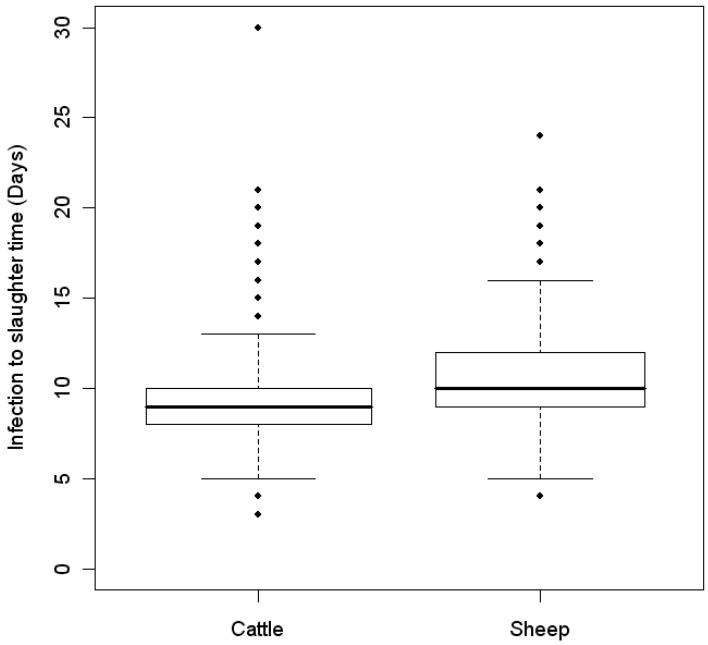


Figure 2.3: Boxplot of infection to slaughter time by species initially infected for 1876 IPs infected after the NMB for which an infection date was available and cattle or sheep were the species initially infected. Box widths represent the number in each group, whiskers represent ±1.5 times the interquartile range, the remaining data points are outliers outside this whiskers.

the reporting date, therefore the slaughter date was used.

2.4.2 Evaluating sources of infection

The VLA database contains a field identifying a likely source of infection for IPs. These were identified by veterinarians and epidemiologists on the ground using information obtained through interviews with the farmer. The DCS database contains a record called the DCS Link which concerns visits to premises where a previous IP had prompted the visit (DC tracing), the IP which prompted the visit and a reason for the visit is recorded in the DCS. Haydon et al. (2003) have constructed epidemic trees of transmission routes using the VLA data using a bootstrap algorithm to fill in gaps in the data. However due to the uncertainty in the data the use of the bootstrap algorithm resulted in wide confidence intervals for the sources of infection.

By combining the VLA and DCS datasets, sources could be identified for 1181 holdings. For the 221 of these for which there was both a DCS and a VLA source,

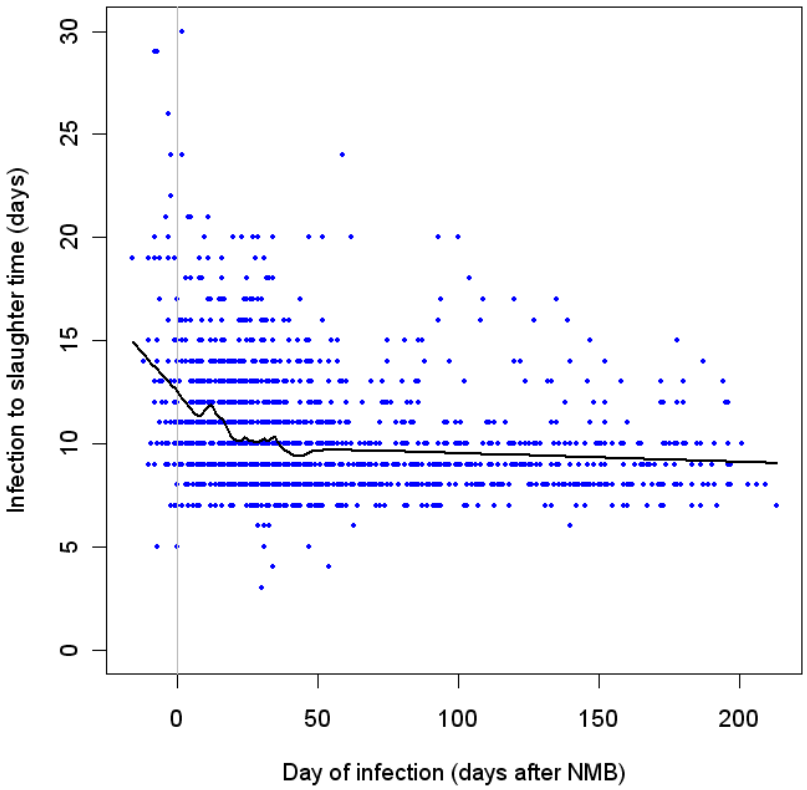


Figure 2.4: Infection to slaughter time against the number of days after the NMB that the holding became infected. The blue points are the data points and the black line is a trendline generated using the supersmooth function in R (R Development Core Team, 2004). Only data for holdings on which cattle or sheep were the species initially infected are plotted.

these sources agreed in 126 (55.6%) instances and did not agree in the remaining 95 instances (Table 2.3). A further 11 holdings had a DC link, however this holding number was the same as the *case* and these data were excluded. The accuracy of these identified sources will be evaluated.

		VLA		
		Y	**N**	
DCS	**Y**	221 (126)	224	**479**
	N	702	860	**1547**
		946	**1080**	

Table 2.3: Source of infection data for IPs broken down by data source (either VLA or DCS) and whether (Y) or not (N) that dataset recorded a source of infection. The value in brackets on the YY cell is the instances in which these sources matched.

Sources of infection for an IP were identified from a number of candidate IPs. Candidate IPs are IPs which could possibly have infected the IP in question. Four

factors will assist in identifying the likely sources of infection from the set of candidate IPs:

1. Some clear route of infection such as receiving infected animals from another IP.

2. Rapid identification of IPs. All candidate IPs must be known in order to accurately identify the most probable source.

3. The number of candidate IPs. Fewer candidates will increase the probability of selecting the correct source.

4. The resources are available to carry out the necessary epidemiological tracings to establish the most probable source.

During the early stages of the epidemic many holdings were infected via movements of infected livestock which provides a clear route of infection and, in addition, there were few candidate IPs. As a result of this, farms infected before the NMB have a greater proportion of identified sources. Of the 78 IPs estimated to have been infected before the NMB, 67 (85.9%) have a source of infection identified, compared to 44.8% of the 1948 IPs infected after the NMB (Table 2.4). This result is highly statistically significant (χ^2_1=45.4, p<0.001).

	Source data			
	VLA	**DCS**	**Both**	**Neither**
pre-NMB	60	1	6 (5)	13
post-NMB	642	223	221 (121)	860

Table 2.4: The origin of source of infection data (DCS or VLA) for IPs against whether the IP was infected before the NMB.

The patterns noted above and the limiting factors for identifying sources are matched by the daily proportions of IPs with identified sources (Figure 2.5). There is an initial peak in the number of VLA identifications as there are few IPs and many are infected by movements. This is followed by a trough as the epidemic peaks. The trough is due to the number of candidate IPs and resource limitations. The line flattens as the epidemic is brought under control, although there is a lack of epidemiological information to reliably identify many sources. This contrasts with

the DCS DC link, during the early stages of the epidemic, when few sources were identified by DC tracing (Figure 2.5) due to a lack of resources. As the epidemic is brought under control and resources (particularly personnel) become available, DC tracings increase and more sources are identified by this route. Throughout the epidemic, however more holdings had a VLA source than a DCS source.

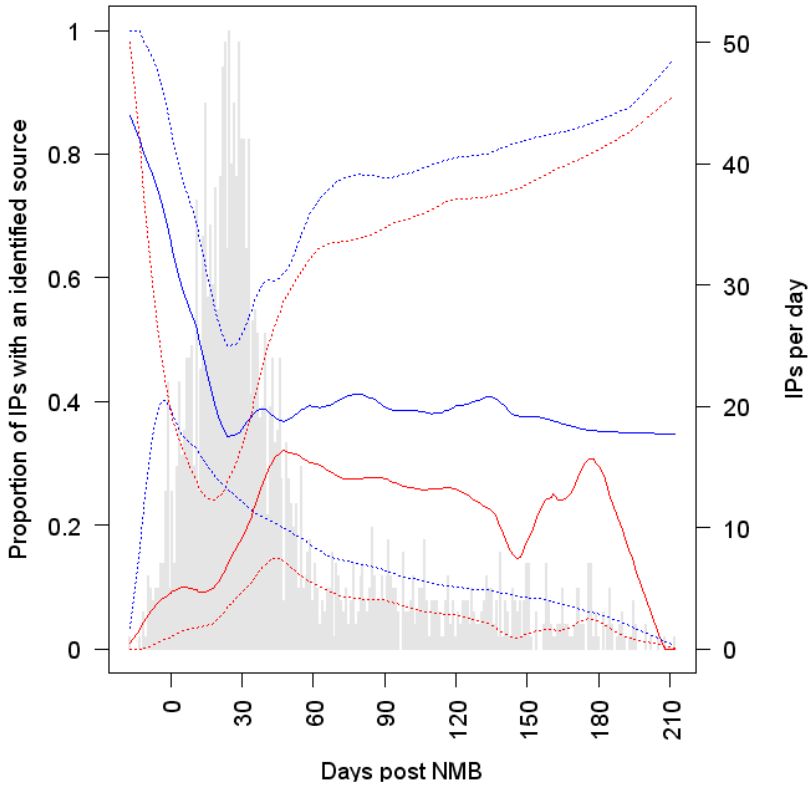


Figure 2.5: The proportion of holdings with an identified source for each day of the epidemic for DCS sources (red line) and VLA sources (blue line). Dashed lines are 95% CIs and the grey bars are the daily number of IPs. CIs for the proportion data were generated using the Hmisc package in R (Harrell, 2007). All lines have been plotted using the supersmooth function in R.

The reliability of identified sources can be evaluated by testing whether the *sources* meet two criteria (Figure 2.6):

1. The *source* must have been infected before the *case* was infected.

2. The *source* must have been slaughtered after the *case* was infected[2].

[2]The virus can survive for up to 60 days in the environment (Bartley et al., 2002) and transmission can occur through fomites so this criteria is not strictly correct. It has been assumed that there was no post-slaughter transmission

Further allowance was made to accommodate estimation errors in the dates of infection so a window of ±2 days was applied to these. Figure 2.6 illustrates the process of validating *sources* of infection:

1. **IP1** could have been infected by the *source*, as it was infected after the *source* was infected, and the *source* was slaughtered after IP1 was infected.

2. **IP2** was infected after the *source* was slaughtered. This remains the case when 2 day confidence intervals are applied to infection dates. This will be referred to as *slaughter error*.

3. **IP3** was infected before the source (including once confidence intervals are applied) and therefore could not be a daughter IP of the source. This will be referred to as *infection error*.

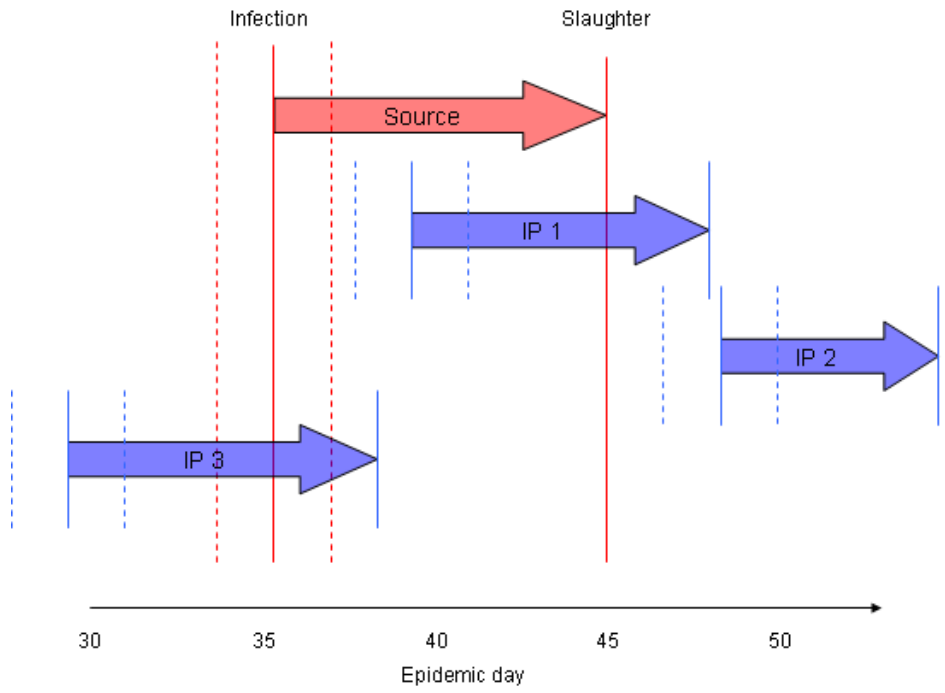
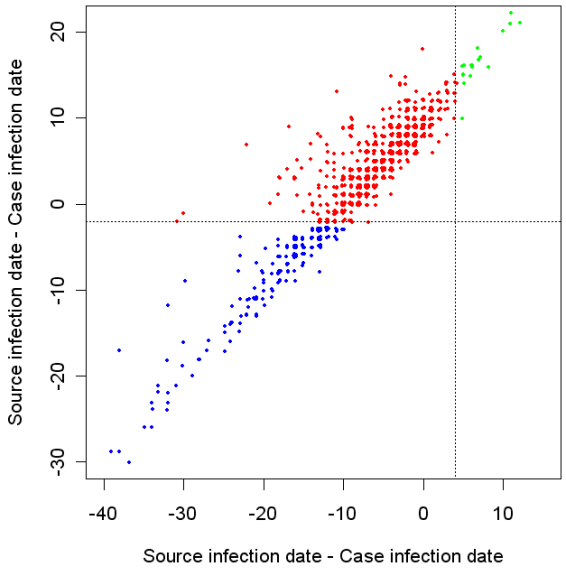


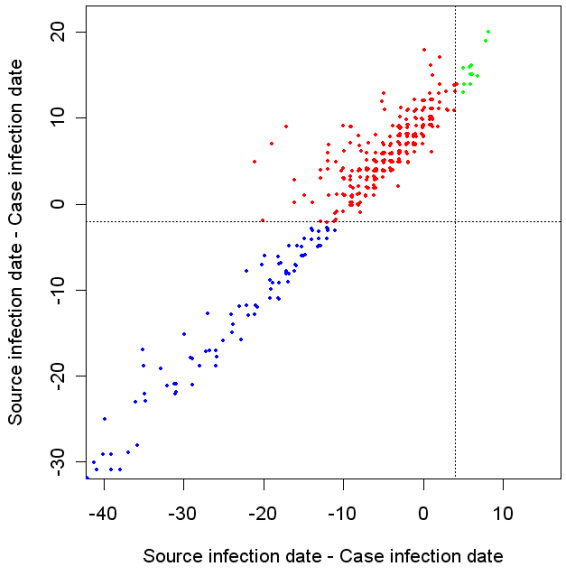
Figure 2.6: Examples of the errors caused when identifying sources of infection. Dashed lines represent the ±2 day sensitivity windows applied to the infection dates.

Evaluation of the positions of *cases* relative to the infectious windows of their *sources* shows that VLA source data identified sources with far greater accuracy than DCS sources (Figure 2.7, Table 2.5). The majority of errors (20% for the VLA

and 32% for the DCS data) resulted from assigning a *source* which had been culled out before the *case* had become infected. Furthermore the time frame over which this occurs is large, with holdings being designated as a *source* when there is over a 30 day gap between the date of slaughter of the *source* and estimated date of infection of the *case* (Figure 2.7). There was relatively little error in the opposite direction, instances of a *source* being assigned which was infected after the *case* are few (Table 2.5). Although DCS links were less accurate than VLA links the greatest percentage accuracy occurred when the DCS and VLA sources agreed.



(a) VLA source



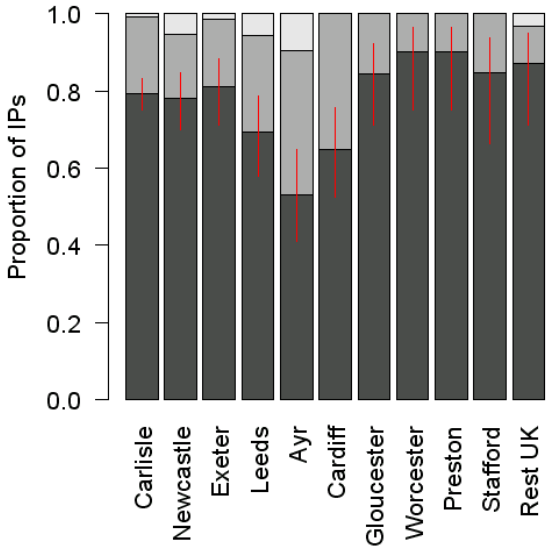
(b) DCS source

Figure 2.7: Difference in slaughter timings against difference in infection timings. The x and y axes represent dates of infection or slaughter as *source-case*. The crosshairs represent the quadrant of possible values (red points), to the right of the vertical line represents 'infection' error (green points) and below the horizontal 'slaughter' error (blue points). The crosshairs do not pass through zero as a sensitivity window was applied to the estimated dates of infection. The axes have been truncated, there were 8 instances in which there was a difference of more than 30 days between the infection of the *case* and slaughter of the *source* for the IP sources. The corresponding number for the DCS sources is 53.

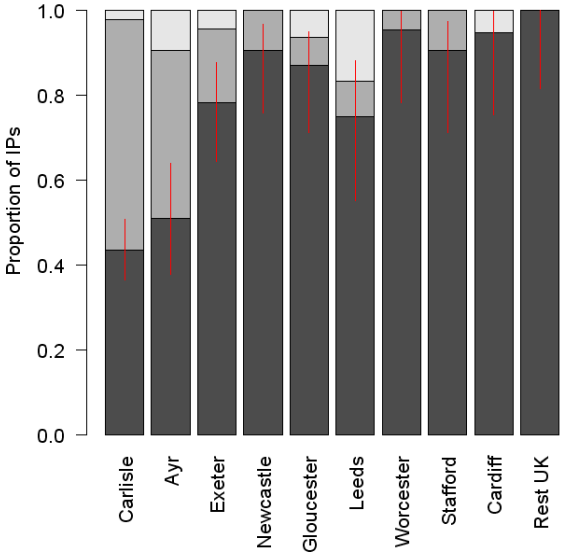
Source of error	Timing	VLA source	DCS source	DCS & VLA agree
Correct	Inf_a-2<=Inf_b+2 and Sla_a>=Inf_b-2	708 (77.3%)	289 (64.0%)	95 (85.6%)
Slaughter	Sla_a<Inf_b-2	187 (20.4%)	144 (31.9%)	13 (11.7%)
Infection	Inf_a-2>Inf_b	21 (2.3%)	18 (4.0%)	3 (2.7%)

Table 2.5: IPs with sources of infection and inaccuracies in source identification. Farm a is the source and farm b the IP. Correct, slaughter and infection sources of error correspond to the red, blue and green points in Figure 2.7. The percentages are the values as a proportion of the column totals.

Analysis of these error categories by DCC reveals considerable differences between DCCs in terms of the proportion of correctly identified *sources* (Figure 2.8). There was significant difference when both VLA links (χ^2_9=36.2, p<0.001 results of binomial generalised linear model) and DCS links were tested (χ^2_8=96.1, p<0.001, results of binomial generalised linear model). Post-hoc analysis using generalised linear hypothesis testing to compare the group means, showed significant differences between Ayr and the other DCCs for the VLA data and Carlisle and Ayr for the other DCCs (p<0.05).



(a) VLA source



(b) DCS source

Figure 2.8: Source identification accuracy broken down by DCC. Dark grey represents correct sources, lighter grey slaughter error, and light grey infection error. The bars are ordered by decreasing n with the exception of 'Rest UK'. Red lines represent 95% CIs for correctly identified sources. Only DCCs for which n>15 are included, the remainder are included in the 'Rest UK' category.

There is greater error in the identification of DCS sources than in the identification of VLA sources. This is primarily as a result of errors in the Carlisle and Ayr DCCs where the epidemic was most intense, and therefore resources having been most stretched and there are the greatest number of candidate sources. As a result of these differences a final source will be designated according to the following criteria:

1. A valid (falling into the *correct* criteria Table 2.5) VLA source. If this is the case and there is a DCS source the DCS source will be ignored.

2. A valid DCS source.

3. If neither criteria 1 or 2 are fully met no source will be assigned.

The results of this process were 953 IPs with a valid source. Of these, 721 were assigned the VLA source and 160 the DCS source, and in 108 instances of the 721 VLA sources the VLA source matched the DCS source (Table 2.5).

The final infection distances identified by this method are shown in Figure 2.9. 71.9% of post-NMB infections were over a distance of less than 3km and 89.8% over a distance of less than 10km. However these numbers are only estimates of the true infection tree as they are based upon and the 881 IPs for which there was an identified source.

Figure 2.10 shows that distances over which holdings were infected are relatively smaller in the DCCs of Carlisle, Newcastle, Leeds and Exeter. The exception is Ayr in which the median infection distance is considerably greater than the national median. Post-hoc analysis was carried out using generalised linear hypothesis testing in the multcomp package (Hothorn et al., 2007) in R. Carlisle was statistically significantly different from Worcester and Ayr ($p<0.001$) and the rest of the UK ($p<0.001$). Similar analysis breaking infection distance down by DCC shows that infection distances are relatively smaller in Cumbria than the 3km in Dumfries and Galloway and the Welsh Borders DCCs of Gloucester, Stafford and Worcester (Figure 2.10).

The source assigned to any IP will come from a set of potential sources all of which were infectious at that time. The distance to these other sources relative to the chosen source may provide information which could be used to assign a probable source for those instances of IPs missing sources. In around 50% of instances

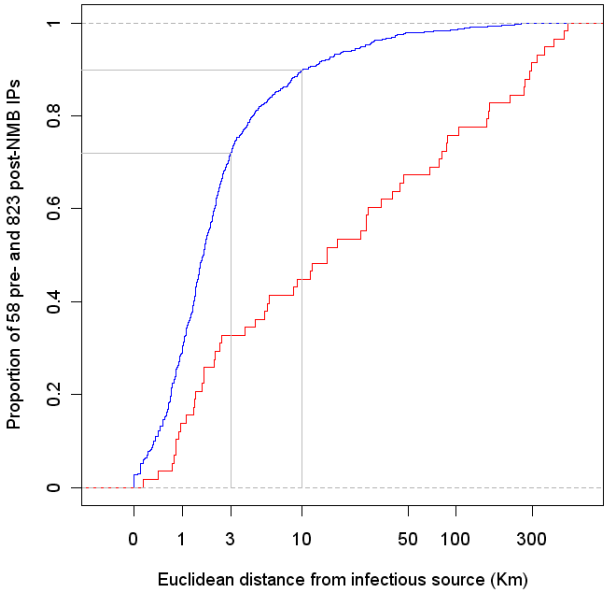


Figure 2.9: The distances from source to case for IPs infected before the NMB (red line) and those infected after the ban. Distances are plotted on a log scale.

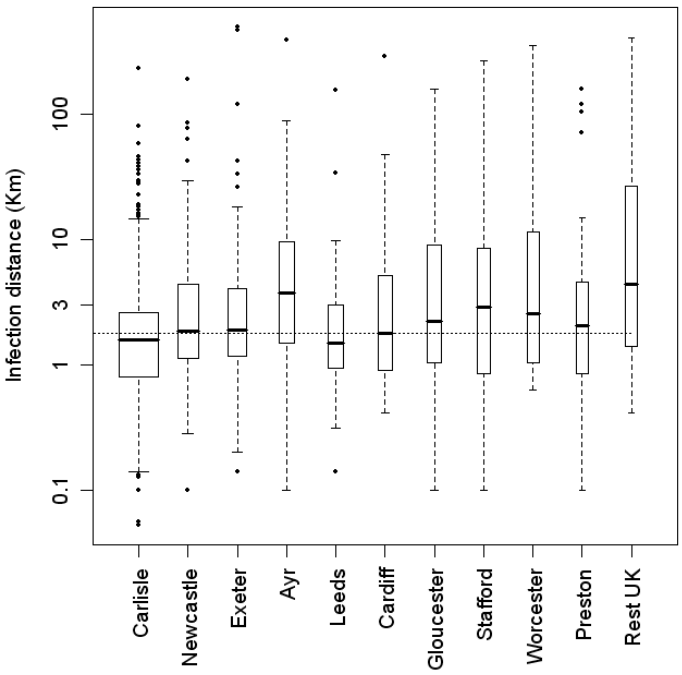


Figure 2.10: The distance to the source of infection broken down by DCC. DCCs are ordered by their number of IPs with identified sources. Box width represents the number of data points in the group. The horizontal dashed line is the national median infection distance. The boxes are ordered by decreasing number of data points in each DCC.

the source assigned was the closest possible source (Figure 2.11). The effect of infections due to animal movements can be seen from the greater proportion of holdings with 10 or more closer sources. 445 (50.5%, 95% CIs=47.2,53.8) holdings had no closer potential source, this percentage increases to 52.1% (95% CIs=48.7,55.5) when only post-NMB infections are analysed. There were 105 (11.9%) which had 10 or more closer potential sources, however 23 of the 105 were infected before the NMB.

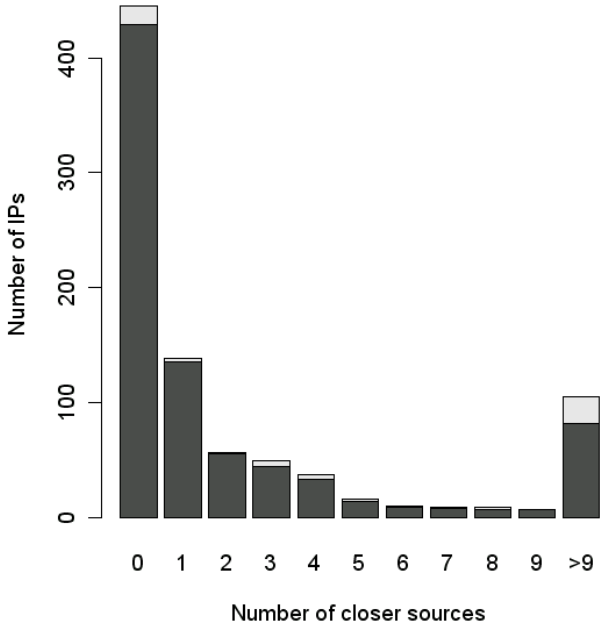


Figure 2.11: The number of closer potential sources for each IP. The black portion of the bar represents IPs infected after the NMB, the lighter section those infected before the NMB.

This chapter has described the collation and evaluation of data on the 2001 GB FMD epidemic. Sensitivities in the data to potential errors in the sources of infection and laboratory testing have been evaluated. To place these data in their context further analysis is required to develop a dataset of farms in the UK and comparison of the demographic components of the two datasets is needed.

Chapter 3

UK farming demographics

3.1 Introduction

3.1.1 Overview

In June 2000 there were 9.5 million heads of cattle, 6 million pigs and 40 million sheep registered on the UK agricultural census (Anderson, 2002). This translates to densities of 39.3 cattle, 24.8 pigs and 165.6 heads of sheep per km^2 in the UK. Given the numbers and economic value of these livestock, the importance of an accurate inventory of all farms has been identified as essential for managing and evaluating epidemics affecting livestock (Durr and Froggatt, 2002; Sanson and Pearson, 1997). Indeed the lack of a farm dataset geared towards disease control was highlighted as a failing by the "FMD Lessons to be Learned Enquiry" (Anderson, 2002, p.72). Such an inventory should include both data relating to the stock on farms, and some spatial reference for each farm (Anderson, 2002). This chapter will evaluate sources of farm data and sources of spatial farm data and compile a full farm dataset for the UK.

The UN Food and Agriculture Organisation (FAO) encourages all countries to conduct an agricultural census at least every 10 years (Stloukal, 1999) and the majority of countries outside Africa do conduct a regular census. The purpose of the census is to collect data on farm holdings, crops, livestock and agricultural inputs (Stloukal, 1999). The majority of countries outside Africa conduct an agricultural census every five or ten years (FAO, 2008). The basic unit in an agricultural cen-

sus is the agricultural holding which is defined by the FAO (1995) as 'an economic unit of agricultural production under single management comprising all livestock kept and used wholly or partly for agricultural production purposes' (FAO, 1995, p.25). The holding may be owned individually, jointly, or by a corporation and the land parcels comprising the holding may be fragmented. The agricultural census is most commonly used for demographic analysis and for forecasting agricultural output. However, census data have been used for spatial epidemic analysis for diseases such as BSE (Stevenson et al., 2000) and FMD (Keeling et al., 2001; Savill et al., 2006; Ward and Perez, 2004). Only New Zealand has developed a full database of farms specifically for epidemic management (Sanson and Pearson, 1997; Sanson et al., 1999), with georeferencing for all farms (Morris et al., 2002).

Although farm demographic data are required for many studies, there have been no comprehensive reviews of sources of demographic data on farms. The majority of demographic data has been acquired from agricultural censuses for studies of BSE in Switzerland, France and the UK (Abrial et al., 2003; Doherr et al., 2002; Stevenson et al., 2000), as well as bovine tuberculosis in the UK (White and Benhin, 2004) and FMD in Uruguay (Rivas et al., 2004, 2003). These studies have been limited by the spatial and temporal resolution of the census data. Other studies at the level of the farm have surveyed farms to gather valid and accurate data on animal numbers (Davison et al., 2003; Morignat et al., 2002). Due to the time and cost of gathering such data this is only possible on small scales, and if the farmers are informed of the benefits of the work to them (Davison et al., 2003). The following sections will describe the development of a farm database including spatial referencing for each farm and demographic data on the farms.

Some accurate means of identifying the locations of holdings has been identified as essential for epidemic analysis (Anderson, 2002; Durr and Froggatt, 2002; Sanson and Pearson, 1997). Durr and Froggatt (2002) analyse the utility of different means of georeferencing farm holdings for epidemiological and biosecurity issues. Although the authors acknowledge the value of digisised farm boundaries these are difficult to maintain in a database and require a large storage capacity and regular updates which are difficult to undertake (Durr and Froggatt, 2002). Therefore the authors

evaluate the accuracy of taking a single coordinate for holdings recorded at different locations on the holding. They conclude that the main farm building is the most appropriate location for georeferencing with a single coordinate.

3.2 Animal husbandry in the UK

This chapter describes the development of a dataset to model livestock farms and livestock production in the UK. Therefore consideration and understanding of livestock farms and livestock production is required in order to fully understand the simplifications and assumptions which are being made when modelling farm holdings.

In spite of the title this section will consider only cattle, sheep and pig production. Poultry species are not susceptible to FMD and poultry production is largely independent of production of FMD susceptible species. Commercial goat and deer production is very small comparative to the production of cattle, sheep and pigs and these species played little part in the UK FMD epidemic.

3.2.1 The structure of the farm holding

The farm holding typically comprises a main residence or farm house. Close to the farm house there will typically be a set of farm buildings including animal housing (barns, cattle sheds, pig units) and storage (for both machinery and feed). Additionally there may (although not always in the case of intensive pig farms) be a set of fields. The fields are not necessarily to be found together around the farm buildings, there may be off-fields or satellites some distance from the main holding. These may represent fields which have been purchased subsequent to ownership of the main farm unit or temporary pasture - land loaned from other holdings to accommodate additional animals on the main holding.

British farms are typically operated as family enterprises and the business is passed down through family generations. As a result the core of the staff tends to be family members who live on site. However, depending upon size and seasonal variations in labour requirements (lambing seasons for instance) the holding may

employ outside labour and indeed may share labour between several holdings. In addition to labour, machinery and feed production equipment may be shared between nearby holdings as a means of saving costs.

The nature, extent and structure of livestock production is heavily influenced by both current subsidies issued as part of the EU Common Agricultural Policy (CAP) and trends initiated by historical subsidies (MAFF, 2000; The Royal Society, 2002). Depending upon the product, these subsidies can provide up to 50% of farm income (The Royal Society, 2002). Furthermore the CAP implements certain trade regulations which influences the distribution of production within the EU as well as the profitability of production.

3.2.2 Cattle production

Cattle production can be divided into two non-independent streams: dairy cattle and beef cattle. The annual production cycle for cattle is not necessarily fixed and can occur at any time of year although many holdings aim to calve their animals in Spring or Autumn, so mating takes place 9 months prior to the desired time for calving (Allen, 1990).

The geographical distribution of dairy farming is very much driven by the distribution of good quality lowland pasture which in turn is very dependant upon rainfall. As a result there is a concentration of dairy farming in the west of England, south-west Wales and south-west Scotland (The Royal Society, 2002). In 2000, there were a total of 33,892 holdings with an average of 72 animals per holding which were actively involved in milk production (The Royal Society, 2002). However, the average number of animals per holding is increasing, whilst the number of herds is decreasing in response to cuts in CAP support (The Royal Society, 2002). As a result of the reliance on milking machinery, dairy farms tend to be intensive, although still requiring large areas for pasture.

Beef farming is also dependent upon large areas of rich pasture, although beef production can be more extensive and is therefore not as reliant on lowland areas as dairy farming. There are two systems of production; the beef-breeding system and dairy breeds. In beef-breeding system calves are reared on site by suckling the

dam. This system is more extensive and more commonly found in more upland areas (Allen, 1990). Such holdings are typically mixed cattle and sheep holdings. The animals are weaned at 7-9 months before being transferred to a specialist finishing unit before slaughter aged between 15 and 30 months. The second system involves rearing dairy cows sired from beef bulls or are male dairy calves not required for breeding (The Royal Society, 2002). Such animals are transferred away from the natural dam during the fortnight after birth to specialist calf rearers where they are suckled by a specialist multiple-suckler cow or fed milk by artificial means (Allen, 1990). Like the beef-breeders, they are slaughtered at 15-30 months old. Therefore most cows will move between farms at least once during their life, and some animals such as bulls may move regularly.

Diet for both dairy and beef cattle principally comprises live grass or silage during the winter months when grass growth is slow (Allen, 1990). The animals may also be fed artificial supplements, particularly during the finishing stages. During winter months many animals are housed indoors and released in the spring (Allen, 1990).

3.2.3 Sheep production

Britain has the largest sheep flock in Europe which has benefited from and expanded under CAP regulations (The Royal Society, 2002). Despite the large number of sheep reared in Britain, the sheep industry is the most extensive of the livestock production systems typically occupying the most marginal pasture either in lowland and upland areas (The Royal Society, 2002). There are a number of sheep breeds, and the breed reared usually depends upon the nature of the sheep farming (lowland, upland or hill) and the geographical location (Fraser, 1987). Within the UK, sheep are found throughout the west coast of England, western Scotland and Wales. Sheep are often the sole species farmed in many upland areas of the country.

The sheep production cycle a holding adopts typically depends on the desired date for sending lambs to slaughter. This is usually determined by whether the farm is a lowland, upland or hill farm (Fraser, 1987). Unlike cattle, sheep are almost exclusively fed on grass and lactating ewes require a large amount of food, therefore lambing typically takes place early in the spring to take advantage of the most prolific

grass growth (Fraser, 1987). The gestation period of a sheep is 21 weeks so tupping takes place during the previous autumn. Ewes give birth to an average of 1.1 lambs per year which are typically born indoors for warmth and hygiene. After giving birth ewes and lambs are quickly returned to pasture. Lowland lambs tend to be weaned after around 2 months and finished and sold in late spring and summer. Hill lambs are weaned later, after around 5 months and sold as store lambs to be finished on grass in lowland farms and slaughtered during the autumn and winter (Fraser, 1987).

3.2.4 Pig production

The pig industry operates very differently to cattle and sheep production. This is partly because pig production is not subsidised and EU markets are not protected by import barriers. Therefore, the industry is highly commercialised with animals kept in large herds and pig production is very sensitive to changes in the pork market. The pig industry is mainly concentrated in eastern and north-eastern England, a distribution driven by the availability of grain for feed (MAFF, 2000). The majority of pigs are housed indoors; in 2000 70% of breeding pigs and 96% of pigs being fattened were kept indoors (The Royal Society, 2002).

The industry is driven by the highly productive nature of pigs. Each sow produces around 20 piglets per year and gilts reach sexual maturity after around 170 days. Therefore a pig can give birth in its first year of life as the gestation period is 115 days (Whittemore, 2006). Pigs are held in units specialising in one particular aspect of pig production in order to minimise disease risk. A farm itself may specialise in one particular aspect of production or contain several units (The Royal Society, 2002). The structure is divided into three broad units: breeding units where piglets are kept until weaning at around 3 weeks, nursery units from 3 to 10 weeks of age and after 10-14 weeks finishing units. Pig diet consists of cereal and protein supplements (Whittemore, 2006). Until May of 2001 licensed holdings could feed a diet of swill (waste food collected from commercial kitchens) which had been cooked for an hour. However, following the suspected role of swill feeding in the beginning of the FMD epidemic this was banned (The Royal Society, 2002).

3.3 Development of the farm dataset

Farms are identified using a county-parish-holding (CPH) number which is a hierarchical identifier referencing the farm to the county at the first level and the parish at the second level. The county and parish of the holding are identified using the address of the holding and the spatial reference derived from this address (see below).

The farm dataset being developed requires 2 components: the farm demographic component (type of farm, species and numbers of livestock kept and farm size) and a spatial reference for each farm. There are several sources of these data each with various levels of accuracy and completeness, the data sources will be evaluated to ensure that the most complete and accurate dataset is developed.

3.3.1 Sources of farm data

There are three datasets in the UK which record farm data at the level of the holding in the UK:

1. **Integrated Administration and Control System (IACS; SIACS in Scotland):** This is a dataset of field boundary polygons. Prior to 2001 this data was collected by DEFRA, and after 2001 the Rural Payments Agency (RPA) in England and the Scottish and Welsh assemblies were responsible for gathering the data. Farms must register a centroid coordinate for each parcel of land for which they are claiming subsidy in May of that year.

2. **The UK farm list:** The farm list serves two purposes: it is a register of farm holdings from which the agricultural census is sent out and it is as an inventory of all parcels of land containing animals that was recorded in the early stages of the 2001 FMD epidemic (Honhold and Taylor, 2006). The data consists of the holding address, CPH number and an OS grid reference generated from the address generated using OS address point data. Furthermore, for some holdings it records species by their presence or absence.

3. **The UK agricultural census:** Is a detailed demographic survey of farms. It is conducted by DEFRA as a complete census every 10 years and as a sample

survey every year. The last full census was carried out in June 2000. The census is a means of monitoring agricultural production and tracking trends in agriculture. The census form is posted in mid-May to allow for completion by June. Farmers have a legal obligation to complete the census form and most return it quickly. For those who do not return the form a series of reminders are sent out via postcard. Accuracy of data is checked by evaluating the consistency of data and by comparing the data with pervious returns. Queries with the responses are raised by telephone. The survey form asks questions about all production, both crops and livestock, including a demographic breakdown of livestock numbers by species and the total area of the holding. Further details concerning changes in land ownership and labour are included. Each holding is recorded as one unit at one location, irrespective of how distributed its fields may be.

DEFRA record the location of holdings as a point rather than a polygon due to the issues of recording and storing polygon data for holdings (Durr and Froggatt, 2002). DEFRA generate point spatial references for holdings using four means (derived from DEFRA (2008)):

1. **IACS data:** The centroids of all parcels of land belonging to a holding are plotted in ArcView. Those centroids which fall within the same parish referred to by the farm's CPH number are identified and the average easting and northing of these centroids is taken, known as the AGR (average grid reference). The centroid of the field nearest to the AGR is used for the farm's location.

2. **RADX database:** A database of farm grid references derived from either the farm address (see below) or a coordinate gathered during a veterinary inspection.

3. **PAF (postal address file) data:** Using the registered address of the holding computer software validates the postcode against the address and where the two agree an easting and northing. As a result these data refer to what is likely to be the main farm building.

4. **Random point within a parish:** A random point within the parish is generated as the grid reference.

Grid references recorded in the census are allocated by stepping through the above stages until a grid reference within the same parish as the farm's CPH number is identified. The farm list records the RADX identifier or the PAF, which in most cases (particularly for census holdings with an accurate record of their address will be the same thing).

The livestock dataset must comprise some spatial reference for the holding and a record of the numbers of animals broken down by species. Furthermore these data must be compatible with the FMD datasets which comprise a single coordinate for each culled premises. If the animals are within 1km of the main farm building the animals are culled under the CPH and map reference of the main holding. If the animals are more than 1km away they are recorded under a different CPH number and map reference (Mansley et al., 2003). The census data is in a similar format to the FMD data except that it records entire farm holdings irrespective of the distribution of fields. The data on the farm list adds little value over the census with the exception that it records the off-fields as they were at the beginning of the 2001 FMD epidemic. IACS data records data at a much finer scale than the FMD data.

The majority of previous analyses have used agricultural census data to provide the demographic component to the analysis (Ferguson et al., 2001a,b; Kao, 2003; Keeling et al., 2001, 2003; Morris et al., 2001; Savill et al., 2006; Tildesley et al., 2006; Wilesmith et al., 2003) where such a component is required. Only two papers (Ferguson et al., 2001b; Honhold et al., 2004b) make reference to use of the IACS data and this is to deal specifically with the issues associated with fragmentation. In spite of the extra spatial data, the IACS dataset offers little of utility in this thesis over and above the point location data. The issue of farm fragmentation can be handelled more effectively using the register of parcels of land recorded in the farm list. Therefore the dataset used in these analyses will be a combination of agricultural census and farm list data.

The spatial component of the data will be taken wherever possible from the farm list address data. This represents a meaningful central geographical point of the farm

enterprise, and the main holding has been identified elsewhere as the best location from which to georeference a holding (Durr and Froggatt, 2002).

3.3.2 Building the dataset

The dataset is being developed for analysis of the UK 2001 FMD epidemic, so it should include all holdings with stock susceptible to FMD in 2001. Some of this data, particularly that relating to farm fragments is on the farm list. The farm list data can be broken down into four categories:

1. Livestock holdings which are also listed on the agricultural census.
2. Farms recorded as containing livestock on the farm list which do not appear on the agricultural census.
3. Farms with no record of livestock which were off-fields of main holdings with livestock listed in the agricultural census.
4. Farms which have no record of livestock because they have no livestock or no longer exist as agricultural enterprises.

Table 3.1 shows the numbers in these groups by cross referencing the census with the farm list. There are a total of 139,195 farms with livestock recorded on the agricultural census. In addition there are a further 43,226 farms with a record of animals on the farm list but not the census.

		Census		
		Present	Absent	Totals
Farm List	Livestock	111,560	43,226	154,786
	No Livestock	27,653	319,244	346,897
	Totals	139,195	362,470	474,134

Table 3.1: The number of holdings by data source broken down by the datasets. There were 556,994 holdings on the farm list however only 474,134 had unique CPH numbers.

The following work was required to ensure the accuracy and completeness of this dataset were maximised:

1. To identify the most accurate coordinate for each holding.

2. Using the FMD data to evaluate the 43,226 farms with livestock on the farm list but not the census and establish whether some or all of these holdings did in fact have livestock which has not been identified by the census.

3. Using the FMD data to identify and populate the proportion of the 319,244 farm list holdings with no livestock on census, or farm list holdings which were satellite fields to main holdings.

4. To assess how accurately the animal numbers assigned to each holding reflect the numbers recorded in the FMD data.

These stages of analysis and validation will form the remainder of this chapter.

3.4 Evaluation of the spatial references

This section will compare the sources of farm spatial data by comparing georeferences from the farm list and census with georeferences from the FMD dataset.

3.4.1 Assigning coordinates

There were three potential sources of coordinates for farms these were:

1. Census coordinate.

2. Map reference from the farm list.

3. A postcode.

However, there remained some 35,660 farms which could not be assigned a coordinate from any of these data sources. This is summarised in Table 3.2 which shows that there is a lot of overlap between the different coordinate sources. Furthermore, for the majority of farms for which a source was available the source could be taken from the census or farm list. This is preferable to using coordinates generated using the postcode as these were assigned to some random point in a postcode district and is not necessarily an accurate representation of the true location of the holding. The process by which farms on the census are georeferenced is validated to ensure the most accurate location is selected (section 3.3.1). Comparison of distances between

the census and farm list georeference and between census and cull data shows there is little discrepancy between census and cull but a much greater degree of error between census and farm list (Figure 3.1). However the farm list and cull datasets are almost identical, only 15 holdings have different coordinates (median difference 63.0km, lower quartile 2.5km, upper quartile 100km) and are likely to be the result of error in recording a grid reference.

Data source	Census	Farm list (livestock)	Farm list (no livestock)	Total
Total	139,195	43,226	319,244	474,134
Census	138,621	NA	NA	
Farm list	137,696	41,084	250,164	428,944
Postcode	126,277	36,936	149,078	312,291
No data	20	393	35,247	35,660

Table 3.2: Breakdown of holdings into the different coordinate types that are available. The farm list categories exclude those holdings which have already been counted in the census column.

The farm list georeference was used in preference to the census coordinate. This is because the farm list is a reference to the main farm building at which most movements onto and off the farm are likely to be centered and it is likely the be the location of the majority of the fields (Durr and Froggatt, 2002). The census coordinates are the mean of the coordinates of the fields, therefore the census coordinate location may bear little relation to the actual location of the farms premises. Indeed, the mean location could be some distance from any property owned by the farm. This coordinate was designated as a way of aggregating agricultural census data to 1km raster grids which the public can access (EDiNA). Therefore, during this process the exact location of the holding is unimportant, the coordinate location is intended as a means of representing the geographical space occupied by the farm. The following hierarchy was used to assign coordinates depending on availability: farm list grid reference, census data, and where neither of these is available, postcode. The numbers of holdings assigned these coordinates are presented in Table 3.3.

Due to rounding of OS grid references to 100m there are 80,566 (25,579 livestock, 54.987 non-livestock) overlaying holdings. These holding locations were adjusted by moving the farms by a random value drawn from a uniform distribution between -50 and 50 meters.

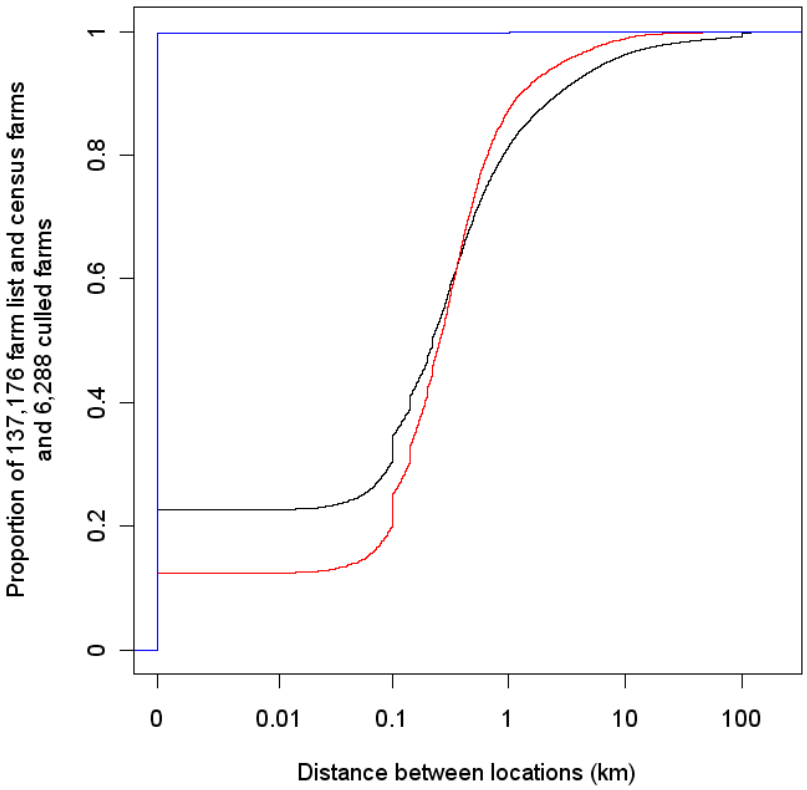


Figure 3.1: Cumulative distribution plot of distances between farm locations according to the census compared to the farm list (black line), between the DCS data and census data (red line) and between the farm list and cull data (blue line). The x-axis has been $log_{10} + 1$ transformed.

Data source	Census	Farm list (livestock)	Farm list (no livestock)
Farm list	137,696	41,084	250,164
Census	1,473	NA	NA
Postcode	6	1,849	5,521
No data	20	393	35,247

Table 3.3: The source of farm data (columns) against the source of the spatial data (rows).

3.5 Evaluation of the demographic component

The demographic components of the dataset can be evaluated by comparing these data with the FMD dataset (Chapter 2). The analysis will comprise three components:

1. Evaluation of how well the census data reflects the FMD data given that there will be movements on and off the holding and animals born (section 3.2).

2. Evaluation of the holdings with livestock present/absent recorded in the farm list but no census records. Identification of the nature of these holdings and whether they represent a substantial number of livestock farms which have been missed in the census.

3. Evaluation of the farm list holdings with no livestock recorded to identify those holdings which are satellite holdings and to populate the satellites.

The breakdown of the reason for culling farms (section 2.2) by the data source is presented in Table 3.4. The greatest percentage of culled farms were from the census (58%), followed by the farm list with no livestock (36%), followed by the farm list with animals (6%). Furthermore 4.5% of all census farms were culled, compared to 1.2% of farms on the farm list and 1.6% of farms on the farm list with livestock presence recorded (Group total row in Table 3.4).

	Farm list			
Reason for cull	No livestock	Livestock PA	Census	Total
IP	348	96	1582	2026
	(17.2)	(4.7)	(78.1)	
SOS	85	17	148	250
	(34.4)	(6.8)	(59.2)	
DC	764	105	554	1423
	(53.8)	(7.4)	(38.9)	
CP	1338	230	2051	3619
	(37.0)	(6.4)	(56.7)	
Other	111	25	132	268
	(41.4)	(9.3)	(49.3)	
3km	1177	150	1653	2980
	(39.6)	(5.0)	(55.5)	
Local	39	58	183	280
	(13.9)	(20.7)	(65.4)	
Column	3862	681	6303	10846
	(35.6)	(6.3)	(58.1)	
Group total	319,244	43,226	139,195	474,134
	(1.21)	(1.58)	(4.53)	

Table 3.4: Numbers of premises by data source broken down by the FMD cull categories described in chapter (Chapter 2). Percentages (derived from row totals) are in brackets.

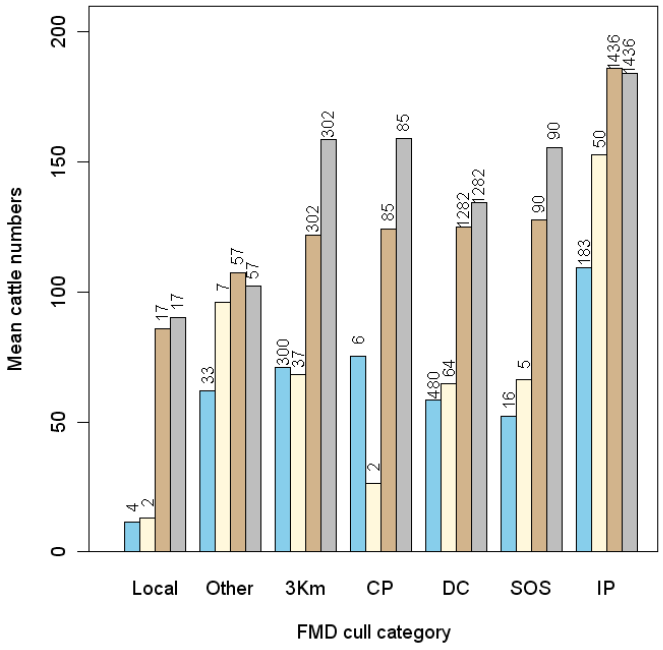
The breakdown of these data groups by FMD cull category and animal numbers is presented in Figure 3.2. The conclusions which can be drawn from this are:

1. The mean cull size was greater on farms with census data suggesting that larger farms are more likely to have census records. This is shown by the differences between the brown, yellow and blue bars.

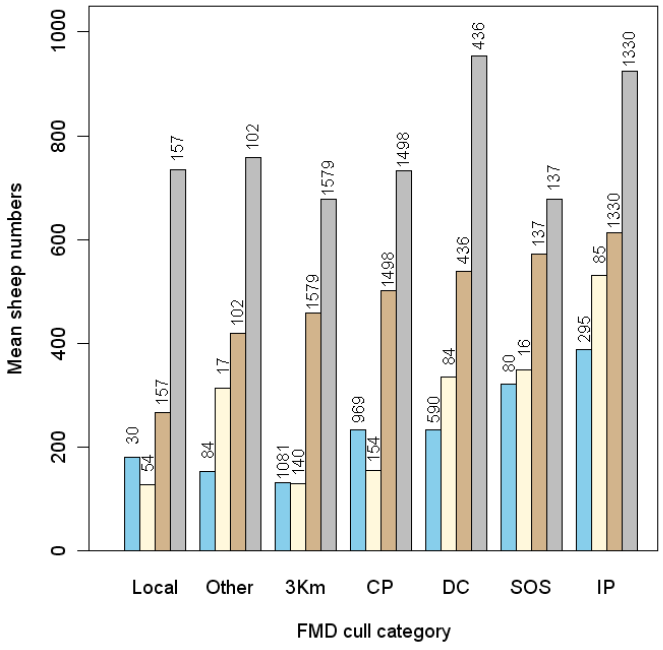
2. Farms culled which were on the census had more animals recorded during the census than the cull. This is particularly the case for sheep but also applies to cattle for some cull categories.

3. Analysis of the data on culled sheep numbers presented in Figure 3.2 against both classification on the census and the reason for the cull using analysis of variance (ANOVA) revealed that there were highly statistically significant differences ($p<0.001$) between sheep numbers for both groups of variables.

4. The relationship described above also applied to cattle, but only when cattle numbers on IPs are compared to all non-IPs and when census farms are compared to non-census farms rather than both groups ($p<0.001$).



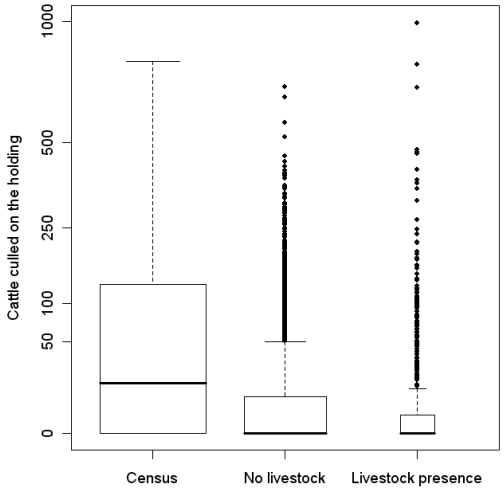
(a) Cattle



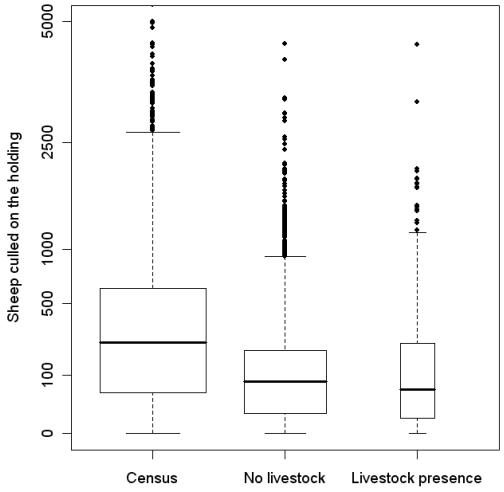
(b) Sheep

Figure 3.2: The mean numbers of animals culled (from the DCS) broken down by data source and species. The blue bar represents farms listed as non-livestock, yellow those listed with animals present on the farm list but not the census and the brown bar those on the census. For the holdings on the census, the mean number of animals recorded on the census are shown (grey bar). Numbers above represent the number of holdings in each category.

Furthermore, Figure 3.2 shows that IPs generally have substantially more animals on them than other culled farms. The distributions of farm sizes across the three data sources is significantly different for cattle and sheep (both $p<0.001$ - Kruskal-Wallis test was used due to non-normal distribution of data). However the difference between the census species numbers and numbers for holdings on the farm list is much greater for cattle than sheep. For cattle the majority (over 50%) of the farm list premises have no cattle (Figure 3.3) and of those with cattle they are typically much smaller than those on the census.



(a) Cattle



(b) Sheep

Figure 3.3: Boxplots of numbers of animals culled by data source and species. The x-axis categories correspond to whether a holding is listed on the census, listed only on the farm list with animal presence ('Livestock presence') or whether is had no animals recorded. Both plots are on a square root scale and the width of the boxes represent the numbers in the groups.

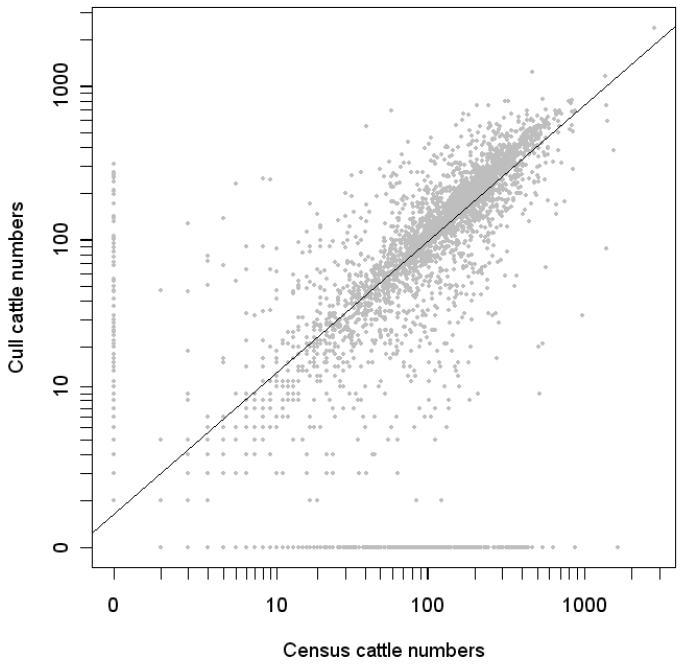
3.5.1 Differences between census and cull numbers

Animal numbers reported during the census were compared to those in the FMD cull data (Chapter 2) and differences evaluated. However, for 3km and local culls just sheep and no cattle were culled. To overcome this, sheep numbers will be compared against all cull data, but cattle numbers will not be compared with 3km and local cull data. The resulting relationship between cull and census numbers is shown by Figure 3.4.

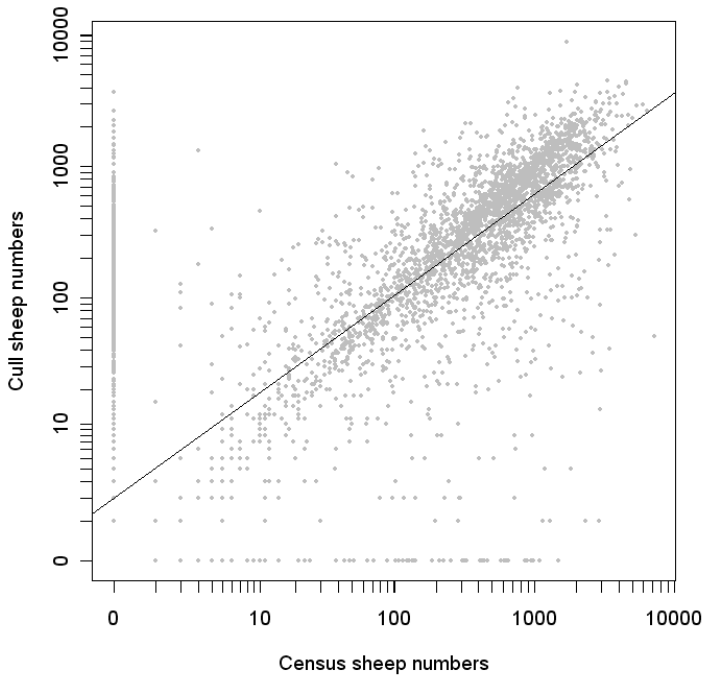
Figure 3.4 shows a positive relationship between numbers on the census and cull. However there are a large group of holdings in Figure 3.4 which have sheep culled which are not recorded in the census or have cattle on the census which were not culled. These data are presented more fully in Table 3.5 which shows that 488 (14.8%) sheep farms had sheep on the cull and not the census. 404 (11.2%) of cattle holdings had cattle on the census and not the cull. The latter can be explained by partial cattle culls on DCs and CPs (Honhold et al., 2004b). Of these 404, 324 (80.2%) were from DCs and CPs but DC and CP culls accounted for only 53.7% of these main holding culls. The large discrepancy in the sheep data may be explained by the different seasons in which the two datasets were collected. In June there would have been lambs either suckling or being weaned in upland pasture (section 3.2). In the winter however, many of these animals are moved to lowland pasture for finishing. It is possible that these data represent lowland holdings which only buy sheep in winter for fattening and potentially use their pasture for cattle (housed indoors in winter) during the rest of the year. This is not as a result of new born lambs in the spring of 2001 as new-born animals are not included in the cull data (Chapter 2).

To test the relationship between cull and census numbers it is necessary to remove the groups of holdings with zeros in the cull or census categories as these groups make the distribution non-continuous. Linear regression analysis shows both relationships are highly statistically significant ($p<0.001$) and the slopes for each are positive (0.887 for cattle and 0.772 for sheep, both on a $log_{10}+1$ scale). However, there is less residual variance in the relationship for cattle ($R^2=0.687$) compared to sheep ($R^2=0.597$).

The difference between numbers on the census minus the cull is presented in Table 3.6. Overall the cattle numbers on the census are a more accurate reflection



(a) Cattle



(b) Sheep

Figure 3.4: Graphs of animal numbers on the cull against the numbers on the census by species. The solid line is the fit of a linear regression model with values of zero removed. Both axes are a log_{10}+1 transformation.

		Cattle Cull		Sheep Cull	
		Present	Absent	Present	Absent
Census	Present	3110	404	2752	57
	Absent	94	0	488	0

Table 3.5: The breakdown of species presence recorded in the cull and census for the data which generated the plots in Figure 3.4.

of the numbers that were culled than is the case for sheep (Table 3.6). Furthermore this difference is at its greatest in Wales where sheep numbers on the census are an over-representation of the numbers present when holdings were culled, cull sheep numbers in Wales are 41.3% of the numbers recorded on the census. Many regions were statistically significantly different from zero at $p<0.05$, this might be expected as a result of annual changes to farm structure during the year as adults give birth. The differences therefore do not necessarily indicate inaccuracies in the datasets. However the results which are statistically significantly different from zero at $p<0.001$ suggests that there are inaccuracies in the data. This is underlined by the strong regional differences in numbers shown in Table 3.6. Which may reflect differences in census data collection or regional differences in farming practices such as regions in which upland farming is practiced.

Species	Region	Difference			Median	Median	Percent	
		25%	50%	75%	cull	census	Difference	p
Cattle	UK	-20	0	21	122	127	96.1	0.729
	England	-21	1	18	121	123	98.4	0.087
	Scotland	-23	7	50	180	206	87.4	0.001
	Wales	-10	2	25	70	80	87.5	0.013
	Cumbria	-25	-2	18	162	167	97.0	0.029
	D&G	-25	7	49	189	208	90.9	0.006
	Welsh Bord	-13	1	20	73	84	86.9	0.155
	Devon	-24	-1	14	101	103	98.1	0.114
Sheep	UK	-73	16	277	300	385	77.9	<0.001
	England	-67	17	245	297	372	79.8	<0.001
	Scotland	-123	-5	224	360	396	90.9	0.034
	Wales	-2	165	795	246	595	41.3	<0.001
	Cumbria	-75	12.5	254	302	350	86.3	<0.001
	D&G	-135	-11	189	353	373	94.6	0.492
	Welsh Bord	-58	38	367	267	371	72.0	<0.001
	Devon	-92	-1	118	300	364	82.4	0.509

Table 3.6: Summary of the distributions of differences between animal numbers on the census and cull datasets. The difference column is the quartiles from the distribution of values generated from census data -cull data. The percentage difference column is the cull numbers as a percentage of the census numbers. p-values are from a two-tailed Wilcoxon Rank Sum test. The regions are defined in Figure 1.9.

The differences in Wales discussed above do not explain the 15% of holdings with animals recorded on the census but not the cull. Analysis of those farms with sheep on the cull but none on the census shows that only 8.3% of sheep farms culled in Wales fall into this category compared to 12.4 and 22.4% in England and Scotland respectively. The alternative explanations are that many more animals in Wales are sent to slaughter between June and February, or that Welsh census data on sheep is inaccurate.

The utility of census data from June 2000 as a representation of animal numbers for 2001 has been evaluated by comparing animal numbers on the census to animal numbers on the cull. Regional differences were greater for sheep than cattle. The results suggest that the census greatly over predicts sheep numbers in England and Wales and that in England these differences are particularly acute in Cumbria.

3.5.2 Farms with only livestock presence recorded

Figure 3.2 demonstrated that at least some of the farms which recorded livestock as either present or absent on the farm list, had livestock during the 2001 FMD epidemic but were not recorded as having livestock in the census. This suggests that the census may not be a complete inventory of all farms with FMD susceptible livestock. However these holdings are generally smaller than those farms on the census and a similar size to those on the farm list with no livestock records. The farms could exist for three reasons, given that the farm list is the circulation list for the census:

1. They are farms which did not return a census form. Completion of the agricultural census is mandatory, however, ensuring that a large number of forms are returned is difficult to implement and it is possible that only larger farms are chased up for census forms.

2. The holdings did not have livestock when the agricultural census was sent out.

3. The holdings may have ceased farming or switched to arable production.

For the 111,560 farms for which there is both census and farm list species data the two agree for cattle and sheep in the majority of cases (Table 3.7). However, there is a large discrepancy for pigs with proportionately many more pig farms recorded on the farm list alone than on the census (Table 3.8). This data could reflect pig farms which previously farmed pigs but ceased farming, possibly as a result of market pressures and the intense commercialisation of the pig industry.

	Census only	Farm list only	Both	Total
Cattle	7507 (9.1)	3383 (4.1)	71197	82,087
Sheep	3583 (4.6)	7604 (9.8)	66397	77,584
Pigs	2471 (14.3)	7298 (42.3)	7464	17,233

Table 3.7: Species recorded on he holdings by data source. Only those with both farm list and census livestock data are presented. The 'Farm list only' and 'Census only' columns relate to holdings on which only that data source records that species, the 'Both' category is instances in which both data sources record that species and therefore both data sources agree. Percentages of the row totals are in brackets.

	Census	Farm list	Total
Cattle	94,710	12,414 (11.6)	107,124
Sheep	82,785	25,687 (23.7)	108,472
Pigs	12,063	9,487 (44.0)	21,550

Table 3.8: Farm composition by species on the farms for the census compared to those on the farm list with only species present. Percentages of the row totals are in brackets.

Holding composition by species for farms with species presence data is broken down by whether the farm is also on the census in Table 3.9. This shows that sheep and pig holdings are over-represented and cattle holdings are under-represented. Furthermore, partly as a result of the under-representation of cattle and therefore drop in the number of mixed cattle and sheep holdings there are many more single species holdings on the farm list (85.1% = 18%+48.9%+17.2%) compared to the census (56.5%).

Species	Farm list only	Farm list and census
Cattle only (C)	7,319 (18.0)	29,420 (26.7)
Sheep only (S)	19,823 (48.9)	29,335 (26.6)
Pig only (P)	6,959 (17.2)	3,481 (3.2)
CS	3,922 (9.7)	36,650 (33.2)
CP	586 (1.4)	3,266 (3.0)
SP	1,355 (3.3)	2,771 (2.5)
CSP	587 (1.4)	5,243 (4.8)
Total	40,551	110,166

Table 3.9: Species composition for those holdings which are listed as species present on the farm list broken down by whether the holding is recorded on both farm list and census or just farm list. Percentages of the column totals are in brackets.

Figure 3.5 shows that the holdings with just a farm list record of livestock are not evenly distributed and are typically found in the areas of eastern England which are more commonly associated with cropping and pig farming (Figure 3.5). Despite the tendency for livestock present holdings to be in different areas to the majority of the FMD cases (Figure 3.5) the livestock present holdings are still under-represented in the three counties with the greatest numbers of FMD cases (Table 3.10).

The following conclusions can be drawn concerning the 43,226 for which only species presence is recorded:

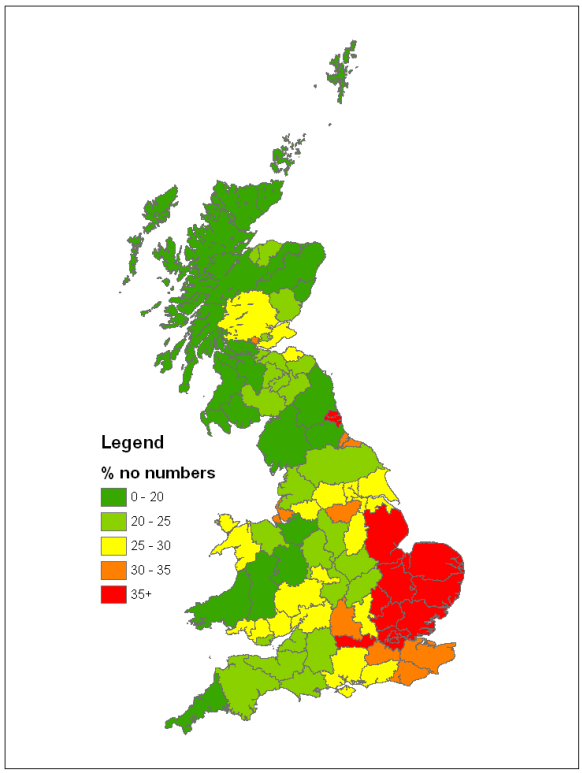


Figure 3.5: Percentage of livestock holdings listed as having livestock on the farm list but with no records on the agricultural census broken down by county.

	Census		Farm list PA		χ^2_1	
County	Total	Culled (%)	Total	Culled (%)	χ^2_1	p
Cumbria	5,114	1,987 (38.7)	1,137	136 (12.0)	168.3	<0.001
Devon	8,404	734 (8.7)	2,272	61 (2.7)	83.7	<0.001
Dumfriesshire	1,085	642 (59.2)	293	74 (25.3)	38.2	<0.001

Table 3.10: The number of culled holdings which were identified as having livestock by the census against those whose livestock were identified only by the farm list. This is for the three counties with the most holdings culled in the FMD control effort. The percentages are of the neighbouring 'Total' cell.

1. From the cull data it is clear that some of these holdings did have animals during the 2001 FMD epidemic. However the cull data also show that it is likely that many did not.

2. When there are animals on the holding the presence data is a reliable indicator of the species actually on the holding.

3. The farms with no census data are small farms. They are in most cases single

species holdings.

4. The farms with no census data are less likely to hold cattle and more likely to hold sheep and particularly pigs.

These findings show that it is an unknown fraction of these holdings which have animals, therefore the population of these farms would have two uncertainties: whether a holding has animals and how many animals that holding has. Furthermore populating these holdings would involve creating data which are not already in the headline statistics for numbers of animals during the 2001 FMD epidemic and increasing these figures. As a result these holdings will not be populated and from here on these holdings will be treated as non-livestock.

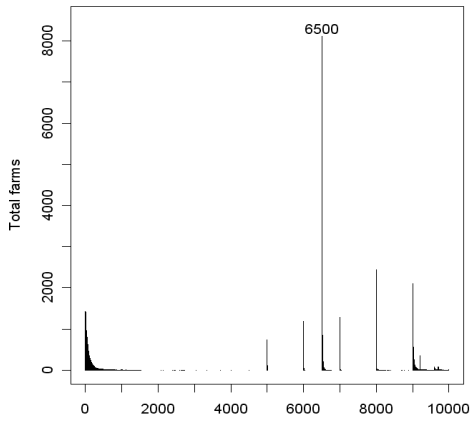
3.6 Linking satellites to main holdings

Section 3.5 identified some 3,610 farms which were listed in the farm list as having no animals but during 2001 were culled and animals removed. This section will further investigate this and demonstrate that what was defined as a holding in the agricultural census and what was defined as a holding during the FMD epidemic are different. Methods will be developed to link some of the farm list holdings (satellite farms) to holdings with livestock (main farms) and these parcels of land will be populated with animals from their linked farm.

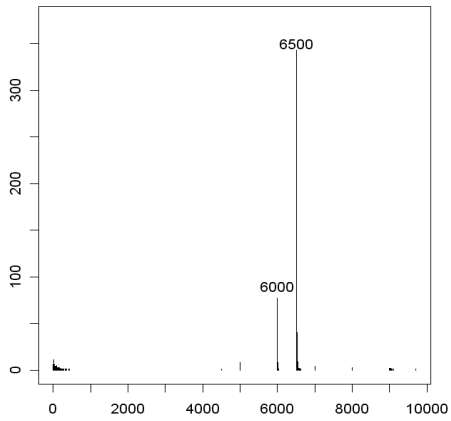
The DEFRA agricultural census records numbers of animals on all the fields a farmer owns to one CPH number. In many cases a single CPH number on the agricultural census really comprises a number of distributed parcels of land. In these cases for FMD control if a field was more than 1km from the main farm holding the field was recorded under a separate CPH number (Gibbens and Wilesmith, 2002). Furthermore, a farmer may rent or lease fields from another holding as temporary pasture for animals. At the beginning of the 2001 epidemic DEFRA conducted an inventory of parcels of land in England and Wales which held livestock, each parcel was assigned an emergency CPH number and its coordinate recorded, although animal numbers were not recorded (Honhold and Taylor, 2006).

The parcels of land identified at the start of the epidemic were assigned certain

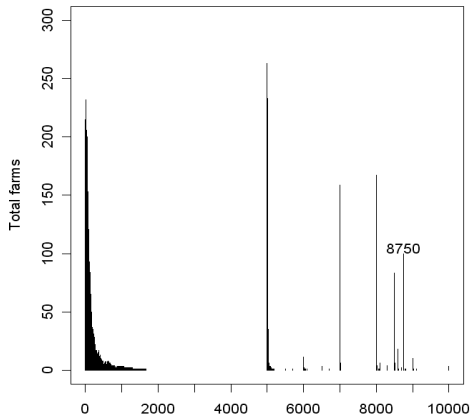
holding (the H component of the CPH number) numbers. The numbers assigned were different for England, Scotland and Wales. Analyses of these holding numbers has been conducted for Cumbria (Honhold and Taylor, 2006), although not for the rest of the UK. For this analysis these holding numbers were identified using data in Honhold and Taylor (2006) and by comparing the holding numbers of 'non-livestock farms' from the farm list with the holding numbers of farms culled which were not on the census (Figure 3.6).



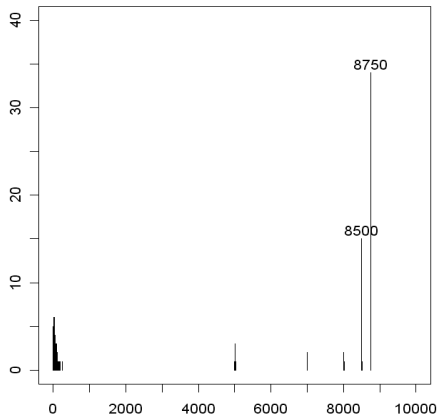
(a) England: non-livestock



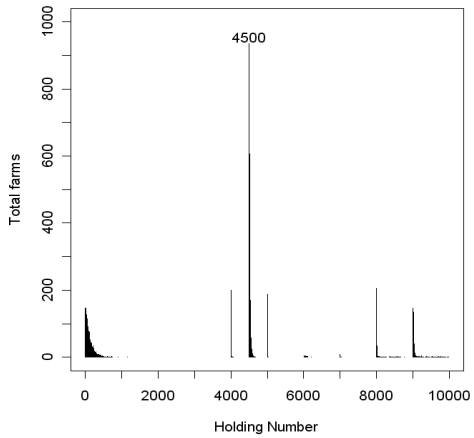
(b) England: Culled



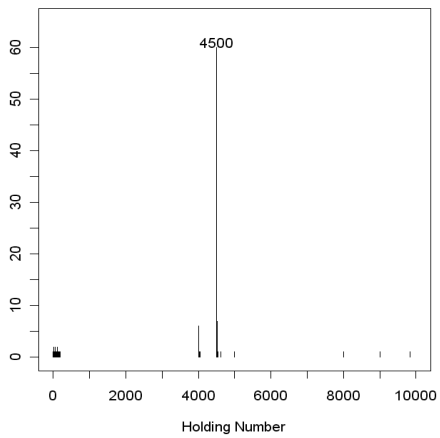
(c) Scotland: non-livestock



(d) Scotland: Culled



(e) Wales: non-livestock



(f) Wales: Culled

Figure 3.6: Holding numbers of all farms with no livestock recorded and all culled farms with no livestock respectively for England (Plot 3.6(a) and 3.6(b)), Scotland (Plot 3.6(c) and 3.6(d)), and Wales (Plot 3.6(e) and 3.6(f)).

The peaks in the cull numbers in Figure 3.6 indicate that parcels of land in England identified as having animals on them were assigned a holding number between

6500 and 6600, those in Scotland a number between 8750 and 8800 and those in Wales a number between 4500 and 4600. This is in partial agreement with Honhold and Taylor (2006) who identify holding numbers in the range 6000 and 6500 as the emergency holding numbers for Cumbria. This is shown by the 2,896 (80.2%) of 3,610 farms culled with no livestock records fall into these holding ranges, whilst farms with these numbers comprise 81,509 (27.4%) of the 297,518 'non-livestock' farms in the database. These newly identified farms are not distributed uniformly across England, Scotland and Wales (Table 3.11).

	Holding range	Farms in range	Total non-livestock farms culled	Total culled in range
England	6500-6600	66125	3085	2619 (84.9%)
Scotland	8750-8800	289	274	94 (34.3%)
Wales	4500-4600	15095	247	183 (74.1%)

Table 3.11: The number of holdings in the ranges of values identified in Figure 3.6 broken down by culls.

Table 3.11 shows that there were far fewer parcels of land identified in Scotland. From the distribution of the parcels of land (Figure 3.7) it appears that only parcels of land within FMD zones were recorded in Scotland, whilst a thorough inventory was carried out for England and Wales covering all areas. This means that not all land that might have housed animals in Scotland were identified. This will introduce bias into any dataset which uses these parcels of land. Furthermore whilst many of the 3,610 culled farms can be identified using their holding numbers there are still a further 714 (19.8%) pieces of land with animals culled land which does not fall into these holding number categories.

A further problem is that the holding number for these satellite pieces of land merely identifies them; it does not indicate which holding owns the animals they stock. However, the farm list records the owner of the parcel of land and this can be used to link potential satellite holdings to main holdings. A Java program was written to compare the owner name field of the livestock farms to the owner name field of non-livestock farms. If several possible matches were found for a non-livestock farm, the closest potential main farm was selected.

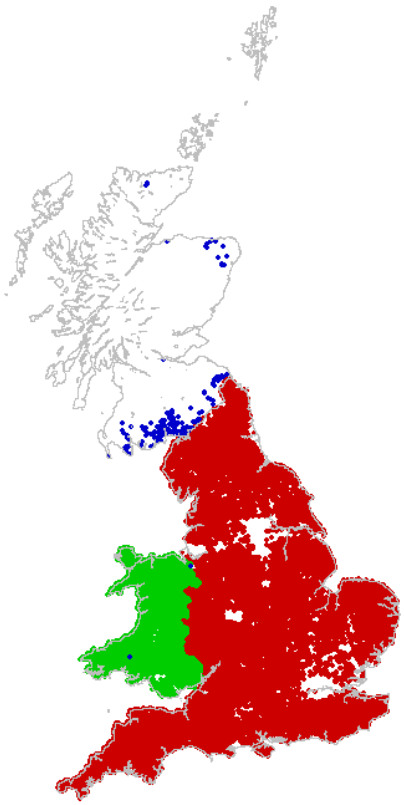


Figure 3.7: Spatial distribution of all farms with holding numbers in the range 6500-6600 (red points), 4500-4600 (green points) and 8750-8800 (blue points).

3.6.1 Evaluation of farm links

Using the owners name 71,911 (88.2%) of 81,509 farms which were identified as being livestock farms by their holding number could be linked to a livestock farm on the census using the owners name. Furthermore Figure 3.8 shows that these farms were substantially closer together than those linked using the owners name which were not identified as holding livestock using the holding number. The differences between the two lines are highly significant using a Kolmogorov-Smirnov test ($p<0.001$).

Based on these analyses a robust set of rules are required for linking satellite holdings in the farm list to main holdings on the census. Given the differences in Figure 3.8, farms which were not identified as being livestock by the holding number can not reliable be linked to a main holding, they are too far away and the identical name is likely to be coincidence. Therefore, only when the satellite farm has a holding number in the defined ranges and an owner link will it be linked to a main farm.

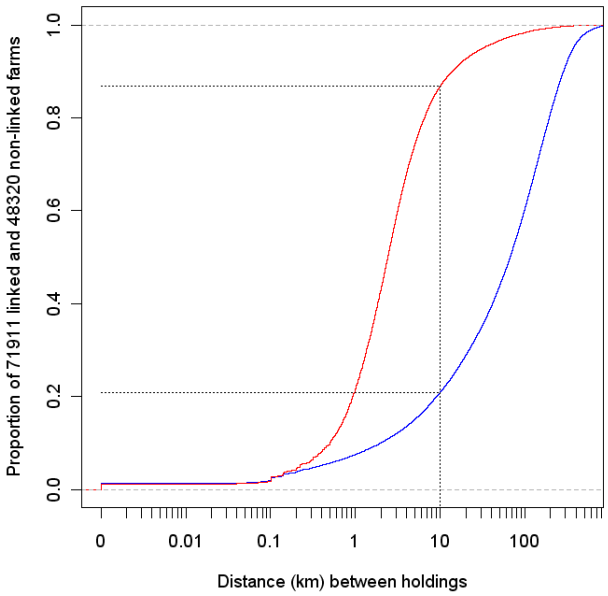


Figure 3.8: Cumulative distribution plot of the distance between the main and satellite farm. In instances where one holding owns both pieces of land and the holding number indicates a livestock (red line), and based just on owners name with no holding number indication (blue line).

Using the criteria described above 71,911 satellite farms were linked to 27,415 main holdings.

3.6.2 Analysis of farm links

Investigations of main holdings reveals that the greater proportion of holdings with satellites are mixed cattle and sheep farms. Furthermore a large number are mixed cattle, sheep and pig farms, although overall these are relatively few in number (Figure 3.9).

The satellite farms were generally smaller than the main holdings and were more likely to have sheep than cattle (Figure 3.10). The majority of satellite farms had no cattle culled, shown by the median on zero in Figure 3.10. This is likely to be a reflection of the fact that cattle were not culled on all farms and possibly that farmers are more willing to put sheep on pasture outside the main holding. This is also a reflection of the fact that the majority of main farms were IPs compared to the majority of satellite farms which are CPs (Table 3.12). This suggests that the satellite farms are close to the main farms as they were apparently taken as a contiguous cull

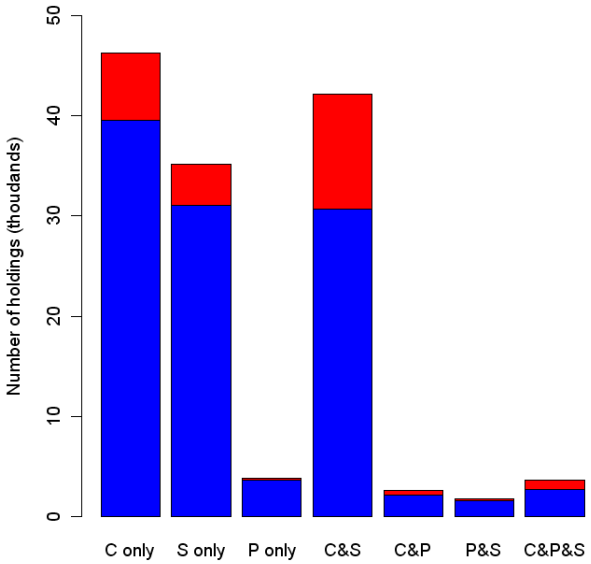
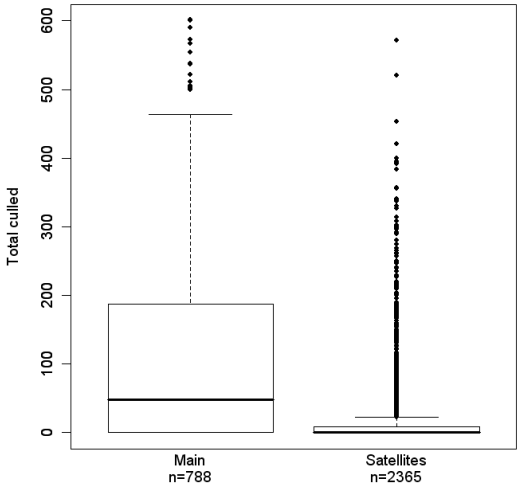


Figure 3.9: Barplot of numbers of main holdings without satellites (blue area) and with (red area) broken down by species composition of the main holding (cattle=C, sheep=S and pig=P).

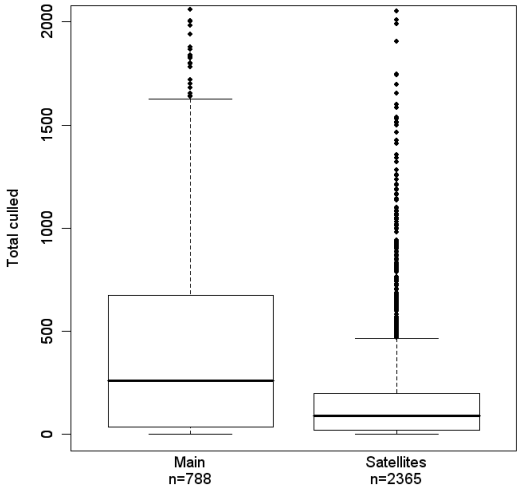
after the main holding had been declared an IP. The culled satellite farms were also a slightly higher proportion of DCs, as they have a link to the main holding if the main holding is declared an IP. Analysis of the reason for culling satellite holdings whose main holding was an IP shows the greatest number were culled as CPs (Table 3.12).

Cull reason	Satellites (%)	Main holding cull (%)	All Culls (%)
3km	776 (31.6)	320 (17.0)	2980 (27.5)
CP	828 (33.8)	218 (11.6)	3620 (33.4)
DC	472 (19.2)	194 (10.3)	1420 (13.1)
IP	240 (9.7)	1110 (58.9)	2026 (18.7)
Local	19 (0.7)	9 (0.4)	280 (2.5)
Other	67 (2.7)	24 (1.2)	268 (2.4)
SOS	50 (2.0)	11 (0.5)	250 (2.3)
Total	2450	1890	10800

Table 3.12: Reason for the cull of satellite holdings where the main holding was also culled.



(a)



(b)

Figure 3.10: Boxplots of numbers culled on main holdings compared to satellites for cattle (plot 3.10(a)) and sheep (plot 3.10(b)).

3.6.3 Populating farm fragments

In order to use the fragment data, census animal numbers from main holdings must be distributed among the satellite holdings. This was achieved using data on culled farms, but only those instances in which the main farm and all identified satellites were culled. Additionally, for cattle farms, none of the component fields may have been culled as part of the 3km cull. Two methods for populating farm fragments were derived using these data:

1. **Entire enterprise sampling** Sampling based upon the proportion of animals belonging to the farm enterprise which are on the satellites (Figure 3.11). The cattle line in Figure 3.11 is composed of 218 data points. Each data point represents a complete farm enterprise with fragments in which the main holding and all identified fragments have been culled but not as part of the 3km cull. The corresponding number of sheep holdings is 455, which is more than double that for cattle because sheep holdings were taken in the 3km culls, so these are included. To populate farm enterprises each derived frequency distribution in Figure 3.11 is sampled (with replacement) and the corresponding proportion of sheep and cattle subtracted from the census number on the main holding and divided evenly among the satellites. For instance consider a farm enterprise with 200 cattle and 800 sheep according to the census which has 2 linked satellites. The 2 distributions in Figure 3.11 are sampled and values of 0.4 for cattle and 0.8 for sheep chosen. $0.4 \times 200 = 80$ and $0.8 \times 800 = 640$. The values of 80 and 640 are subtracted from the 200 and 800 and divided by 2 to give the numbers on each satellite. The enterprise now consists of a main holding with 120 head of cattle and 160 sheep and 2 satellites each comprising 40 cattle and 320 sheep.

2. **Satellite level allocation.** The farms on which sampling is based must be instances in which the entire farm enterprise (main holding and all off-fields) was culled. Thus, the same rules that were applied to entire enterprise sampling still apply. However, the enterprise on which the sample is based is taken from a subset of enterprises with the same number of satellites as the enterprise to be populated. The fragments are then populated according to the proportions

of cattle and sheep on each fragment on the sampled premises. So for the enterprise of 200 cattle, 800 sheep and 2 satellites an enterprise is sampled from those culled with 2 satellites. The proportion of the enterprises cattle on the satellites are 0.5 and 0 and 0.1 and 0.6 for sheep on the two satellites respectively. So the enterprise is repopulated thus: the main holding contains 100 cattle and 240 sheep, the first satellite 100 cattle and 80 sheep and the second satellite 480 sheep and no cattle.

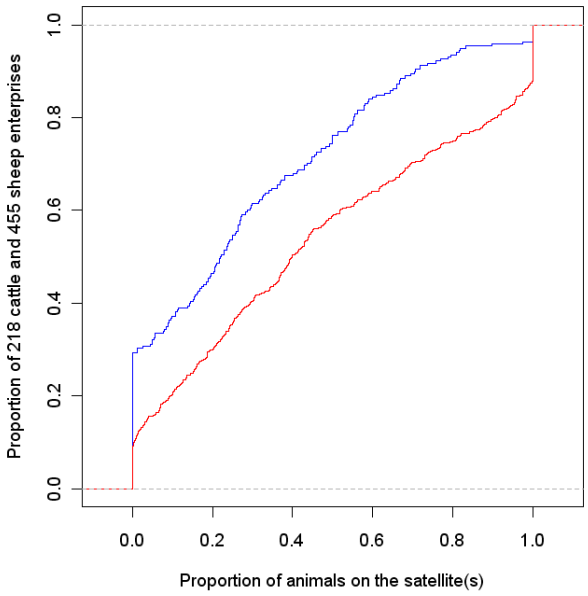


Figure 3.11: Cumulative distribution plot of the proportion of cattle (blue line) and sheep (red line) belonging to the entire farm enterprise which were on fragments during the 2001 FMD epidemic.

The second approach would be the optimal solution as it is based upon the actual satellite numbers rather than values pooled across the entire farm enterprise. It makes allowances arising from differences in the numbers of satellites in relation to the proportions of animals on the satellites. However there are very few entire culled holdings with more than two satellites culled (Figure 3.12). As this method is dependent upon matching to holdings with the same number of satellites this would introduce bias as enterprises with several fragments would be sampled from a very small population.

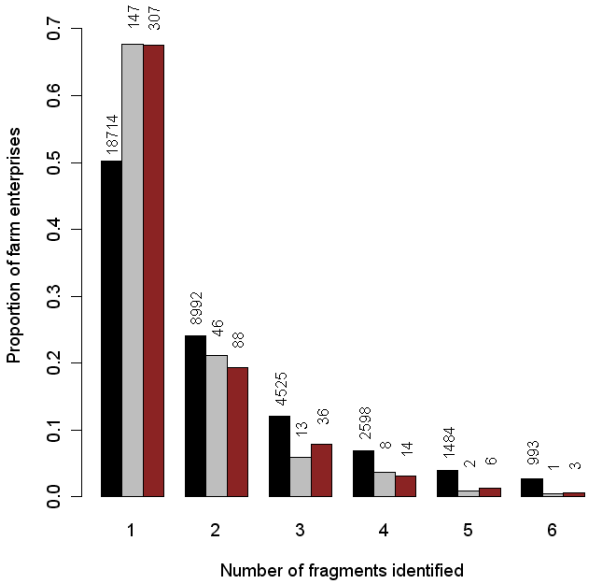


Figure 3.12: Proportions of farm enterprises by the number of satellites linked to them for the entire dataset (black bars), entire enterprises culled with cattle (grey bars) and entire enterprises with sheep culled (brown bar). Numbers above the bars are the number in each group.

As a result of these limitations with the satellite level allocation the entire enterprise sampling method was used to allocate animals recorded on the census to satellite holdings. However there remain limitations in this method of allocation:

1. Equal numbers of animals are distributed among all satellites.

2. Fragments are left with no animals. This occurs because values of zero can be sampled for both cattle and sheep and as a result all satellites for an individual holding are left without animals. Another possibility is a value greater than zero is sampled but there were no animals of that species recorded on the census for the receiving enterprise. This was the case for 16,334 (22.3%) of fragments.

3. Farms with more fragments have a greater proportion of their animals distributed among their fragments (Figure 3.13). This was analysed using a generalised linear model with binomial errors result in which the outcome is the proportion of animals on he fragment and the predictors the number of fragments. The estimate for cattle was 0.268 ($p<0.001$) and sheep 0.259, ($p<0.001$). However, for the reasons described above there is limited data on complete en-

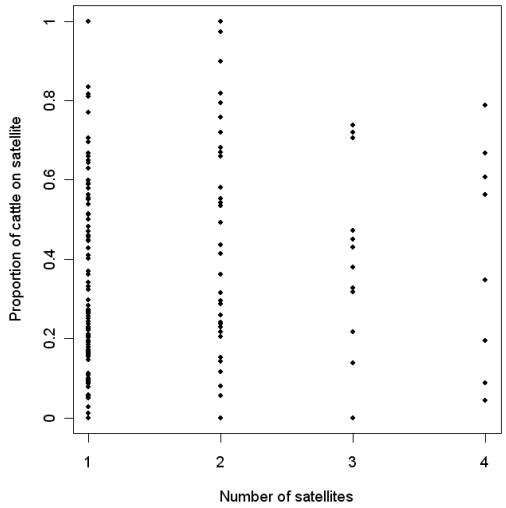
round	All Culled	entire farm enerprise culled	%
Cumbria	3872	313	8.084
Devon	1218	19	1.560
Dumfriesshire	910	21	2.308
North Yorks	785	26	3.312
Northumberland	435	15	3.448
Hereford &Worc	395	6	1.519
Powys	392	9	2.296
Durham	384	16	4.167
Gloucestershire	363	7	1.928
Kircudbright	324	8	2.469
Lancashire	275	8	2.909
Gwynedd	238	1	0.420
Wigtown	214	5	2.336
Staffordshire	167	6	3.593
Gwent	149	8	5.369
Shropshire	127	0	0.000
rest of UK	598	18	3.010
Total	10846	486	4.481

Table 3.13: The number of entire farm enterprises culled during the course of the 2001 epidemic for counties in which more than 100 premises were culled.

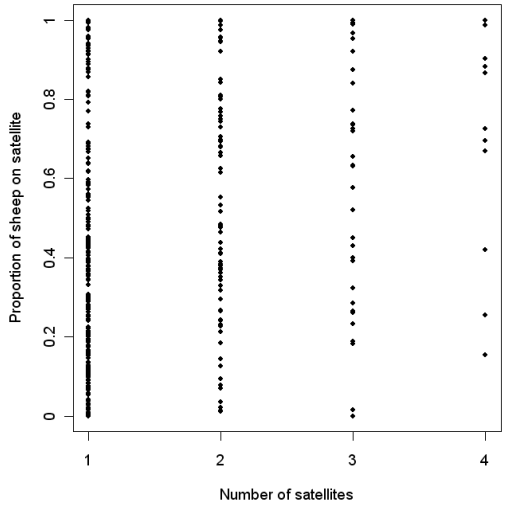
terprises culled with more than one fragment and as a result this can not be controlled for.

4. Bias in the spatial distribution of complete farm enterprises culled. The greatest proportion of farms on which all animals on the main and satellite holdings were taken are in Cumbria (Table 3.13). As no county other than Cumbria has more than 30 complete holdings culled using different distributions for different regions is not statistically feasible, so data from the whole of GB was used. Therefore all populations are based upon data largely from the Cumbria disease cluster and spatial variations in fragment population can not be accounted for.

A method has been derived for dividing animals recorded in the agricultural census among satellite holdings where these satellites have been previously identified. However the method chosen was driven by data limitations and the method has many limitations. Given additional data on animal numbers on farm enterprises and data on how these animals are distributed among the component units of the enterprise more optimal solutions could be devised.



(a) Cattle



(b) Sheep

Figure 3.13: Proportion of animals on satellites by the number of fragments owned by the farm enterprise for cattle (plot 3.13(a)) and sheep (plot 3.13(b)).

3.7 Conclusion

A farm dataset has been constructed primarily using the agricultural census with supplementary data on satellite fields from the farm list. A spatial reference was selected using the farm list reference which identifies the main farm building and matches the reference from the FMD data.

The demographic component on census holdings was evaluated to evaluate how accurately the data reflects the data on the cull. Differences exist in the sheep data

probably as a result of lamb sales for meat and movements of sheep from upland to lowland pasture between June and February. Cattle analysis is complicated by partial culls of cattle, otherwise the census was an accurate representation of cattle numbers.

There were a number of holdings for which the farm list recorded species presence but not numbers. These farms represent holdings which have ceased livestock production, or did not have stock at the time of the census or holdings which did not return a census form. Therefore due to the uncertainty in how many holdings had stock and how much stock they held these holdings were not populated.

Satellite holdings of main holdings were identified wherever possible and were populated by sampling from distributions from the cull. However, although the locations are known the population from which sampling was conducted was limited.

Chapter 4

Risk factors for holding level susceptibility

4.1 Introduction

In the event of a future outbreak of FMD the immediate response must be the minimization of further local and long range transmission events (Anderson, 2002). This will be expedited by imposing a movement standstill on susceptible livestock and elevating biosecurity measures (Anderson, 2002). Following these measures, there may be continued local spread albeit at a reduced rate. Some holdings will be more susceptible to infection by local spread and these holdings may be concentrated in certain areas. Therefore if a model can be developed to identify holdings and areas at high risk, infection control measures and resources can be targeted most effectively. Furthermore, such a model can be used retrospectively to understand patterns of transmission through the parameters which constitute the model.

Previous modelling studies have demonstrated the importance of spatial distribution of species numbers as a predictor of FMD incidences in the UK 2001 outbreak (Keeling et al., 2001, 2003; Tildesley et al., 2006) and other outbreaks (Haydon et al., 1997; Perez et al., 2006; Ward and Perez, 2004). In the models of Keeling et al. (2001, 2003) and Tildesley et al. (2006) two measures of animal numbers are important in determining FMD transmission:

1. Given that holding i is not infected, the number of cattle and sheep on i deter-

mines the farms susceptibility to infection.

2. Given that holding i is infected, the number of cattle and sheep in the neighbourhood of i determines the potential of i to infect further holdings.

Additionally, the model takes into account the distance between the infected farm i and a source of infection j. The probability that holding i will become infected on day D is given by (Tildesley et al., 2006):

$$p_{i,D} = 1 - exp\left(-(s_c n_{c(i)} + s_s n_{s(i)}) \sum_{j \in \forall infectious(D)} (t_c n_{c(j)} + t_s n_{s(j)}) K(d_{ij})\right) \quad (4.1)$$

where n_c and n_s are the number of cattle and sheep respectively, s_c and s_s are susceptibility parameters and t_c and t_s are transmissibility parameters for cattle and sheep respectively. Transmission is calculated based on all holdings which are infectious on day D. d_{ij} is the Euclidean distance between i and j, and $K(d_{ij})$ is the transmission kernel that weights transmission based on the distance between i and j.

The model described above is a dynamic spatio-temporal model of disease spread which is designed for forward projection of epidemic progression. It is a means of assessing the progression of the epidemic one generation at a time and is not intended to give a start of epidemic measure of infection risk. Some means of assessing the risk of infection for an individual holding would be invaluable for assessing areas most at risk from infection before an epidemic. This would be particularly valuable if such an assessment can be made statically or given a particular epidemic situation so that resources can be targeted effectively. Any such index would ideally comprise three measures:

1. **Euclidean distance to an infectious source.** This captures the spatial dependency of FMD.

2. **The number and species of animals on the individual holding.** This describes the number of animals on the holding which can become infected. Additionally greater numbers of animals are likely to result in a greater throughput of personnel, equipment, feedstuffs and machinery which could bring infection onto the holding.

3. **The number of animals in the local area.** The majority of infection originates from nearby IPs via local spread, therefore some measure of the potential of the local neighbourhood to spread infection is required.

This thesis chapter will develop a static measure of risk of FMD infection for individual holdings based upon these measures. The study will have the following aims:

1. To evaluate whether these variables act as predictors for FMD cases in the 2001 FMD epidemic.

2. To derive a predictive measure of risk of FMD infection based on the variables described above for the 2001 epidemic.

4.2 Methodology

4.2.1 The Model

Eight variables were calculated for each holding. These variables fall into four groups:

1. **Distance to seed.** The distance to the nearest initially infected infectious source which are defined as the 78 IPs estimated to have been infected before the imposition of the NMB.

2. **Holding population.** The number of cattle and sheep on each holding. However, this cannot be modelled as two continuous variables as there are a large number of sheep only (22.4%) and cattle only (31.5%) holdings, so there are a large number of zeros in the data. Therefore these data will be modelled as species on the holding as a three level variable with mixed, cattle only and sheep only holdings. Mixed is the reference level against which other levels will be compared. A separate model for mixed holdings with numbers included will be developed to investigate the effect of numbers of animals on the holding. An additional predictor is the area of the holding in hectares. All data is taken from the June 2000 agricultural census.

3. **Country.** Differences at the national level in the nature of farm holdings and possibly the nature of census data collection have been described previously

(Chapter 3). Furthermore, the epidemic was managed by different governments in England, Scotland and Wales who exercised some independence in epidemic management. Therefore, the country of the farm (England, Scotland or Wales) will be included as a factor with England as the reference level.

4. **Farming in the locality.** Cattle and sheep numbers in the neighbourhood of each holding. Calculations of these densities will use a kernel as described in Section 4.2.4. Furthermore the total area (in Ha) under livestock (cattle and sheep) production will be calculated.

The outcome for these analyses is whether a farm is an IP or not an IP. There are two formal statistical methods for differentiating between two groups of values using a set of predictor variables. These are logistic regression and linear discriminants analysis (LDA). Both methods perform similar analysis, the major difference is that LDA seeks to place the data into the two groups (in this case IPs and non-IPs (Armitage and Berry, 1987), and logistic regression is based on the likelihood of each farm becoming an IP. The purpose of LDA is "to set up some rule which will allocate further individuals to the correct population of origin with minimal probability of misclassification" (The International Statistics Institute, 2006, p.116). However the process of infection during the 2001 epidemic was highly stochastic, and if the epidemic were repeated a different set of holdings are likely to become infected. Therefore, assigning individuals to a particular group is not appropriate. There are likely to be identifiable factors which predispose holdings to infection and determine the likelihood of infection, identification of these factors is the aim of this chapter. As a result logistic regression will be used rather than LDA.

There is spatial dependency in the outcome and in some of the predictors. This is because IPs are spatially clustered and among the predictors are the numbers of animals in the locality of the holding, distance to an infectious seed and country of the holding, all of which are also spatially dependant. The significance of the spatial dependance will be tested using Moran's I statistic on the residuals of the GLM. Moran's I will be calculated for a range of nearest neighbours using the *spdep* package in the R statistical environment. Significant autocorrelation in the residuals will be

corrected by imposing a 5km wide hexagonal lattice upon the spatial distribution of the data. The unique ID of the hexagon into which each farm falls will be assigned as a random effect to form a generalised linear mixed model (GLMM, with binomial errors). Previous studies (Haine et al., 2004; Pfeiffer et al., 2007) used political boundaries as the random effect. However, political boundaries are uneven in size and shape, so by imposing an arbitrary network of evenly sized polygons upon the data, some of these problems have been overcome. The GLMM will be generated using the *lme4* package in the R statistical environment and the residuals of the GLMM will be tested for spatial autocorrelation in the manner described above.

4.2.2 Data

The model requires data regarding numbers of sheep and cattle on IPs and non-IPs. However as discussed in Chapter 3 there is no way of matching these datasets perfectly. Three variations on the demographic and FMD datasets could be used to generate the dependent variables, these variations are described in Table 4.1.

Demographic data	FMD data	Number of non-IPs	Number of IPs (non-seeds)
Census holdings	Census IPs	135,142	1582 (1534)
Census holdings	DCS data	135,142	2026 (1946)
Census holdings + fragments	Census IPs + fragments	197,525	2026 (1946)

Table 4.1: Combinations of possible datasets for these analysis.

There are areas of incompleteness, inaccuracies or biases with each demography and FMD data combination, these are:

Census holdings & census IPs: Farm fragments are not included, therefore numbers of animals on holdings are over represented and many potential IPs are missing. However, data are accurate and there is no new bias being introduced as census data is used for all analysis. Analysis would be of the subset of IPs which are on the agricultural census.

Census holdings & DCS data: This includes data on all 2026 IPs. However those 2026 include farm fragments which are not included in the non-IP data,

therefore bias is potentially introduced. Furthermore data on animal numbers are from two different sources.

Census and fragment data: This includes data on more IPs than the census (but not all IPs) and most fragments. However data on animal numbers on fragments is generated by bootstrapping (Section 3.5) and data is only partially complete for Scotland.

As the census data is the dataset which comes from one source and is the most robust for such analyses (Chapter 3) this will be used in these analyses, furthermore census data has been used in previous analyses (Ferguson et al., 2001b; Kao, 2003; Keeling et al., 2003; Matthews et al., 2003). After the initial stages the epidemic predominantly involved cattle and sheep. Table 4.2 demonstrates that the proportion of the national stock of pigs which was culled on IPs was an order of magnitude lower than the corresponding proportions of cattle and sheep. Therefore farms without cattle or sheep were excluded from the analysis. The 78 seed IPs will not be included in the analysis as some of these were infected by processes such as animal movements which were not operating in later stages of the epidemic. After removal of these farms there were 130,136 non-IPs and 1,527 IPs in the dataset. Results will be compared to analysis using the modelled fragment dataset (Chapter 3) to compare the census dataset with data with fragments incorporated. The two models will be compared by checking for an overlap in the confidence intervals of the odds ratios of the predictors from the two models. Further analysis will be carried out by taking as IPs just those 1,025 which were laboratory positive.

Species	National stock	On IPs	% on IPs
Cattle	9,501,181	294,716	3.10
Sheep	39,580,631	974,785	2.46
Pigs	6,317,475	20,475	0.32

Table 4.2: The percentage of the national stock by species which was culled on IPs only.

Values for species composition can be taken directly from these data, and distance to an infectious seed can be generated using the FMD data. However the kernel densities for cattle and sheep must be derived using these data. The outputs of

different kernel shapes described in Section 4.2.4 will be tested.

4.2.3 Logistic regression

The binary outcome variable (whether the holding was an IP) will be analysed by univariate logistic regression analysis using each independent variable described in Section 8.2 as the predictor. The logistic regression will be carried out using the steps described in Hosmer and Lemeshow (2000, p.92-98):

1. **Univariate analysis**. Will be carried out for all predictors against the outcome. The distance to seed variable will be tested against a range of possible transformations, as there is no reason to assume that the relationship between distance to seed and the outcome will be linear. The different transformations to be be tested are: log_{10}, $x^{1/n}$ and x^{-n}. Univariate analysis will be conducted in the groups defined in Section 8.2.

2. **Multivariate model development**. All univariate predictors with p<0.25 are entered into a multivariate model. Terms non-significant at p<0.05 will be removed using manual backward stepwise model fitting to generate a most parsimonious main effects model.

3. **Preliminary main effects model**. Comparisons of the estimates of risk factors in the multivariate model and univariate model will be carried out to look for significant changes.

4. **Variable checking**. This involves checking for linearity of the logits of the predictors in the multivariate model and will be carried out by plotting the variable against the logits of the model which are generated in R using the predict.lm function. Non-linear logits will be adjusted using a transformation and where this does not work by dropping the variable from the model. The resulting model will be called the main effects model.

5. **Interactions** Biologically plausible interactions will be tested. An interaction will be retained in the model if $p < 0.05$. If the inclusion of a subsequent term changes the p-values of another interaction to a value of p greater than 0.05 the

non-significant term will be removed. The impact of removing interactions on the p-values of the remaining interactions will be checked. Biologically plausible interactions to be considered are:

(a) **Holding level interactions**: Between all three (Species and farm area) predictors at the holding level. Holdings with more animals may occupy a greater area, may be more likely to disperse animals on off-fields and may be more likely to stock multiple species.

(b) **Density interactions**: Between cattle and sheep densities.

(c) **Distance density interactions**: How animal densities vary with distance to the seed.

(d) **Species level interactions**: How species on the holding interact with animal densities.

(e) **National level interactions**: Differences in the other predictor variables at the national scale. This is with the exception of distance to the seed with which any interaction would reflect the distribution of seeds.

Models will be compared using the Akaike Information Criterion (AIC). AIC is calculated as (Crawley, 2002):

$$AIC = -2 \times l + 2(p + 1) \tag{4.2}$$

where l is the log likelihood and p the number of parameters. The model with the best fit is the model with the smallest value of AIC.

Further model comparison will be conducted using the area under the Receiver Operator Characteristic (ROC) curve (Dohoo et al., 2003) and is based upon the modelled logistic regression prediction value. The area under the ROC curve is calculated as (derived from Hosmer and Lemeshow (2000)):

$$A_{ROC} = \frac{\sum_{i=1}^{N} r_{+i}}{n_+ \times n_-} \tag{4.3}$$

where n_+ is the number of true positives and n_- the number of true negatives and $N = n_+ + n_-$. r_+ for true positive i is the number of true negatives which returned

a lower modelled value than i. Possible values of A_{ROC} range from 0 to 1, values for A_{ROC} should be interpreted as (adapted from Hosmer and Lemeshow (2000)):

ROC ≈ 0.5 there is no discriminatory power.

$0.7 \leq ROC < 0.8$ Acceptable discrimination.

$0.8 \leq ROC < 0.9$ Excellent discrimination.

≥ 0.9 Outstanding, almost total discrimination.

The model will be further evaluated by testing the model's ability to predict the actual IPs. This will be done by selecting the n holdings with the highest modelled values and evaluating how many of these were actually IPs where n is the number of IPs in the outcome variable. This model will be compared to a second model which will use just IPs which were infected during phase 1 of the epidemic. This is to allow for any temporal effect resulting from culling policies which changed over the course of the epidemic.

Sensitivity analysis will be carried out by substituting the outcome variable for all holdings which have been culled and laboratory positive holdings to test how effective the model is at predicting culled and laboratory positive holdings. The analysis will be carried out on the GLMM and differences in the risk factors evaluated by plotting the odds ratios of the models being compared. Risk factors with odds ratios whose 95% CIs overlap are not significantly different.

4.2.4 Calculating animal densities

Methods of aggregating point data to create generalised density surfaces are discussed in great depth in the literature on spatial data processing and analysis (Bailey and Gatrell, 1995; O'Sullivan and Unwin, 2003; Worboys, 1995). For this analysis various measures could be derived to describe the density of holdings at a particular point, these measures are discussed in Table 4.3 and Figure 4.1.

For the purpose of this analysis some measure of density centered on the individual holding is required, therefore the circular density measure is most appropriate. For density calculation a radius or bandwidth (τ) is required. Some means of weighting

Statistic	Advantages	Disadvantages	Parameters
Nearest neighbour distance (a)	Describes the proximal relationship of farms	Does not incorporate attribute data and only informs on one farm.	None
Density - areal unit (b)	Creates a raster dataset which can be displayed and analysed in conjunction with other variables using overlay analysis	Prone to the modifiable areal unit problem (MAUP) (O'Sullivan and Unwin, 2003). Square units can be a poor representation	Dimension x
Density - circular from each point (c)	Describes a neighbourhood of interest whilst avoiding the MAUP. Maintains the original units of study.	Does not generalise. Difficult to display.	Radius r

Table 4.3: Different methods of describing the farming landscape and in particular the density of holdings

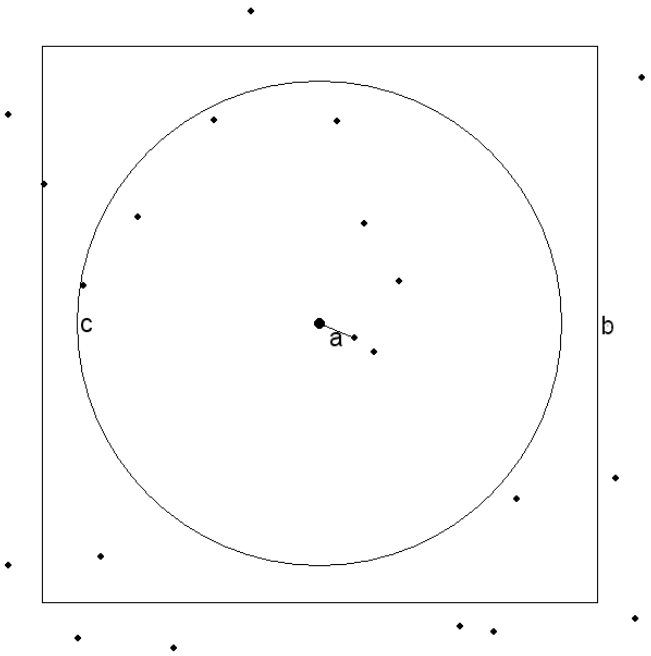


Figure 4.1: Illustration of methods for describing the farming landscape. The large black point is the holding in question, smaller points represent farm holdings and the polygons are units in which density is calculated. Labels correspond to labels in Table 4.3.

closer values is needed, so a kernel is applied to the density calculations. The kernel is used to calculate some coefficient (y) based upon the distance from the central point to the point of interest (h). Typical kernels include (see Figure 4.2 for kernel

shape):

1. **Flat kernel**: $y = 1$ where $h < \tau$, 0 otherwise.

2. **Quartic kernel**: $y = \frac{3}{\pi}(1 - (\frac{h^2}{\tau^2})^2)$ (from Bailey and Gatrell (1995)).

3. **Exponential kernel**: $y = e^{-kh/\tau}$. A value of 4.5 was chosen for k as this is the other extreme to the quartic and flat kernels with a rapid drop off at a relatively short distance followed by a long tail.

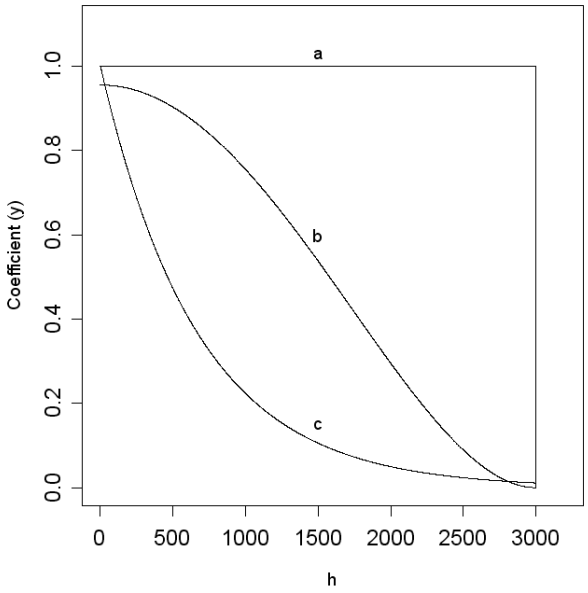


Figure 4.2: Examples of kernel functions where $\tau = 3000$. The lines describe changes in the coefficient (y) with increasing distances between farms (h). a is a flat kernel, b is a quartic kernel, c is an exponential function.

The range of kernel functions and values of τ will be tested to evaluate the effects of different kernels on the statistical significance of cattle and sheep densities and total farmed area. Values of τ will be investigated at distances up to 30km. The different values of τ will be compared using the AIC value generated by running logistic regression analysis of all three predictors (total cattle, sheep and area) against the outcome at each value of τ.

4.3 Results

4.3.1 Holding level analysis

All holding level predictors were highly significant in univariate logistic regression analysis (Table 4.4). Although all relationships are significant the protective effect of being a sheep only holding versus a cattle only holding is significantly greater. Just over one-third of holdings are cattle only, a similar proportion mixed cattle and sheep and around a quarter sheep only (Table 4.4). To achieve linearity in the logit holding area was $log_{10} + 1$ transformed. Holding area is a significant predictor of FMD risk, which may be due to a positive relationship with animal numbers.

Predictor	Unit	Non-IPs	IPs (%)	OR (95% CIs)	z value	p
Species	Mixed	45,232	996 (2.20)	1	-	-
	Cattle	48,508	408 (0.84)	0.38 (0.34,0.42)	-16.5	< 0.001
	Sheep	35,843	123 (0.34)	0.15 (0.13,0.19)	-19.5	< 0.001
Area	$log_{10}(Ha + 1)$	NA	NA	2.84 (2.61,3.09)	24.6	< 0.001

Table 4.4: Univariate logistic regression results of holding level predictors. The percentages are a percentage of the Non-IPs cell.

4.3.2 National level differences

At the country level, there is a small protective effect of a holding being in Scotland and larger protective effect of being in Wales (Table 4.5). This might be explained by considering the country level distribution of the 1527 IPs with respect to the susceptible population (Table 4.5). Table 4.5 shows that there are considerably fewer IPs in Scotland than England and fewer again in Wales. To fully understand the significance of this relationship it must be tested in a multivariate model in which distance to seed is included to allow for differences in the distribution of seeds.

Predictor	Unit	Non-IPs	IPs (%)	OR (95% CIs)	z value	p
Country	England	84,765	1,315 (1.53)	1	-	-
	Scotland	21,450	141 (0.65)	0.42(0.35,0.50)	-9.65	< 0.001
	Wales	21,484	71 (0.33)	0.21(0.16,0.27)	-12.70	< 0.001

Table 4.5: Univariate logistic regression results of the country variable. The percentages are a percentage of the Non-IPs cell.

4.3.3 Density measures

Three neighbourhood variables were calculated for each data point. These were: total (kernel transformed) cattle, sheep and farm area. However total farm area was closely correlated with total cattle and total sheep, and preliminary analysis of a logistic regression model containing the three predictors revealed that total farm area was not linear in its logit, instead the shape of the relationship was an inverted 'V'. As a result this predictor was dropped from subsequent analysis. To ensure linearity in the logits both total cattle and total sheep were square root transformed.

Values of AIC for the three methods of calculating densities described in section 4.2.4 were compared across a range of kernel bandwidths (τ) from 500m to 30km (Figure 4.3). Values of τ of between 5 and 7.5km return similarly low values of AIC for both the quartic and flat kernels, for all of these values cattle and sheep densities were highly significant ($p < 0.001$). The exponential kernel returns much higher values of AIC at smaller values of τ, but returns very low values at greater distances. The exponential kernel had a trough in AIC at 15km, which was the lowest value returned by any model, therefore, this was the kernel selected for use in subsequent analyses. Results of univariate analysis of these three predictors with the exponential kernel at a bandwidth of 15km is presented in Table 4.6.

Predictor	Unit	OR (95% CIs)	z value	p
Cattle density	$\sqrt{10^{-3}head}$	4.90 (4.46,5.38)	33.46	< 0.001
Sheep density	$\sqrt{10^{-4}head}$	2.85 (2.64,3.10)	25.72	< 0.001

Table 4.6: Logistic regression results of density based predictors.

4.3.4 Distance to seed

A square root transformation of distance to the seed returns the lowest AIC of all the transformations which were tested (Table 4.7). A linear function returns a substantially lower AIC than inverse transformed values of distance to the seed (Table 4.7). The square root transformed distance to seed was linear in its logit and was used in subsequent analysis.

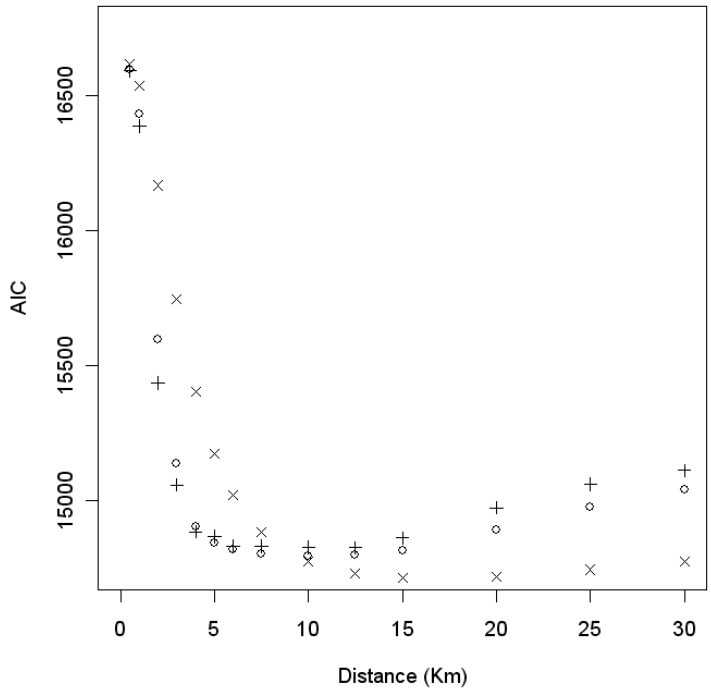


Figure 4.3: AIC values for univariate GLMs of animal densities where the model is a binomial model with cattle and sheep densities as the predictors and whether the holding is an IP as the outcome. Circles represent quartic kernel densities, flat crosses the flat kernel and diagonal crosses the exponential kernel.

Unit	OR (95% CIs)	z value	p	AIC
km	0.928 (0.923,0.932)	-33.240	< 0.001	14200
$log_{10}(km+1)$	0.056 (0.049,0.063)	-46.904	< 0.001	14315
$km^{1/2}$	0.525 (0.508,0.541)	-40.206	< 0.001	14157
$km^{1/3}$	0.225 (0.211,0.241)	-43.038	< 0.001	14193
km^{-1}	2.352 (2.033,2.722)	11.474	< 0.001	16390
km^{-2}	1.029 (1.014,1.043)	3.898	< 0.001	16573
km^{-3}	1.001 (1.000,1.003)	2.364	0.018	16582

Table 4.7: Logistic regression results for distance to the seed.

4.3.5 Multivariate model

All predictors had p<0.25 in the univariate analysis and therefore all were entered into the baseline multivariate model. The baseline multivariate model is shown by Table 4.8. The area under the ROC curve is 0.908 and the model correctly predicts 385 (25.2%) of 1527 IPs.

In the model in Table 4.8 the distance to the nearest seed is the principal risk factor, as it has the greatest value for the Wald statistic (z-score). The effect of this

Predictor	Unit	OR (95% CIs)	z value	p
Intercept		NA	-34.067	< 0.001
dist to seed	$\sqrt{km}$	0.605 (0.584,0.627)	-27.959	< 0.001
	England	1	-	-
Country	Scotland	0.964 (0.799,1.162)	-0.389	0.698
	Wales	0.124 (0.094,0.163)	-14.819	< 0.001
	Mixed	1	-	-
Species	Cattle	0.653 (0.573,0.744)	-6.407	< 0.001
	Sheep	0.579 (0.474,0.708)	-5.329	< 0.001
Farm area	$log_{10}(Ha+1)$	4.290 (3.776,4.873)	22.388	< 0.001
Cattle density	$\sqrt{10{-}3}Head$	4.326 (3.832,4.884)	23.666	< 0.001
Sheep density	$\sqrt{10^{-4}}Head$	4.128 (3.563,4.782)	18.892	< 0.001

Table 4.8: Logistic regression output for the main effects of the full multivariate model.

predictor can be further investigated by considering a model without distance to seed (Table 4.9), the remaining predictors are still highly significant and this model has strong predictive power (area under ROC curve=0.878).

Predictor	Unit	OR (95% CIs)	z value	p
Intercept		NA	-52.502	< 0.001
	England	1	-	-
Country	Scotland	0.719 (0.598,0.864)	-3.524	< 0.001
	Wales	0.060 (0.046,0.078)	-21.041	< 0.001
	Mixed	1	-	-
Species	Cattle	0.659 (0.578,0.750)	-6.289	< 0.001
	Sheep	0.574 (0.470,0.701)	-5.453	< 0.001
Farm area	$log_{10}(Ha+1)$	4.260 (3.756,4.832)	22.550	< 0.001
Cattle density	$\sqrt{10{-}3}Head$	6.965 (6.200,7.823)	32.724	< 0.001
Sheep density	$\sqrt{10^{-4}}Head$	6.717 (5.850,7.711)	27.039	< 0.001

Table 4.9: Logistic regression output for the multivariate model without the distance to seed variable.

A comparison between the model without distance to seed and the full model is presented in Figure 4.4. The principal change is in the animal density predictors which are much stronger risk factors without distance to seed. However, the country risk factor, in particular Wales also changes significantly. There is no change in the holding level factors (species and farm area).

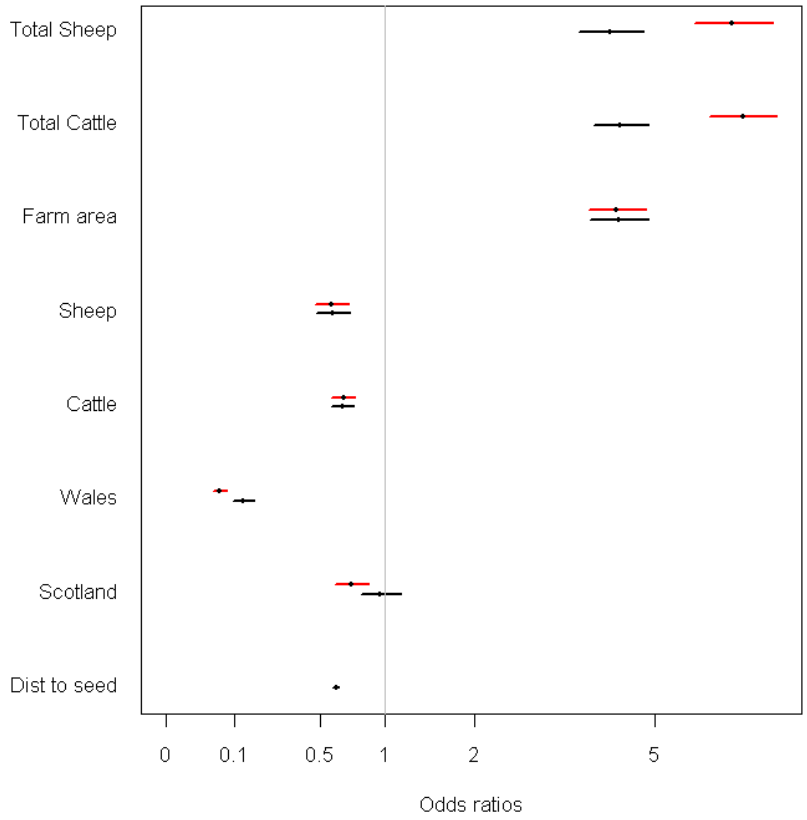


Figure 4.4: Odds ratios (points) and 95% CIs (lines) for the full model (Table 4.8, black lines) and the model without distance to seed (Table 4.9, red lines). The x-axis is square root transformed.

There was significant spatial autocorrelation in the models presented in Table 4.9 and Table 4.8, this was corrected by overlaying a 5km hexagonal lattice on the distribution of farms and inserting the hexagon to which each farm belonged into the model as a mixed effect. The median number of farms in each hexagon was 8 (25th percentile = 4, 75th percentile = 15, Figure 4.5). The area under the ROC for this model is 0.981 and it correctly identifies 46.6% of IPs.

The results of the GLMM are presented in Table 4.10 and comparison of the risk factors from the GLMM and the GLM is presented in Figure 4.6. The confidence intervals for the predictors in which there is some spatial dependance are considerably wider for the GLMM compared to the GLM. These variables are the distance to the seed, country and cattle and sheep densities. The point estimates change slightly for species and holding area are slightly different, but the confidence intervals are the same. The change in the spatially dependant variables reflects that there is little

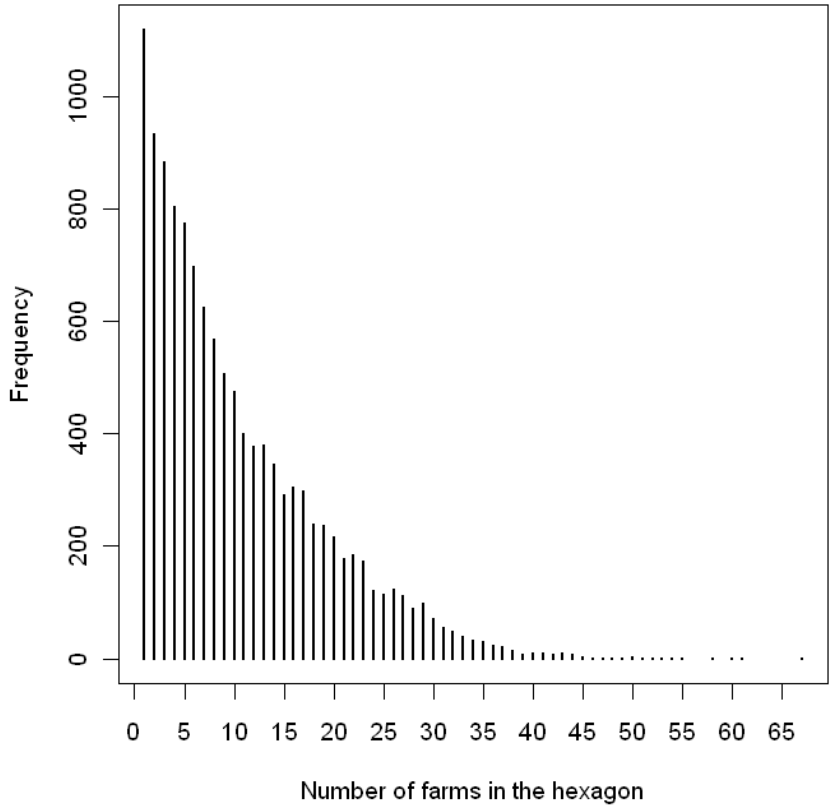


Figure 4.5: Distribution of the number of farms in each hexagon.

within-hexagon variation in the spatially dependant predictors.

Predictor	Unit	OR (95% CIs)	z value	p
Intercept		NA	-14.55	< 0.001
dist to seed	$\sqrt{km}$	0.525 (0.456,0.604)	-8.944	< 0.001
Country	England	1	-	-
	Scotland	1.209 (0.607,2.406)	0.539	0.590
	Wales	0.188 (0.082,0.431)	-3.951	< 0.001
Species	Mixed	1	-	-
	Cattle	0.611 (0.518,0.722)	-5.809	< 0.001
	Sheep	0.530 (0.415,0.678)	-5.061	< 0.001
Farm area	$log_{10}(Ha+1)$	4.715 (3.980,5.586)	17.93	< 0.001
Total cattle	$\sqrt{10^{-3}head}$	3.207 (2.112,4.870)	5.470	< 0.001
Total sheep	$\sqrt{10^{-4}head}$	3.246 (1.916,5.499)	4.377	< 0.001

Table 4.10: Multivariate logistic mixed model of risk factors for being an IP.

Comparison of the main effects model (Table 4.10) with a model in which the outcome is IPs infected during phase 1 (Table 4.11) of the epidemic is shown by Figure 4.7. The only significant difference between the two models is in the distance

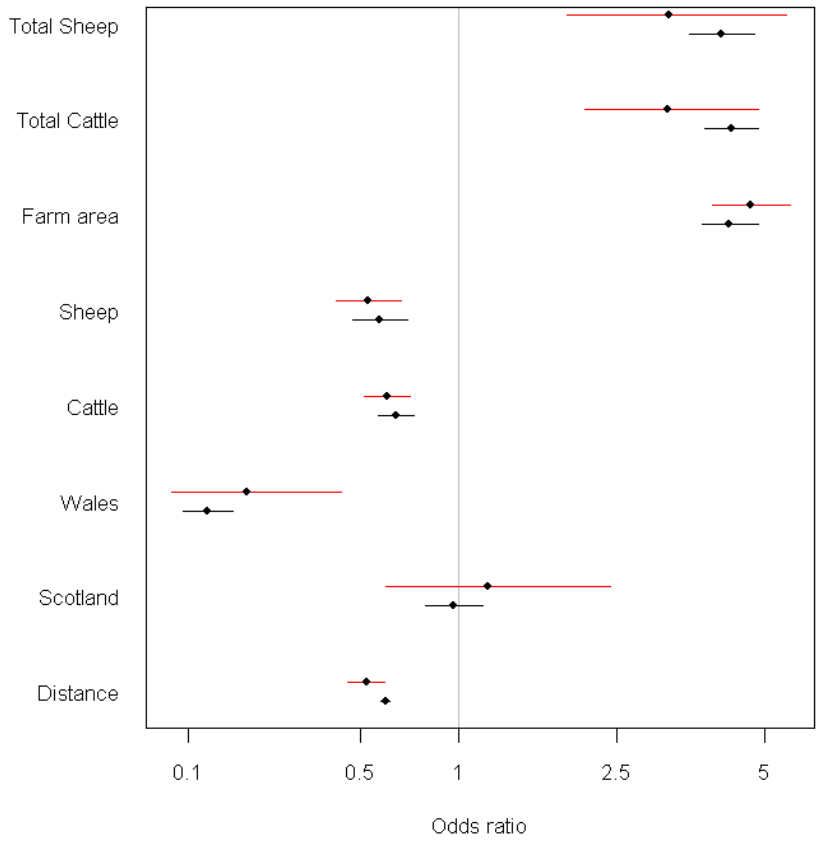


Figure 4.6: Odds ratios (points) and 95% CIs (lines) for the logistic GLM (Table 4.8 - black lines) and the logistic GLMM (Table 4.10 - red lines).

to seed predictor which is a significantly stronger risk factor in the phase 1 only model. These are likely to be linked, as the majority of cases during phase 1 were in Wales and Wales has abnormally high sheep densities, which in other models were not a major risk factor. The area under the ROC for this model is 0.989 and it correctly predicted 294 (41.0%) of 717 phase 1 IPs.

Predictor	Unit	OR (95% CIs)	z value	p
Intercept		NA	-9.009	< 0.001
dist to seed	$\sqrt{km}$	0.359 (0.285,0.452)	-8.683	< 0.001
	England	1	-	-
Country	Scotland	1.316 (0.526,3.288)	0.587	0.557
	Wales	0.532 (0.171,1.653)	-1.091	0.275
	Mixed	1	-	-
Species	Cattle	0.518 (0.413,0.649)	-5.698	< 0.001
	Sheep	0.492 (0.341,0.709)	-3.803	< 0.001
Farm area	$log_{10}(Ha + 1)$	4.704 (3.694,5991)	12.55	< 0.001
Cattle density	$\sqrt{10^{-3}head}$	3.647 (1.993,6.676)	4.195	< 0.001
Sheep density	$\sqrt{10^{-4}head}$	1.048 (0.465,2.363)	0.113	0.910

Table 4.11: Multivariate logistic GLMM of being an IP in phase 1.

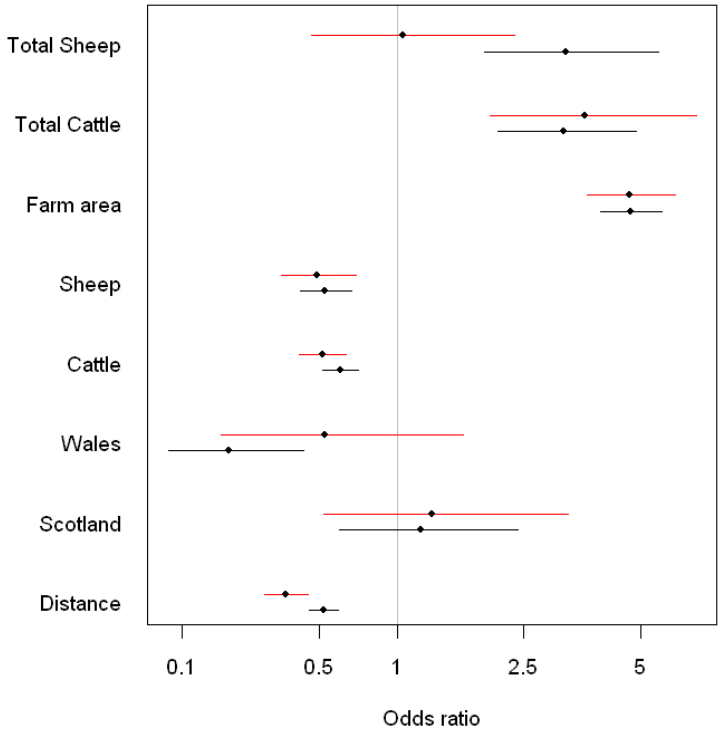


Figure 4.7: Odds ratios (black points) and 95% confidence intervals (lines) for the main effects model presented in Table 4.8 (black lines), and the model of being an IP in phase 1 (red lines). Odds ratios are plotted on a square root transformed scale.

4.3.6 Sensitivity analysis

The logistic regression model parameterised with all culled holdings is presented in Table 4.12. Comparing the odds ratios for the risk factors in this model with those for the full model (Figure 4.8) shows changes in some of the predictors:

1. **Country.** Being in Scotland is not significantly different to England for the IP model, however it is a significant risk factor for being a culled holding and significantly different to its value in the IP model.

2. **Species.** Sheep only farms are at significantly greater risk in the culled model, however it remains a protective effect relative to mixed holdings.

3. **Farm area.** Whilst still a risk factor farm area is significantly less of a risk factor in the culled model.

The model has an area under the ROC of 0.956 and correctly identifies 67.5% of culled holdings.

Predictor	Unit	OR (95% CIs)	z value	p
Intercept		NA	-20.72	< 0.001
dist to seed	$\sqrt{km}$	0.490 (0.466,0.515)	-27.90	< 0.001
Country	England	1	-	-
	Scotland	4.463 (3.484,5.717)	11.84	< 0.001
	Wales	0.161 (0.118,0.219)	-11.66	< 0.001
Species	Mixed	1	-	-
	Cattle	0.488 (0.443,0.536)	-14.82	< 0.001
	Sheep	0.811 (0.728,0.904)	-3.769	< 0.001
Farm area	$log_{10}(Ha + 1)$	2.118 (1.963,2.287)	19.26	< 0.001
Total cattle	$\sqrt{10^{-3} head}$	2.551 (2.184,2.979)	11.83	< 0.001
Total sheep	$\sqrt{10^{-4} head}$	4.284 (3.501,5.243)	14.12	< 0.001

Table 4.12: Multivariate logistic regression model of risk factors for being a culled holding.

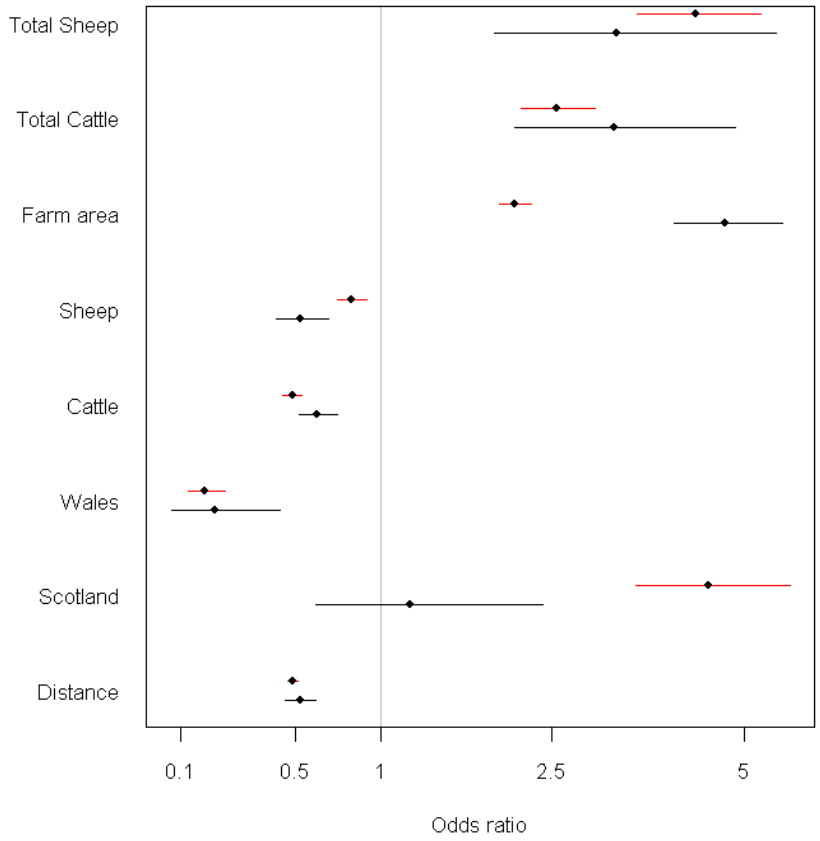


Figure 4.8: Odds ratios (black points) and 95% confidence intervals (lines) for the full single term model presented in Table 4.8 (black lines), and the model of all culled holdings presented in Table 4.12 (red lines). Odds ratios are plotted on a square root scale.

Analysis discounting laboratory negative IPs did not produce a significantly different model compared to that of all IPs (Figure 4.9, Table 4.13 for non-laboratory negative IPs and Table 4.8 for all IPs). The area under the ROC for this model is 0.987 and correctly identified 43.7% of IPs. None of the predictors for the non-negative IPs were significantly different to those for all IPs although there is a large difference in the odds ratios for the cattle (presence and density) and farm area have substantially greater odds ratios whilst distance to seed is less of a risk factor.

Predictor	Unit	OR (95% CIs)	z value	p
Intercept		NA	-12.14	< 0.001
dist to seed	$\sqrt{km}$	0.583 (0.476,0.714)	-5.224	< 0.001
	England	1	-	-
Country	Scotland	0.854 (0.325,2.247)	-0.320	0.749
	Wales	0.117 (0.027,0.500)	-2.898	0.003
	Mixed	1	-	-
Species	Cattle	0.747 (0.613,0.910)	-2.893	0.003
	Sheep	0.335 (0.232,0.483)	-5.845	< 0.001
Farm area	$log_{10}(Ha+1)$	6.116 (4.893,7.645)	15.91	< 0.001
Cattle density	$\sqrt{10^{-3}head}$	3.246 (1.790,5.888)	3.876	< 0.001
Sheep density	$\sqrt{10^{-4}head}$	4.062 (1.862,8.862)	3.522	< 0.001

Table 4.13: Multivariate logistic regression model of risk factors for being a laboratory positive IP.

The results of all models are summarised in Table 4.14, this shows that the model fit and the accuracy of IP prediction of the models do not always match. Some models (phase 1, laboratory positive) have a larger area under the ROC but a lower predictive accuracy than the full model. Whilst others (culled holdings) work the other way around.

Model	Area under ROC	% correctly identified
GLM	0.908	25.2
GLMM	0.981	46.6
Laboratory positive IPs	0.987	43.7
All culled holdings	0.956	67.5
Phase 1 IPs	0.989	41.0

Table 4.14: Summary of results of different multivariate models developed.

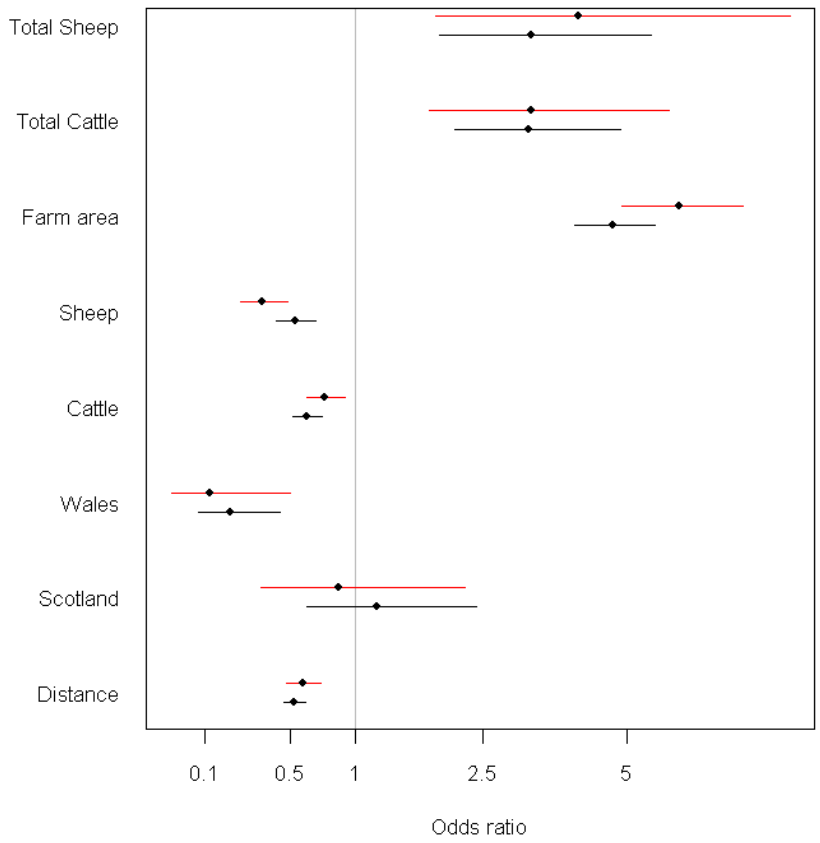


Figure 4.9: Odds ratios (black points) and 95% confidence intervals (lines) for the main effects model presented in Table 4.10 (black lines), and the model of laboratory positive holdings presented in Table 4.13 (red lines). Odds ratios are plotted on a square root scale.

4.4 Discussion

A static model of holding level susceptibility has been developed using a set of variables similar to those used by Keeling et al. (2001) and Tildesley et al. (2006) but implemented in a random effects logistic regression framework. Such a model is of value for developing understanding of the factors responsible for driving and determining the course of the 2001 UK FMD epidemic. Additionally, this model could be used to identify holdings at greatest risk of experiencing a local epidemic in the event of a future FMD outbreak. The model was an excellent fit (area under the ROC of 0.981) and was able to predict 46.6% of holdings which would go on to become IPs, this represents an improvement on the model of Tildesley et al. (2006) whose figure was 15% of IPs correctly identified (Tildesley et al., 2008).

The principal risk factor for FMD infection was the distance to the nearest of the 78 seeds infected before the imposition of the NMB. This variable describes the locations of holdings in relation to the distribution of the FMD virus at the beginning of the epidemic. Previous studies (Pfeiffer et al., 2007) have dealt with the location of a virus introduction for Highly Pathogenic Avian Influenza (HPAI) by including region into the model where the virus can be seen to be clustered in certain region(s). This paper develops this by treating disease spread as a set of point sources.

Analysing solely species on the holding and holding area also produces a strong predictive model, because it was larger holdings with cattle (and greater risk if sheep were also present) which became infected. Animal numbers in the locality represent the infection challenge to the susceptible holding and are significant predictors, cattle numbers more so than sheep numbers. Densities were calculated by applying a kernel to the farming landscape and a range of kernel shapes and sizes were evaluated. An exponential kernel with a bandwidth of 15km performed slightly better than either a quartic or flat kernel (Figure 4.3). Examination of Figure 4.3 shows that the exponential kernel had not reached its minimum by 15km. Whilst the shape of the plots for the quartic and flat kernels are similar, the shape of the exponential kernel is very different with a much greater range of AIC values, this is shown by the minimum of the quartic and flat kernels which was at 5km whilst the exponential kernel is at

15km. The shapes of these kernels are shown in Figure 4.10. The area in the 2 or 3km immediately surrounding the IP are very important for determining the transmission challenge and hence exponential kernels at smaller bandwidths which ignore this area have poor predictive power. Additionally, the area beyond the immediate 2 or 3km has some influence on susceptibility, underlining the importance of a 'sea' of livestock when determining risk of FMD infection. Hence capturing this area is necessary, and the exponential kernel outperforms the other shapes in this respect.

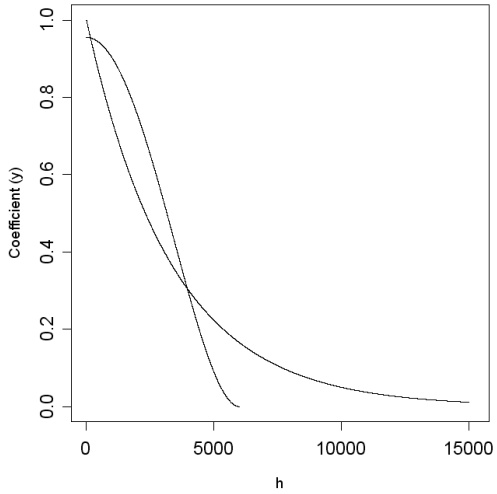


Figure 4.10: Comparisons of the shapes of the quartic and exponential kernels at 6 and 15km respectively, the values for bandwidth at which they returned the minimum AIC value.

Both cattle and sheep densities were significant predictors, although cattle densities are more important ($F_{1,125,223}$=710 and $F_{1,125,223}$=368 for cattle and sheep respectively), probably as a result of the quantity of virus excreted by cattle (Donaldson et al., 2001). There is little correlation between cattle and sheep densities (r=0.084) because at the high end of the distribution of cattle numbers sheep numbers are lower, simply because in areas of intense cattle production there is little room for extensive sheep production. Areas of high cattle density were generally close to a seed and it was areas close to a seed with a lot of cattle which were at greatest risk.

In a multivariate model, species presence on the holding and animal densities gives significant predictive power over just distance to seed. Due to the different transformations applied to the variables in Table 4.8 direct comparison of the effects of variables using odds ratios is difficult. However, analysis of the z-scores shows that

distance to seed is the most important predictor, cattle densities are a stronger risk factor than sheep densities. This again suggests that cattle are major driving forces of the epidemic probably as a result of their greater susceptibility and potential to transmit the virus (Donaldson et al., 2001).

There is a strong country level effect with being in Wales or Scotland apparently being protective relative to England. The protective effect of being in Scotland is only a factor once the larger (area) farm holdings in its FMD regions have been taken into account. The high number of large farms in these regions should result in higher predicted risk, however this was not the case and hence there is a large country level effect for Scotland and interaction with farm area. Furthermore the nature of farming in the FMD regions in Scotland is very different to farming throughout the rest of the country, where holdings tend to cover large areas and are sparsely populated, unlike the high intensity holdings in the South West. Some of the protective effect of being in Wales may be explained by discrepancies between census animal numbers and those recorded in the FMD data (Chapter 3) with sheep numbers apparently greatly over reported in the census. There were also differences in animal densities observed at the country level, in England IPs are characterised by higher cattle and sheep densities, in Wales by higher cattle densities and in Scotland there are no differences between IPs and non-IPs. This may reflect the degree to which the virus was seeded in the respective countries and therefore the nature of the FMD regions. In England the virus was spread widely through the country including to regions such as Leicestershire, Essex and Kent where there is little livestock production. However, in Scotland and Wales the virus was only seeded in major livestock production regions.

As a result of the inherently spatially autocorrelated nature of the FMD epidemic there was significant spatial autocorrelation in the residuals of the initial GLM. This was corrected by inserting a random effect in the form of a hexagonal lattice. Similar approaches have been adopted in previous studies (Pfeiffer et al., 2007), however rather than use political boundaries this study used arbitrary even sized spatial units. The addition of the random effect did not make a significant difference to the estimates of the risk factors. However, for the predictors which are spatially dependant the addition of the random effect did increase the standard errors of the predictors,

this would be the equivalent of reducing the degrees of freedom.

The model is contingent upon the control policies which were implemented. Farms which were culled as non-IPs may have been sub-clinically infected and therefore the spatial and temporal pattern of culling will have influenced the pattern of disease spread. Phase 1 of the epidemic was a period during which non-IP culling was low and was implemented in a relatively uniform manner. An accurate model was parameterized to predict the probability of being an IP during phase 1. This model differed from the main effects model in three ways, namely that the protective effect of being in Wales was reduced and the risk associated with sheep densities was also reduced probably in response to the drop in risk in North Wales (Figure 4.7). Furthermore, as the IPs during this period were tightly clustered around seeds the 'distance to seed' protective effect is greater.

The same model framework can also be parameterised to predict holdings which were culled, not just IPs. In the model in question (Table 4.8) sheep (both presence and densities) are a greater risk factor than cattle reflecting the greater numbers of sheep culled on non-IPs relative to cattle. This is as a result of the tendency for smaller sheep holdings to be culled out pre-emptively and also the use of culling of sheep (particularly 3km culls) as a means to reduce the density of the susceptible population in areas of high animal densities. Importantly the ability to develop a single model which can predict both IP and non-IP culls with some accuracy demonstrates that targeting of non-IP culls was effective by targeting those which were more likely to become infected which were being culled. Furthermore if laboratory positive only holdings are considered the model develops better predictive power as a result of the tendency for laboratory positive holdings to be large cattle holdings (McLaws et al., 2006), rather than sheep holdings which were more likely to be misidentified due to mild clinical signs in sheep.

The model was found to be robust to sensitivity to the input data by performing analysis using the fragment data (Table 4.14, which as discussed in Chapter 3 is not a complete inventory of all farms in the UK. In spite of the modelled nature and the fact that a lot of smaller farms were included in the analysis. This still only reduced the area under the ROC by 0.035, so this methodology is not dependant

upon the use of agricultural census data and the larger farms this data represents.

A highly accurate predictive model of the risk factors for holdings becoming IPs has been developed. The distance to an infectious seed was the principal predictor. After distance is taken into account, animal densities, in particular high stocking densities are important predictor variables. Larger cattle farms are also important. This model is robust to sensitivities in the data and potential autocorrelations in predictor variables. Furthermore it is a model which could be developed further to generate a measure of risk in different epidemic situations.

Chapter 5

Risk factors for local transmission

5.1 Introduction

Epidemiological studies frequently focus on factors which predispose members of the susceptible population to potentially contract an infectious disease if exposed (French et al., 1996; Halliday et al., 2006; Odiit et al., 2006, Chapter 4). However, equally valuable would be an understanding of the members of the population most likely to transmit infection. However, only a few studies have investigated individual farm level risk factors for transmitting infection once an individual has become infected. Such analyses have been applied to *E.coli* super shedders (Matthews et al., 2006; Chase-Topping et al., 2007) and HIV transmission networks (Latora et al., 2006). Although studies have attempted to look for important nodes in FMD transmission networks (Haydon et al., 2003), no studies have attempted to charachterise these farms.

This chapter will develop a model of risk of transmitting infection rather than acquiring infection. The study population is the subset of those farms which had FMD and were therefore capable of transmitting infection. In this respect this model is complementary to the susceptibility model (Chapter 4); together they provide information on the risk that a farm will contribute to the spread of infection.

5.2 Methodology

5.2.1 Data

IP data

The IPs were the 1948 IPs estimated to have been infected after the NMB. For the 37 IPs for which there was no estimated date of infection the estimated date of infection was taken as the reporting date minus 10 days as has been done previously (Savill et al., 2006) and described in Chapter 2. Animal numbers on the IPs were taken as the numbers slaughtered. Data on whether the IP was a dairy holding was taken from the farm list data (Chapter 3).

Non-IP data

Data for non-IPs was taken from the June 2000 agricultural census (Chapter 3), all holdings with cattle or sheep recorded were included.

5.2.2 Variables

Dependent variables

The ideal dependent variable would comprise:

1. *Cases* (hereafter referred to as infectors): IPs that have infected another holding.

2. *Controls* (non-infectors): IPs that have not infected another holding.

Such a variable requires a complete epidemic tree, this is because data on which farm infected other farms is required to generate the data on infectors. To generate the non-infectors, data on which farms did not infect other farms is required, therefore the epidemic tree is still needed. Data for infectors can be derived from the transmission tracing data (Chapter 2). However, the tracing data is only partially complete and, as a consequence, non-infectors can not be reliably identified.

For the purpose of these analyses the source of infection for a holding will be taken as the nearest possible source. A daughter IP (the secondary IP resulting

from an *infector* (the parent)) must be estimated to have been infected within the temporal infectious window of a source IP (Chapter 2). Euclidean distance is the strongest determinant of local FMD transmission (Chapter 4, Keeling et al. (2001)); 52% of sources of infections identified by the tracing data was the nearest possible source (Chapter 2). This method of nearest neighbour method matching is one of the methods employed by Haydon et al. (2003) to construct epidemic trees (Chapter 1).

This methodology ensures that all IPs are assigned a parent IP. Further selection removes all infection events over distances greater than 3km. This is to ensure that only local transmission is analysed. This selection process gives rise to an epidemic tree with a number of disconnected branches (Figure 5.1). Local spread is analysed because as described previously (Chapter 2) the infection processes operating locally and at larger scales may be very different.

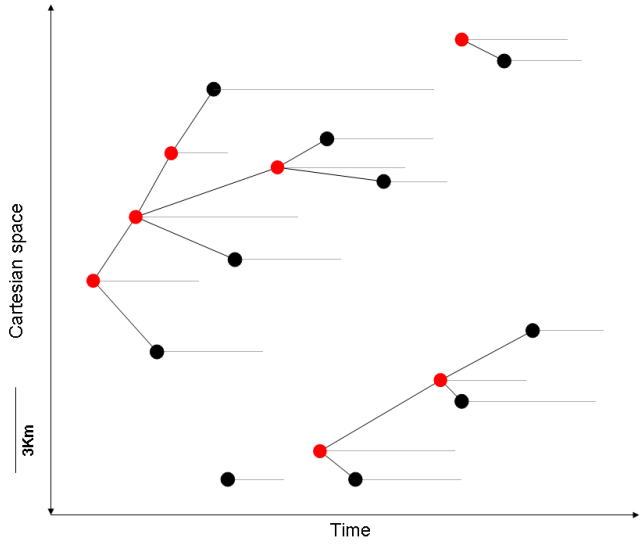


Figure 5.1: An example of an epidemic tree created using the source selection process. The y-axis represents the spatial component to the selection process and the x-axis the temporal component. Red points represent IPs which have generated subsequent IPs (infectors) and black points IPs which have not (non-infectors). Black lines are routes of infection over distances shorter than 3km and grey lines are the infectious period of the IPs.

As a result of this methodology the model is not strictly a model of FMD transmission, rather it is a model of the likelihood of a subsequent IP occurring within 3km of an IP. However, this is not a fundamental problem because in the event of a future epidemic being able to predict which IPs are most likely to trigger further

IPs is of less value than simply identifying the likelihood of there being an IP in the locality. This is because when targeting resources, the source of a case is of little importance, rather, authorities need to know where that case is likely to be, and this model could be used to derive putative case reproduction values. In this respect the analyses in this chapter are building upon the model of susceptibility (Chapter 4), which is looking at static start of epidemic risk of infection. Rather than a static measure, this chapter will build a generation by generation model of infection risk.

Confounders

The daughter IP must fall within the infectious window of the parent IP, therefore the longer the infectious window of the parent IP the greater the probability of finding a daughter IP. To allow for this the duration of the infectious period was included in these analyses as a covariate. Infectious period may also be a determinant of transmission, however it is not possible to identify whether any effect this predictor has is a result of the effect of infectious period on transmission or its influence on how the outcome was derived.

Predictor variables

The predictor variables fall into three broad categories: control effort factors, holding factors and locality factors. The control effort can be modelled as a three level factor corresponding to the three stages in the culling strategy (Chapter 2), the stages are as follows:

1. **Phase 1**: IP and DC culling. 23rd February to 26th March
2. **Phase 2**: Full CP and extended culling. 26th March to 29th April.
3. **Phase 3**: Reduced CP culling. 29th April onwards.

Holding level factors comprise:

1. The laboratory result of the IP. Three levels: positive, not tested and negative, with positive the reference level (Chapter 2).

2. The DCC of the IP (Chapter 2). Due to the large number of DCCs many of which have relatively few IPs (Chapter 2) only those with more than 100 IPs will be used in these analysis. The remainder will be grouped into a Rest of the UK category. This creates a six level factor: Carlisle, Ayr, Exeter, Newcastle, Leeds and Rest GB, with the Rest GB the reference level.

3. The distance to a seed IP (see Chapter 4).

4. Species on the holding. Three level factor: whether a farm is mixed cattle and sheep, cattle only and sheep only. Mixed is the reference level. Numbers of animals cannot be used due to the high number of IPs with either no cattle (319, 16.3%) or no sheep (295, 15.2%).

5. Whether there are identified satellite holdings (Chapter 3) linked to the IP.

6. Whether the holding is listed on the June 2000 agricultural census. This is because not all IPs were identified on the agricultural census (Chapter 3) and there may be differences between those which were on the census and those which were not.

Factors which relate to the locality of the holding are the density of animals of each species and the density of susceptible holdings in the vicinity of each IP which were still standing when the IP was slaughtered. The 'vicinity' was defined in two ways:

1. **3km of the IP**. All holdings within a 3km ring of the IP are weighted equally. This method describes the locality of the IP as defined by the epidemic protection zone and gives a measure of the nature of farming in the area into which infection can be transmitted according to the outcome variable. The total number of sheep, cattle and susceptible premises within 3km is divided by $\pi \times 3^2$ to give a measure of number of animals per km^2.

2. **Voronoi polygons around the IP.** Voronoi polygons assign all land to the nearest holdings. The effect of overlaying 3km rings onto this is described by Figure 5.2. The 3km ring and Voronoi polygons were intersected to form aerial units in which the boundary was either the 3km ring or the polygon vertex.

The polygon vertex was used if it was less than 3km away and the 3km ring otherwise. This was implemented using Arc Macro Language (AML) script in ESRI ArcInfo v9.0. For each day of the epidemic, the script generates Voronoi polygons around all IPs infected on that day. For all IPs which were estimated to have been infected on that day 3km buffers are generated. These are clipped using the Voronoi polygons and the remaining area is the area of influence. This ensures that each farm is only counted once in the predictor variables and thus ensures independence by having no overlapping 3km rings.

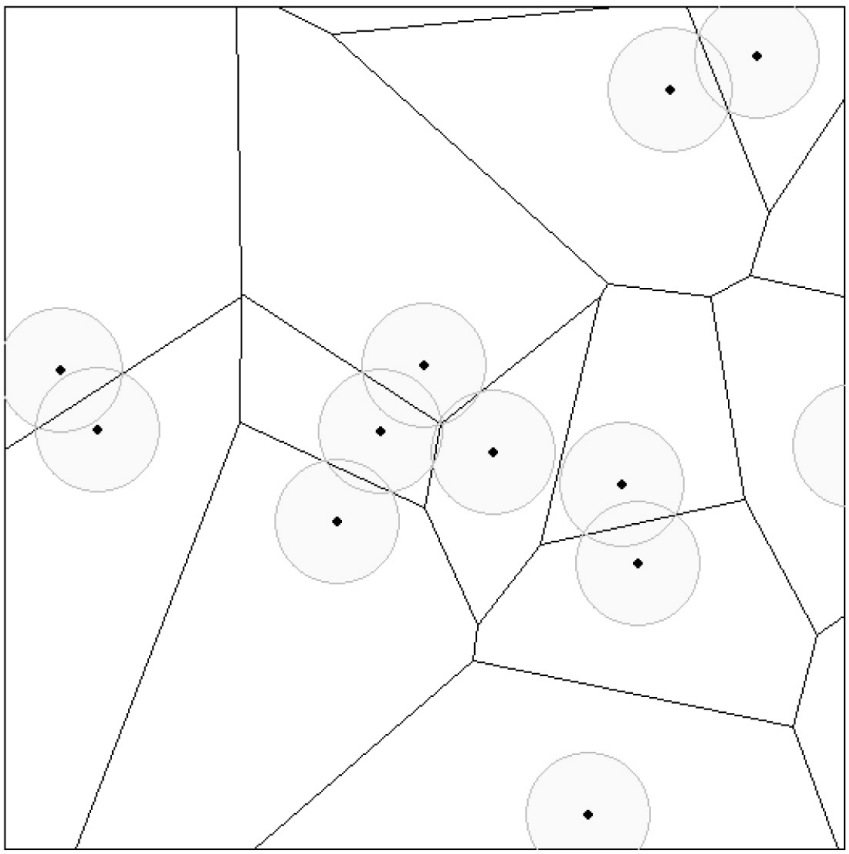


Figure 5.2: An example map of the distribution of IPs in an area of Cumbria in relation to polygons defining their area. Polygons represent Voronoi polygons constructed for each IP (black points) and circles represent 3km rings around each IP.

The principal method of describing the locality will be the 3km ring as this is simpler and more intuitive. However the overlap of rings is a concern regarding independence as farms can be counted several times. As a result, the results of the Voronoi polygon analysis will be compared to results of modeling using the 3km rings.

5.2.3 Statistical analyses

The transmission risk model will be developed by analysing the outcome variable against the predictors in a logistic regression framework. The methodology will follow that employed in section 4.2.3 with univariate analysis, construction of the main effects model and finally incorporating interactions whilst checking linearity in the logits. Model fit will be assessed using the area under the ROC curve (section 4.2.3).

To test the validity of the outcome variable the epidemic tree constructed in these analyses will be compared to the tracing data to analyse differences between the datasets. Sensitivity analyses using the tracing data are complicated because the tracing data does not match all IPs to a source and could result in a model of the probability of having a source identified rather than risk of transmission as such. To overcome this, only infectors in the form of IPs with a daughter IP within 3km will be identified using the tracing data. Therefore, a second outcome will be defined in which non-infectors will be holdings for which both the tracing and nearest neighbour methodology record no daughter within 3km. The infectors will be those for which both methods identify a daughter in 3km. This outcome variable will be used for sensitivity analysis.

Model sensitivities to the source data and to the different areal units for animal densities and laboratory confirmation of IPs will be checked using odds ratios for the predictors. The sensitivity analysis looks for a statistically significant shift in the odds ratios, therefore a term in 2 models is significantly different if the 95% confidence intervals of the odds ratios do not overlap.

Potential spatial autocorrelation between predictors will be tested by running a model which includes a term which is whether there is another infector within 3km. Comparing the odds ratios for a model including these predictors will allow for an assessment of spatial autocorrelation in the locations of infectors versus non-infectors.

Modelled values will be generated for each data point using the predict function in R and plotted over the time course of the epidemic. This generates a likelihood of being an infector for each data point. Plotting modelled values for each data point against date of infection would be difficult to interpret because there are many data points for most days. To overcome this, the supersmooth function in R (R

Development Core Team, 2004) is used to interpolate between the points and create a line which summarises the data points. Continuous predictor variables can be overlayed by taking the value for each data point as a proportion of the maximum value for that variable, and smoothing the data to create a line. Discrete variables must be handled as binary variables in which positives are treated as 1s and negatives as 0s and the data smoothed using those variables.

5.3 Results

5.3.1 Outcome variable

There were 867 (44.5%) of the 1948 IPs estimated to have been infected after the NMB which were the nearest possible sources of infection for another IP within 3km (infectors). A source within 3km was identified for 1255 of the 1948 IPs. For the remaining 693 IPs the nearest possible source was more than 3km away. The majority (603, 69.6%) of holdings with daughter IPs have only one daughter (Figure 5.3).

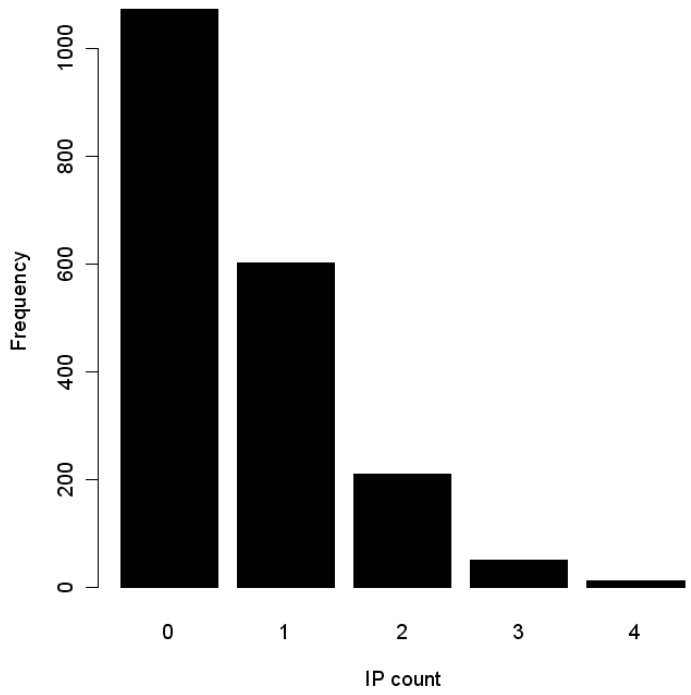


Figure 5.3: Frequency plot of the numbers of daughter IPs for each putative source infected after the NMB.

The peak and trough in numbers of infectors (IPs with near IPs) occurred earlier in the epidemic than for non-infectors (Figure 5.4); this is statistically significant (Kolmogorov-Smirnov D = 0.143, p<0.001). This is because the outcome is whether there is a subsequent generation and therefore the epidemic curve for non-infectors will inevitably be at least one generation after that of infectors.

Cumbria (represented by the Carlisle DCC) contains the greatest proportion of infectors (Table 5.1). There are statistically significant differences between all DCCs (χ^2_5=31.1, p<0.001) with Cumbria being statistically significantly different when compared to all other regions of GB (χ^2_1=26.4, p<0.001) as well as when compared to the other major DCCs (ie excluding the Rest of the GB category) (χ^2_1=13.5, p<0.001). However, comparisons between the Rest of the GB and DCCs excluding Cumbria reveals no statistically significant difference (χ^2_1=3.35, p=0.067).

The methodology used in this study identified a source of infection for 1,255 IPs. The corresponding number identified using the tracing data (chapter 2) was 953, of which 567 were infected over less than 3km. There were 192 instances of the nearest neighbour source being the same as the tracing source (Table 5.2), in a further 304

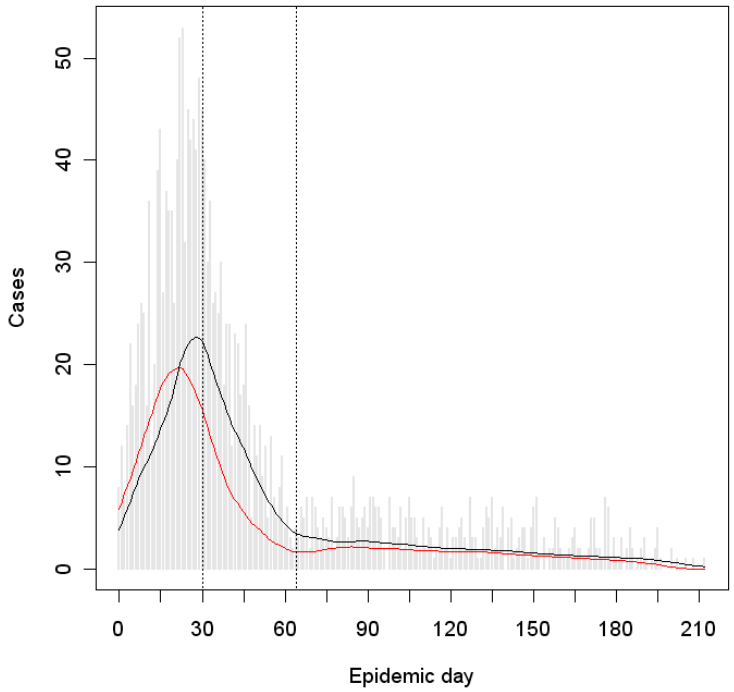


Figure 5.4: Epidemic timeline of the numbers of IPs per day (grey bars). Overlayed onto this are supersmoothed lines of numbers of infectors (red line) and non-infectors (black line). Dashed vertical lines are the cut-offs between epidemic phases.

DCC	non-Infector	Infector	% Infector (95% CIs)
Carlisle	427	451	51.4 (48.1,54.7)
Rest GB	269	154	36.4 (32.0,41.1)
Newcastle	105	73	41.0 (38.0,48.4)
Ayr	93	75	44.6 (37.3,52.2)
Exeter	100	62	38.3 (31.1,45.9)
Leeds	79	60	43.2 (35.2,51.5)

Table 5.1: The number of infectors and non-infectors by DCC. The percentage column is the percentage of IPs which were infectors within each DCC.

instances the tracing data and nearest neighbour data both identified a source within 3km but identified a different source. In 228 instances the tracing data recorded a source greater than 3km away and in 75 of these instances the nearest neighbour data matched to a source within 3km. There were 71 instances of the tracing data matching to a source within 3km where the nearest neighbour data did not find a source. This is because when the tracing data was being cleaned (Chapter 2) confidence intervals were placed on dates of infection. This was not done when identifying sources in this chapter because in this analysis, sources were to be identified with certainty and

therefore margins to allow for uncertainty were not included.

Near source YN	Tracing source YN	Tracing distance	Agreement	Count
Y	Y	<3km	Match	192
Y	Y	<3km	No match	304
Y	N	NA	No match	683
Y	Y	>3km	No match	75
N	Y	<3km	No match	71
N	Y	>3km	No match	153
N	N	NA	No match	469

Table 5.2: Whether an IP had a source identified by the nearest neighbour method used in this chapter (near source), whether it had a source identified in the tracing data (tracing source) and whether the tracing and nearest neighbour sources match (agreement).

The statistic that is important when comparing the tracing data to these nearest neighbour methods is the percentage agreement between the outcome as defined using nearest neighbour method and the outcome as defined by the tracing data. This is defined as all IPs which are the identified source of another IP within 3km. A total of 454 IPs were sources of IPs within 3km according to the tracing data, of these 357 (78.6%, 95% CIs = 74.6,82.2) were also identified as infectors by the nearest neighbour method (Table 5.3). The row percentage column in Table 5.3 shows that a significantly greater percentage of the tracing identified sources were also sources by nearest neighbour matching than those which were not.

		Tracing data		
		source in 3km	no source in 3km / no data	% (95% CIs)
Nearest neighbour	infector	357	510	41.2 (37.9,44.5)
	non-infector	97	982	9.0 (7.4,10.8)
	Total	454	1492	

Table 5.3: Accuracy of identification of the outcome variable by comparing the nearest neighbour data to the tracing data. The percentage column is the source column as a percentage of the row total.

5.3.2 Univariate analysis

5.3.3 Infection factors

The results of univariate analysis of the predictor variables are shown in Table 5.4.

Predictor	Unit	non-inf	inf (%)	OR (95% CIs)	z value	p
Infectious Period	log_{10} Days	NA	NA	17.9(6.91,46.5)	5.931	< 0.001
Phase	1	428	467(52.2)	1	-	-
	2	385	201(34.3)	0.51(0.41,0.63)	-6.120	< 0.001
	3	260	207(44.3)	0.72(0.58,0.91)	-2.815	0.005
DCC	Rest GB	269	154(36.4)	1	-	-
	Carlisle	427	451(51.4)	1.89(1.49,2.40)	5.216	< 0.001
	Newcastle	105	73(41.0)	1.26(0.88,1.80)	1.244	0.214
	Ayr	93	75(44.6)	1.49(1.04,2.15)	2.159	0.031
	Exeter	100	62(38.3)	1.06 (0.73,1.55)	0.317	0.751
	Leeds	79	60(43.2)	1.33 (0.90,1.97)	1.438	0.150
dist to Seed	$\sqrt{km}$	NA	NA	0.92 (0.87,0.98)	-2.772	0.006
Lab Result	Positive	648	612(48.6)	1	-	-
	No result	171	129(43.0)	0.77 (0.60,0.99)	-2.011	0.044
	Negative	257	131(33.8)	0.53 (0.41,0.67)	-5.274	< 0.001
Species	Mixed	711	621(46.6)	1	-	-
	Cattle	172	123(41.7)	0.82 (0.63,1.06)	-1.535	0.125
	Sheep	196	123(38.6)	0.72 (0.56,0.92)	-2.593	0.010
Satellites	N	393	296(43.0)	1	-	-
	Y	683	576(45.8)	0.93 (0.77,1.12)	-0.813	0.416
Susceptible density	holds/km^2	NA	NA	1.64 (1.36,1.98)	5.183	< 0.001
Cattle density	head/km^2	NA	NA	1.19 (1.15,1.23)	9.546	< 0.001
Sheep density	head/km^2	NA	NA	1.02 (1.01,1.04)	3.226	0.001

Table 5.4: Univariate logistic regression analysis of all predictors. For categorical-scale factors the numbers of infectors and non-infectors which fall into each factor level are given by the non-inf and inf columns.

Infectors had significantly longer infectious periods than non-infectors (Table 5.4, Figure 5.5). However the magnitude of the difference is relatively small, the difference between infectors and non-infectors was 0.7 days when the mean values are compared, and 1 day if the median values are compared (Figure 5.5). A log_{10} scale was applied to this variable to ensure linearity in the logits and this is reflected in the per unit odds ratios.

Analysis of the three epidemic phases shows that the second and third phases of the epidemic contained statistically significantly fewer infectors compared to the first phase (Table 5.4), as well as statistically significant differences between the first and third phase although the magnitude of this effect is not as strong. There are also

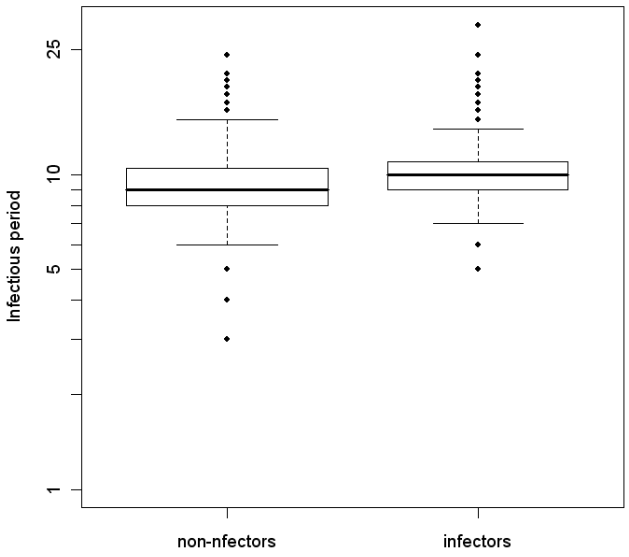


Figure 5.5: Boxplot of infectious period against outcome, plotted on a log_{10} scale. The widths of the boxes represent the proportion of the data in that group. One infector with a value of 70 has been left off this plot.

statistically significantly more infectors in phase 3 compared to phase 2 (OR = 1.52, 95% CIs = 1.19, 1.96).

For DCCs Carlisle and Ayr were significantly different to the Rest of GB category and were at higher risk (Table 5.4). The remaining DCCs were not statistically significantly different. Post-hoc analysis was carried out by merging the Rest of GB category with the non-significant DCCs. The Ayr DCC was no longer statistically significantly different (p>0.2), however Cumbria remained statistically significant (OR=1.62, 95% CIs = 1.35,1.94, p<0.001).

IPs which were either negative or recorded no result for FMD in the laboratory were statistically significantly less likely to be an infector than those which were positive (Table 5.4). However the magnitude of the effect is much greater between positive and negative than between the positive and no result groups. Furthermore the no result and negative groups were statistically significantly different (p=0.017) in terms of the outcome.

The distance to an initially infected seed was statistically significantly less for infectors than non-infectors (Table 5.4). Infectors are on averaged 1.5km closer to a seed than non-infectors (Figure 5.6).

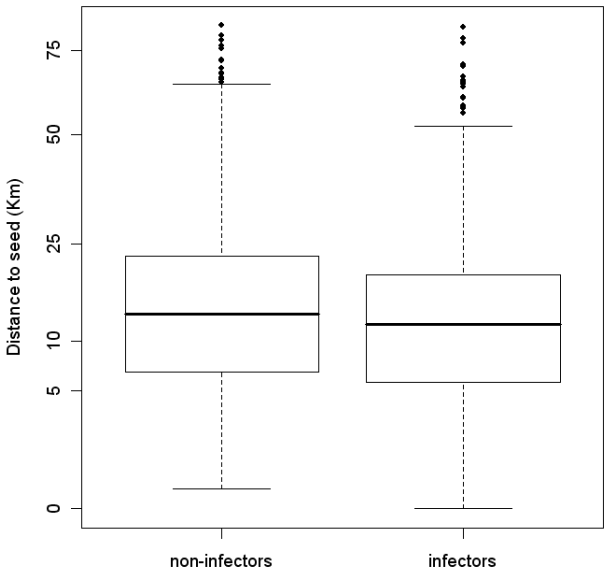
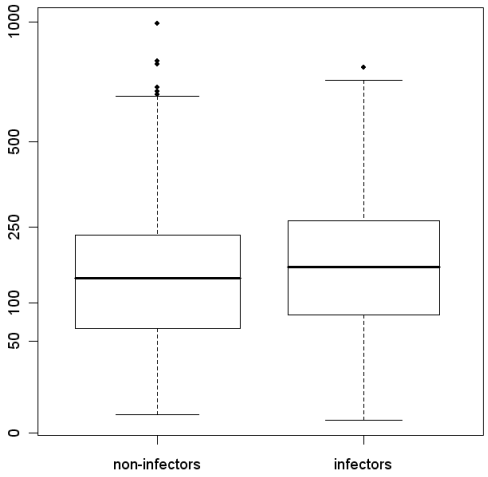


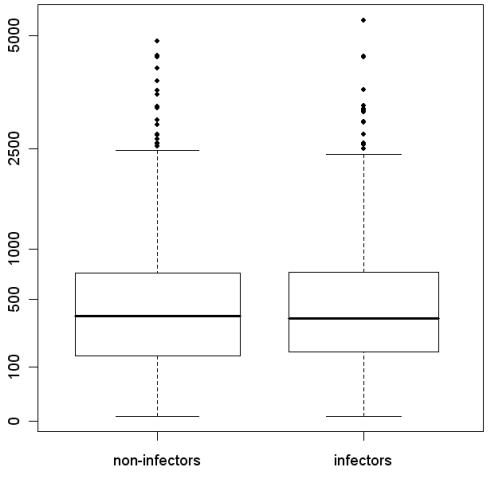
Figure 5.6: Boxplot of distance to an infectious seed against the outcome. The y-axis is square root transformed.

Sheep-only holdings are at significantly lower risk than mixed holdings (The species variable in Table 5.4). When animal numbers are analysed there were statistically significant differences when cattle numbers were analysed on those holdings which contain cattle (p=0.004) in a positive direction (OR=1.35, 95% CIs=1.12,1.64), although there were no significant differences for sheep (p>0.05) (Figure 5.7). This becomes evident when comparing holdings above and below the population mean (Table 5.4).

Total numbers of cattle within 3km are also statistically significantly higher on infectors compared to non-infectors (Table 5.4, Figure 5.8). There was also a significant difference when sheep are analysed, albeit the strength of the effect is of a substantially smaller magnitude (Table 5.4).

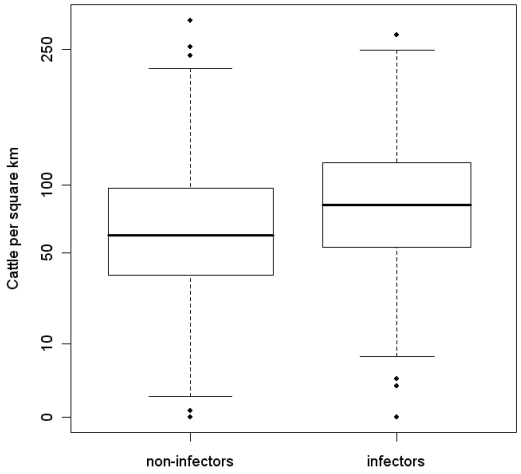


(a) Cattle

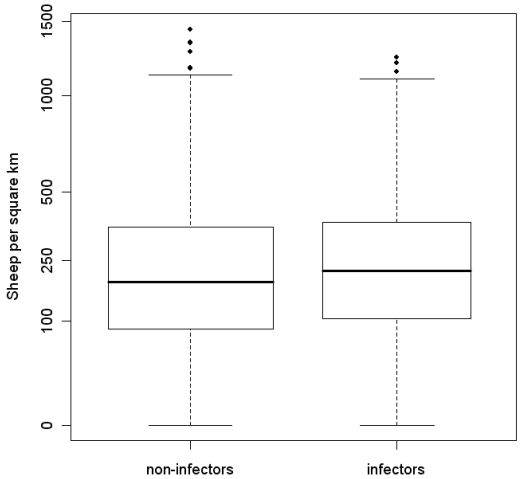


(b) Sheep

Figure 5.7: Boxplots of Cattle and sheep numbers by outcome. Data points are holdings which have cattle or sheep respectively. Widths of boxes represent the number of observations in each group. The y-axes have been square root transformed.



(a) Cattle



(b) Sheep

Figure 5.8: Boxplots of total cattle and sheep densities against the outcome. Widths of boxes represent the number of observations in each group. The y-axes is square root transformed.

5.3.4 Multivariate modeling

The results of multivariate analysis of the predictors entered as single terms are shown in Table 5.5. Cattle densities, IPs in Cumbria (the Carlisle DCC) and the length of the infectious period were highly significant ($p<0.001$). Further post-hoc analysis was carried out by merging the non-significant DCCs with the Rest of GB category. This resulted in the Newcastle DCC becoming non-significant. Furthermore there was a statistically significant difference between epidemic phase 1 and 2, but not between 1 and 3. Laboratory negative IPs remain highly significantly different from laboratory positive IPs, whilst those not tested were not significantly different in terms of the outcome.

Predictor	Unit	OR (95% CIs)	z value	p
Intercept		NA	-8.561	< 0.001
Infectious Period	log_{10}Days	41.40(13.96,122.78)	6.712	< 0.001
Phase	1	1	-	-
	2	0.71(0.55,0.91)	-2.733	0.006
	3	1.13(0.83,1.55)	0.769	0.442
DCC	Rest GB	1	-	-
	Carlisle	1.95(1.43,2.65)	4.223	< 0.001
	Newcastle	1.51(1.01,2.25)	2.030	0.042
	Ayr	1.37(0.86,2.18)	1.33	0.185
	Exeter	0.95 (0.64,1.42)	-0.227	0.820
	Leeds	1.50 (0.94,2.39)	1.711	0.087
dist to Seed	$\sqrt{km}$	0.96 (0.88,1.04)	-1.076	0.282
Lab Result	Positive	1	-	-
	No result	0.77 (0.58,1.02)	-1.831	0.067
	Negative	0.57 (0.43,0.74)	-4.074	< 0.001
Species	Mixed	1	-	-
	Cattle	0.94 (0.71,1.24)	-0.432	0.665
	Sheep	1.00 (0.75,1.33)	0.005	0.996
Satellites	N	1	-	-
	Y	0.95 (0.78,1.16)	-0.494	0.621
Susceptible density	$holdings/km^2$	1.12 (0.80,1.58)	0.673	0.500
Cattle density	$head/km^2$	1.18 (1.11,1.24)	5.642	< 0.001
Sheep density	$head/km^2$	1.01 (0.99,1.03)	1.213	0.225

Table 5.5: The main effects model derived using logistic regression analysis.

Biologically plausible interactions between the variables in Table 5.5 were tested and the resulting model is shown by Table 5.6. The area under the ROC curve for the main effects model is 0.692 and the AIC is 2478.8, whilst the values for the model with

interactions are 0.705 and 2457 respectively suggesting that the interaction model is a better fit.

Predictor	Unit	OR (95% CIs)	z value	P	Fig
Intercept		NA	-9.252	<0.001	
Infectious Period	log_{10}Days	44.77(14.94,134.2)	6.788	< 0.001	1
Phase	1	1	-	-	
	2	0.47(0.34,0.65)	-4.499	< 0.001	2
	3	0.95(0.67,1.33)	-0.325	0.745	3
Cumbria	No	1	0	0	
	Yes	1.16(0.84,1.59)	0.910	0.363	4
Species	Mixed	1	0	0	
	Cattle	0.57(0.19,1.68)	-0.611	0.541	5
	Sheep	5.30(2.37,1.19)	4.061	< 0.001	6
Lab Result	Positive	1	0	0	
	Unknown	0.82(0.43,1.56)	-0.611	0.541	7
	Negative	1.18(0.62,2.26)	0.505	0.614	8
Cattle density	$head/km^2$	1.23(1.17,1.30)	8.320	< 0.001	9
Sheep density	$head/km^2$	1.03(1.01,1.05)	2.776	0.006	10
Phase × Cumbria	1 × Cumbria	1	0	0	
	2×Cumbria	2.20(1.37,3.53)	3.252	0.001	11
	3×Cumbria	1.49(0.91,2.44)	1.565	0.118	12
Cumbria × Species	Cumbria × Mixed	1	0	0	
	Cumbria × Cattle	1.85(1.06,3.21)	2.164	0.030	13
	Cumbria × Sheep	0.61(0.34,1.09)	-1.658	0.097	14
Species× Cattle density	Mixed × $head/km^2$	1	0	0	
	Cattle × $head/km^2$	1.02(0.91,1.14)	0.352	0.725	15
	Sheep × $head/km^2$	0.83(0.75,0.91)	-3.848	< 0.001	16
Lab Result× Sheep density	Positive × $head/km^2$	1	0	0	
	Unknown × $head/km^2$	0.99(0.95,1.03)	-0.426	0.670	17
	Negative × $head/km^2$	0.95(0.91,0.99)	-2.610	0.009	18

Table 5.6: Full multivariate logistic regression model of holding level FMD transmission. The Fig column corresponds to the axis labels in Figures 5.9, 5.10, 5.11.

There were no significant differences between phases 1 and 3, however the effect variable was maintained as a three level factor as phases 1 and 3 are not coincident. Merging the untested and positive levels of the laboratory result category resulted in an increase in AIC, therefore laboratory result was maintained as a three level factor.

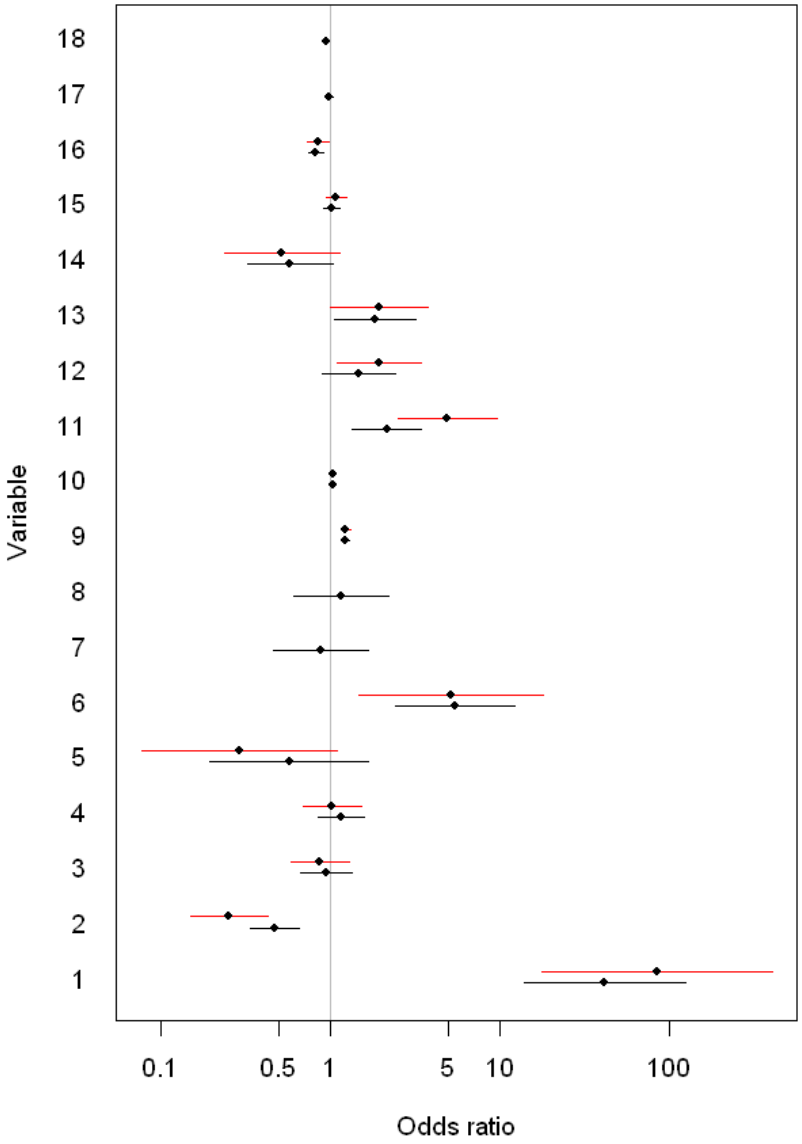


Figure 5.9: Plots of the odds ratios and 95% CIs for the predictors described in Table 5.6 (black lines) and for the model with just laboratory positive holdings (red lines). Variable numbers on the y-axis correspond to numbers in the Fig column in Table 5.7. Terms relating to laboratory result (variables 7,8,17 and 18) have no red lines because they were not in the model. The x-axis is log_{10} transformed.

There were four interactions identified in the model in Table 5.6:

1. Phase 2×Cumbria (OR=2.20): There was lower transmission during phase 2, however, during phase 2 transmission underwent less of a decline in Cumbria compared to to other DCCs.

2. Cumbria×Species (Cattle) (OR=1.85): Cattle holdings are at elevated risk of transmitting infection in Cumbria relative to other DCCs.

3. Species (sheep)×Cattle density (OR=0.83): Infectors are in areas of higher cattle numbers than non-infectors when there are cattle on the farm. However, infectors and non-infectors with sheep are in areas of similar cattle numbers.

4. Laboratory result (negative)×sheep density (OR=0.95): When the result is either positive or not tested, infectors tend to be in areas of higher sheep numbers than non-infectors. Laboratory negative holdings however, are in areas of equally low sheep numbers for both infectors and non-infectors.

Further analysis was conducted by rerunning the model without the laboratory negative predictor and analysing only the subset of the IPs which were not laboratory negative. This model had an area under the ROC curve of 0.683 and the risk factors in this model were not statistically different from the full model. This is shown in Figure 5.9 where there is no major shift in any odds ratio as a result of the exclusion of non-positive IPs, however phase 2 has a greater protective effect.

The source selection was checked by rerunning the model with only those cases and controls whose designations agreed with designations made using DCS and VLA the tracing data, this data comprised 324 infectors and 961 non-infectors. The results of these analyses are presented in Table 5.7, the odds ratios from these analyses overlap the odds ratios from the full model (Figure 5.10) showing that there is no significant difference between the predictors from the two models.

The use of 3km rings for calculating animal densities compared to the Voronoi polygons (section 5.2.2) resulted in no significant change in the odds ratios and the widths of the credible intervals for the predictors remained the same. This shows that the model is insensitive to the area used to calculate animal numbers.

Spatial autocorrelations were tested by comparing the full model in Table 5.6 against a model with the presence of IPs within 3km entered as the first term (section 5.2.3). The model with the spatial term was not significantly different either in terms of the widths of the credible intervals, also there was no shift in the odds ratios (Figure 5.11).

Predictor	Unit	OR (95% CIs)	z value	P	Fig
Intercept		NA	−7.627	<0.001	
Infectious Period	log_{10}Days	22.33(5.66,88.10)	4.436	< 0.001	1
Phase	1	1	-	-	
	2	0.78(0.52,1.16)	−1.245	0.213	2
	3	1.39(0.91,2.11)	1.519	0.129	3
Cumbria	No	1	0	0	
	Yes	1.34(0.90,1.98)	1.442	0.149	4
Species	Mixed	1	0	0	
	Cattle	0.63(0.17,2.40)	−0.678	0.498	5
	Sheep	3.82(1.34,10.9)	2.507	0.012	6
Lab Result	Positive	1	0	0	
	Unknown	1.15(0.52,2.53)	0.347	0.729	7
	Negative	1.36(0.61,3.03)	0.747	0.455	8
Cattle density	$head/km^2$	1.21(1.15,1.29)	6.489	< 0.001	9
Sheep density	$head/km^2$	1.05(1.02,1.07)	3.891	0.006	10
Phase × Cumbria	1 × Cumbria	1	0	0	
	2 × Cumbria	1.50(0.85,2.66)	1.396	0.163	11
	3 × Cumbria	0.85(0.46,1.58)	−0.508	0.612	12
Cumbria × Species	Cumbria × Mixed	1	0	0	
	Cumbria × Cattle	1.34(0.67,2.66)	0.822	0.411	13
	Cumbria × Sheep	0.94(0.45,1.95)	−0.166	0.868	14
Species × Cattle density	Mixed × $head/km^2$	1	0	0	
	Cattle × $head/km^2$	1.03(0.89,1.18)	0.367	0.714	15
	Sheep × $head/km^2$	0.82(0.72,0.93)	−3.182	0.001	16
Lab Result × Sheep density	Positive × $head/km^2$	1	0	0	
	Unknown × $head/km^2$	0.98(0.93,1.03)	−0.801	0.423	17
	Negative × $head/km^2$	0.94(0.90,0.99)	−2.516	0.012	18

Table 5.7: Full multivariate logistic regression model using only infectors and non-infectors where the tracing data agrees with the nearest neighbour identification. The Fig column corresponds to the values in the y-axis of Figure 5.10.

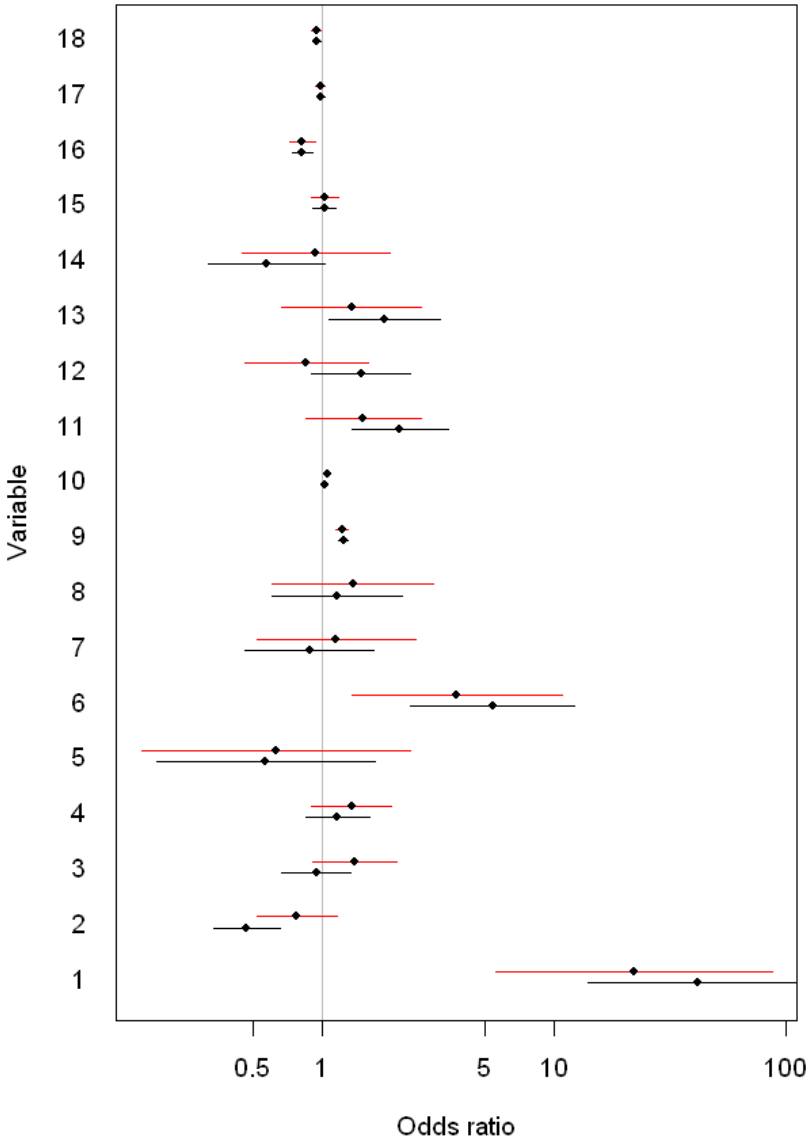


Figure 5.10: Plots of odds ratios and 95% CIs for the predictors described in the model parameterised with tracing data (Table 5.7, red lines) against the full model (Table 5.6, black lines). Variable numbers on the y-axis correspond to numbers in the Fig column in Table 5.7. The x-axis is log_{10} transformed.

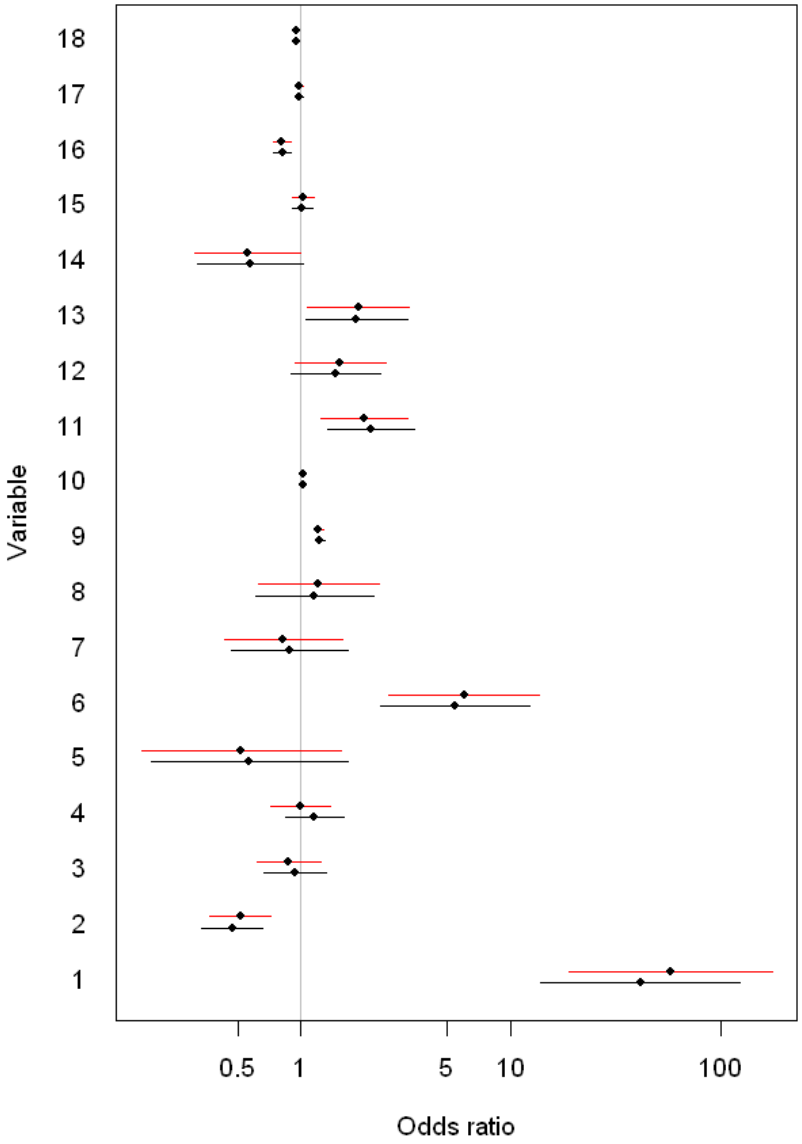


Figure 5.11: Plots of odds ratios and 95% CIs for the predictors described in Table 5.6 (black lines) and a model with the spatial dependency term added (red lines). Variable numbers on the y-axis correspond to numbers in the Fig column in Table 5.6. The x-axis is log_{10} transformed.

5.4 Discussion

A novel statistical model of FMD transmission has been constructed. The model does not describe true transmission because the true pattern of transmission is unknown, and more importantly the farms which did not infect others are not known with certainty. Instead of using incomplete tracing data the source of infection for each IP was taken to be the nearest possible source IP. As a result it is a model of finding next generation IPs in the vicinity of an IP rather than modeling transmission explicitly. The distance between source and potential daughter was capped at 3km to ensure that only local spread was analysed. Some cap was required because the mechanisms behind local and longer range spread are likely to be very different. Long range transmission is likely to be determined by the efficiency of biosecurity measures, as virus transport over distances greatest than 3km in the absence of effective wind-borne carriage must involve some human element, either on personnel, on vehicles, through movements of animals (legal or illegal) or via fomites (National Audit Office, 2002). Many such mechanisms of spread can be stopped with stringent biosecurity, through the disinfection of people and machinery and minimising non-essential services coming onto farm holdings. The mechanisms of local spread are not so clear; it may be the result of various mechanisms which are harder to control through biosecurity such as third party vehicles passing between holdings as well as short range wind-borne ('over the fence') spread, wind-borne fomites and escaped animals (National Audit Office, 2002). Therefore as local spread is the more difficult to control it is this which should be understood in order to manage future FMD epidemics.

Using this methodology 1255 IPs were assigned to 875 sources, the remaining 693 IPs did not have a candidate source within 3km. Of the 454 IPs for which there was both a source as defined in this chapter and a source identified by tracing within 3km, the sources agreed in 192 instances (38.7%, Table 5.2), but the assigned outcome agreed in 357 (78%) instances (Table 5.3). This is the statistic of importance as this is being tested in this analysis. The risk factors of the logistic regression model parameterised with these data were not significantly different from the full model

(Figure 5.10). Other methods have been used to reconstruct epidemic trees, the gold standard is that of sequencing virus samples (Cottam et al., 2006). However, this technique has only been applied to 22 IPs. Therefore it was necessary to derive a new way of reconstructing the pattern of transmission.

Following a methodology involving univariate analysis and variable selection followed by the building of a main effects model and the incorporation of interaction terms a multivariate model with 7 main effects and four interaction terms was developed. The final model had an area under the ROC curve of greater than 0.7, showing that this is an acceptable model fit (Hosmer and Lemeshow, 2000). Both in univariate and multivariate modeling, IPs with cattle were a greater risk factor than sheep only IPs. This was in terms of presence on the IP for which IPs with sheep were significantly different from both cattle only and mixed IPs. Sheep presence was a risk factor in the final multivariate model after the effect of sheep presence in relation to cattle densities was accounted for as an interaction term.

Cattle density was also significantly more important in determining transmission than sheep density, although both were significant (at the 99.9% level for cattle and 95% for sheep). This is likely to reflect differences in susceptibility because the number of animals in the locality are the animals at risk. In spite of the significance of animal densities, the number of susceptible premises is not a significant risk factor in multivariate analysis. The non-significance of the number of susceptible premises further underlines the importance of the numbers of susceptible hosts, in particular cattle rather than the number of at-risk premises (the true susceptible population for the purposes of these analyses) in determining FMD spread.

The infectious period was a highly significant predictor, however this was a confounder variable because the process for selecting daughter IPs was dependent upon the infectious period of the IP. Although it is unclear whether infectiousness changes over time (Savill et al., 2007a) the daily virus production (and therefore its potential to transmit infection) of a farm which is infectious for a long period may be less than that of a farm which is infectious for shorter period and the total virus production over the period may be similar. This is because less infectious farms are likely to have fewer animals infected and may therefore take longer to detect disease on these

farms.

Epidemic phase was a significant univariate predictor, both phases 2 and 3 were significantly different from phase 1 with a protective effect. However, in multivariate analyses phase 3 was not significantly different from phase 1, instead phase 2 is associated with a protective effect over phase 1 and 3. The significant protective effect reflects the epidemic curve falling during phase 2 whilst the epidemic was in decline and the case reproduction ratio (R) was less than 1 whilst phase 1 represents a period in which R was greater than 1 and phase 3 a period in which R fluctuated either side of 1. The significance of the protective effect of phase 2 may be associated with culling being undertaken during that period, although no firm conclusions can be drawn from these analyses. Another contributing factor may be the geography of the 2001 epidemic. Phase 1 represents the beginning and expansion of the North Cumbria, Devon and Dumfries and Galloway clusters. Phase 2 represents the decline in the Devon and Dumfries and Galloway clusters and the slowing down of the Cumbria cluster as it spread into Southern Cumbria and phase 3 the slower "rumbling" epidemics in South Cumbria, Settle and North East England. The slowing down rather than dying out of the Cumbria epidemic as it moved south explains the interaction between phase 2 and Cumbria.

Laboratory test-negative holdings were less likely to transmit infection, and holdings which were not tested in the laboratory produced similar results to positive farms in multivariate modeling, but were at significantly lower risk in univariate analysis. Further validation of the model by analysing only those holdings which were not laboratory test-negative produced a set of predictor variables which were not significantly different from the full model. This further validates the outcome selection processes by demonstrating that those holdings which could not transmit infection because the virus was not present on the farm were at significantly lower risk of transmitting infection than those which could.

The reason for Cumbria being a risk factor is less clear. The 'being in Cumbria' variable is a proxy for some biological or epidemiological process and could be a proxy for many factors. The interactions with cattle infections shows that Cumbria was more of an epidemic of cattle rather than sheep when compared to the rest of

GB. The 'Cumbria effect' is explained by its interaction with phase 2 of the epidemic when as noted above the epidemic continued in Cumbria whilst it declined in other regions. However after taking account of the interactions with phase 2, the main effect of being in Cumbria remains a risk factor and the reasons for this are unclear.

A novel approach has been developed to overcome the problem of uncertainties in the transmission structure of this epidemic to develop a model of the likelihood of infected farms to transmit infections to other farms in their locality. Distance between holdings is the principal risk factor for FMD transmission (Chapter 4) and is accounted for implicitly in the outcome by assigning the nearest neighbour as the source. The model is robust to potential sensitivities, and the fit is reasonable, but not excellent due to stochasticity in virus transmission and factors which could not be modelled in this analysis such as contacts with other farm holdings. The model further underlines the importance of cattle in this epidemic as well as the unusual nature of the epidemic in Cumbria which shows a very different transmission pattern compared to the rest of the country.

Chapter 6

Geographic and topographic determinants of local FMD transmission

6.1 Introduction

Chapter 5 described the development of a model of the local and IP specific risk factors for disease transmission whilst Chapter 4 described the development of a model of local and farm specific risk factors for FMD infection. These are models which describe the likelihood of an individual farm transmitting or contracting infection but do not incorporate dynamic processes to describe the likelihood of infection between an infectious and susceptible premises pairs. This process is spatially dependant and has been simplified to a transmission kernel for the purposes of mathematical modelling (Keeling et al., 2001, 2003; Tildesley et al., 2006, Figure 6.1) and is based on the tracing data described in Chapter 2. The kernel describes gradually declining infection risk with distance to an infectious source.

The transmission kernel (Figure 6.1) assumes that the farming landscape can be modelled as a homogenous surface and features of the landscape do not influence the pattern of transmission. However it is possible that potential transmission routes are influenced by geographic features such as roads which may act as conduits for transmission, whilst rivers, railways and in a different way roads may act as barriers

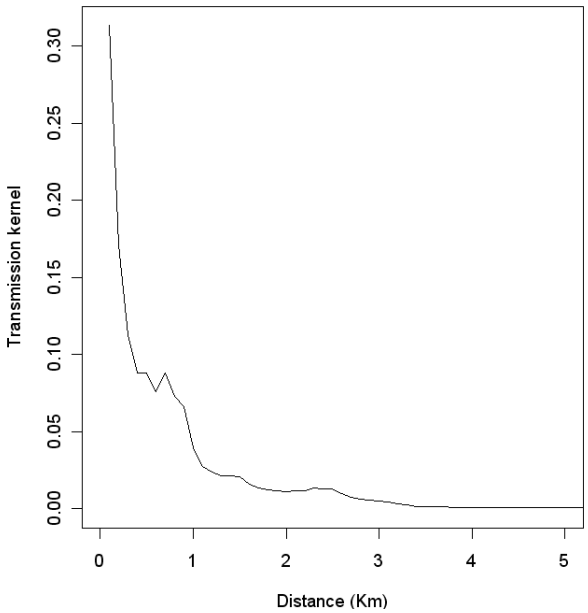
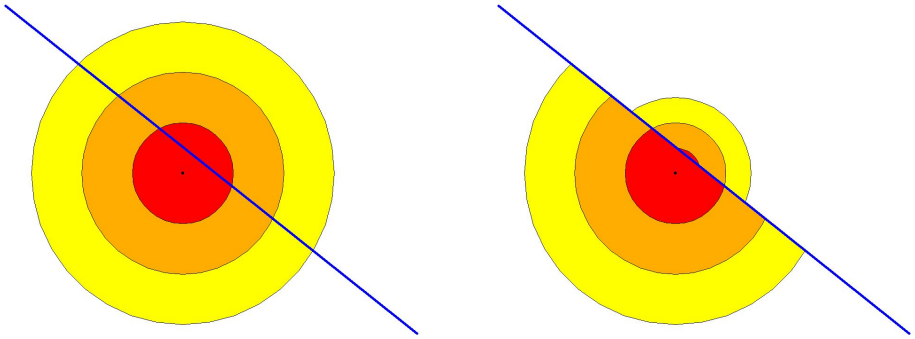


Figure 6.1: Transmission kernel showing the likelihood of a holding becoming infected against the distance from an infectious source. Kernel derived from Tildesley et al. (2006).

to transmission (Figure 6.2). The importance of geographical features as barriers to epidemic progress has been previously demonstrated for a rabies epidemic in which large rivers act as "semi-permeable barriers" to transmission causing a 7-fold reduction in the rate of spread of the epidemic (Smith et al., 2002). Understanding and quantifying the roles of geographic features in the spread of FMD would add vital information to our understanding of how FMD is transmitted between farms.



(a) Homogenous landscape (b) Influence of barriers

Figure 6.2: Diagram illustrating spatial dependency of transmission when a linear feature does not influence the pattern of transmission (Plot 6.2(a)) and when the feature reduces transmission by 50% (Plot 6.2(b)). The central point is the IP and coloured bands represent bands of equal transmission risk.

The roles of geographic features in the transmission of FMD have been partly investigated by Savill et al. (2006) who compared road distance against Euclidean distance as a measure of virus transmission. The authors did not use the tracing data and calculated the road and Euclidean distance for each IP to every other IP within a 10km radius and from each IP to every other uninfected susceptible holding within 10km. Comparing the distributions of road to Euclidean distances for the IP-IP pairs against the IP-susceptible pairs they find no statistically significant difference between the two groups. This is with the exception of instances in which the road distance is inflated by the presence of a river estuary, where the barrier increases the road distance between holdings. The analysis of Savill et al. (2006) suggests that road distance is less valuable as a predictor of transmission than Euclidean distance but does suggest that major river estuaries can influence virus transmission. However this analysis was undertaken at a coarse scale in terms on the numbers of pairs of farms which were included and the analysis does not make use of data on known transmissions.

This study will carry out a finer scale study of barriers to transmission. A case-control methodology has been developed which selects an IP (*source*) which has infected a daughter IP (*case*) and has not infected one or more *control* farms to form groups of farms. A single *source-case* pairing and a single *control* form a triplet. The controls are matched to the *source-case* pairing such that the *controls* are a similar distance to the *source* as the *case* and have a similar composition in terms of animal numbers. Controlling for distance and numbers of animals is necessary to ensure that susceptibility to FMD infection is similar on the *control* and the *case*. Intervening geographic features between the *source* and *case*, and *source* and *control* can then be evaluated. Therefore the aims of the chapter are:

1. To identify a workable set of *cases* and *controls* which have been matched on distance to an infectious source and farm composition.

2. To compare animal numbers on *sources, cases* and *controls* when distance is controlled for.

3. To identify potential barriers or conduits to transmission and measure these

between the farms in the group.

4. To statistically analyse the relative importance and the direction of influence of the barriers and conduits to transmission

6.2 Methodology

The data for this analysis consisted of groups of farms comprising an IP (the *source*) which infected a second IP (*case*). The *case* is matched to one or more *control* farms which share some similar characteristics to the *case*. The methodology must take account of the spatial dependency of FMD transmission during the 2001 outbreak and ensure that *controls* were selected which were a similar distance to the *source* as the *case* and therefore were equally as exposed to infection as the *case*. To ensure analysis of local spread the distance between *source* and *case* will be capped at 3km.

6.2.1 Case Selection

Transmission dynamics before and after the NMB were different (Chapter 2), therefore the 78 IPs which were infected before the NMB will not be included as animal movements mean there are a greater range of infection routes between *source* and *case*. Chapter 2 discussed the generation of reliable data on the source of infection for 823 IPs. However, only 554 of these IPs were infected over distances of 3km or less. The selection processes are summarized in Figure 6.3.

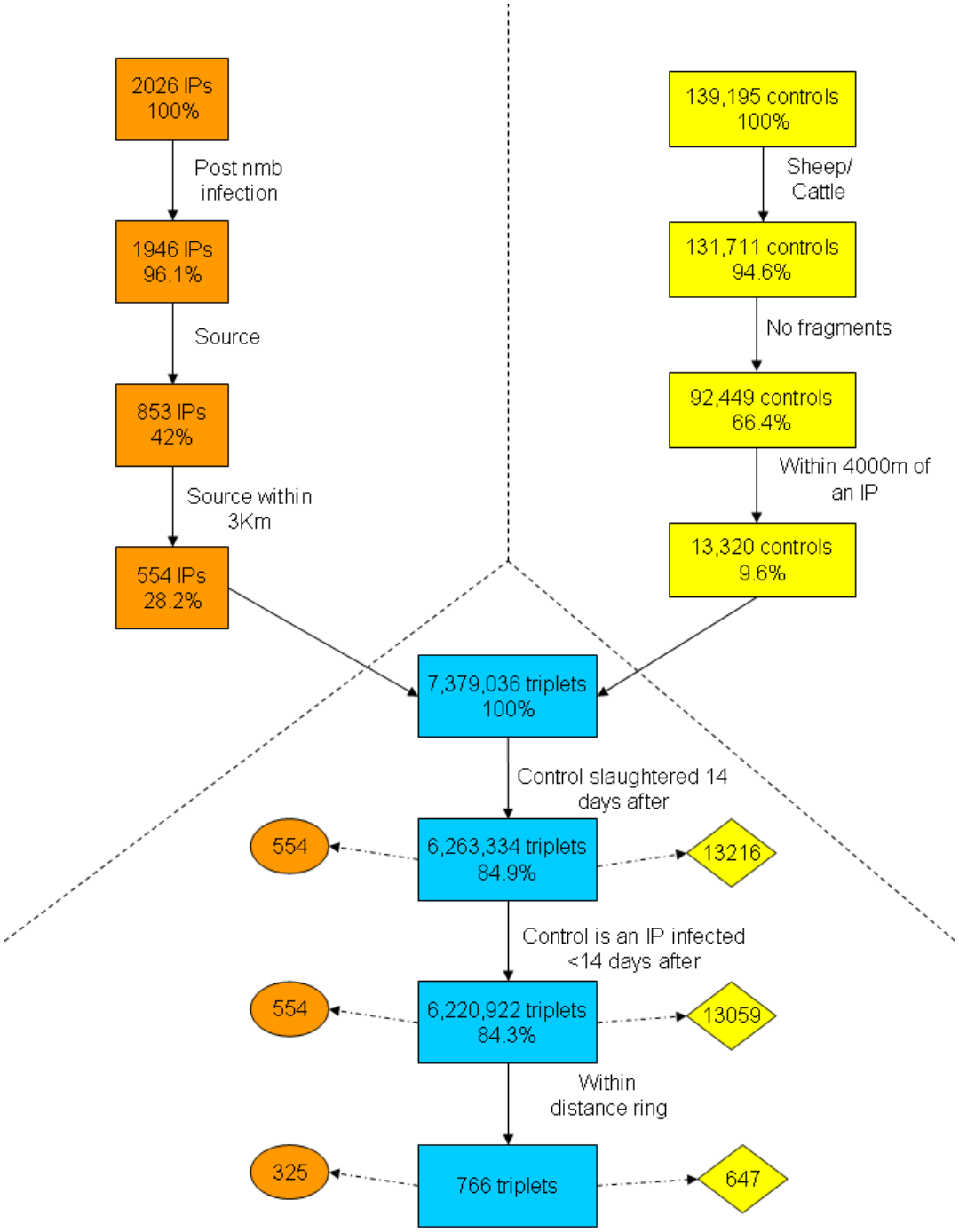


Figure 6.3: Flow diagram explaining the identification of groups. The red section refers to the selection of IPs, the yellow section to the identification of controls and the blue section to the identification of triplets. The orange diamonds and yellow ovals relate to the number of candidate IPs and controls remaining at each stage of group identification, the number of IPs in the orange ovals represents the number of groups.

6.2.2 Control selection

The controls were livestock holdings taken from the 139,195 georeferenced holdings on the June 2000 agricultural census (Chapter 3). As sheep and cattle were the main agents in virus transmission in the outbreak only *controls* which held either sheep and/or cattle were considered. Removing all pig-only holdings reduced the number of holdings to 131,711. Chapter 3 identified livestock farms with field fragments for which animal numbers could only be simulated and animal numbers on the main holding were inflated. To prevent bias being introduced by lack of accurate data on animal numbers on the main holding, these fragmented holdings were not included in the *control* set (Figure 6.3), leaving 92,449 *controls*. As the *controls* are distance matched, only those *controls* which were within 4km[1] of an IP were used. The final set of potential *controls* consisted of 13,320 holdings (Figure 6.3).

6.2.3 Identifying groups

If every one of the 554 *source-case* pairings could form a group with every one of the 13,320 *controls* there would be 7,379,036 *source-case-control* units, but only 554 groups as there are still just 554 *source-case* pairs. However, as some of the potential *controls* were culled as part of the disease control effort it is essential to ensure that the selected *controls* were culled after the *case* and they were culled a sufficient time after the *case* to ensure that there was no chance of the *control* harbouring a subclinical infection originating from the *source*. Experimental laboratory infections with the UK2001 FMD virus strain suggest that 14 days is the maximum estimate for the period from infection to clinical symptoms in cattle and 10 days is the equivalent value for sheep (Kitching, 2002; Kitching and Hughes, 2002). Therefore a rule was applied stipulating that if the *control* was culled it must have been slaughtered at least 14 days after the *source* was culled. Thus any incubating infection originating from the *source* to become clinical and be identified when the holding was visited for culling. This process reduced the number of triplets to 6.26 million, a 15.1% reduction. In some instances the *control* is a farm which became an IP later in the

[1] 4km allowed some margin for *control* selection either side of the 3km cut off for distance from *source* to *case*

epidemic. In these instances it is essential to ensure that there is no possibility that the IP was infected by the *source*. In such instances the estimated date of infection for the *control* must be at least 14 days after the *source* was culled. The 14 day window allows for difficulties in assessing the date of infection for IPs and thus minimises the likelihood that the *control* was infected by the *source*.

Spatial dependence in FMD transmission was controlled by matching for Euclidean distance. The distance from *source* to *control* (d_n) needed to be similar to that between *source* and *case* (d_c), and therefore controls were selected if d_n was within a certain distance (d_p) of d_c. However, simply taking $d_c \pm d_p$ will introduce bias as the area of $d_c + d_p$ will be greater than $d_c - d_p$ and therefore there are more *controls* for which $d_n > d_c$ (Figure 6.4). *Controls* will be matched to *cases* where:

$$\begin{aligned} d_c - d_p &< d_n \\ d_c + d_a &> d_n \end{aligned} \tag{6.1}$$

d_a is derived from the area (A) between the rings generated by d_c and d_p where A is given by:

$$A = \pi d_c^2 - \pi(d_c - d_p)^2 \tag{6.2}$$

so d_a is given by:

$$d_a = \sqrt{\frac{A + \pi d_c^2}{\pi}} - d_c \tag{6.3}$$

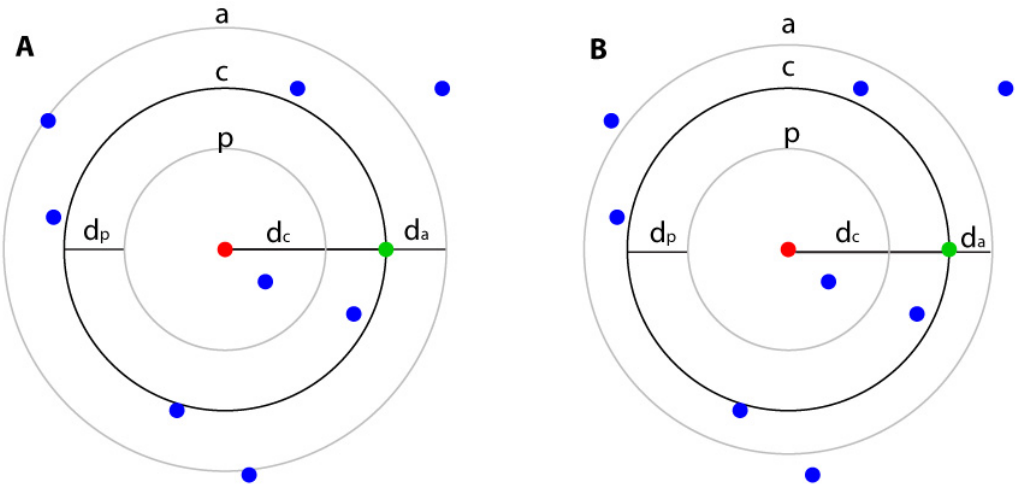


Figure 6.4: Diagram illustrating adjustment of *control* selection distances. The *case* (green dot) is d_cm from the *source* (red dot), this distance is shown as the solid ring (c). In the situation illustrated by A *cases* are searched for using percentage bounds of d_c the area outside (shown by the ring a) is much greater than that inside (shown by the ring p), therefore the probability of identifying a *control* (blue dots) whose distance from the *source* (d_n) is far greater when $d_n > d_c$ than when $d_n < d_c$ and therefore the *control* identification will be biased. To remove this selection bias the scenario illustrated in B is used where the outer ring is fixed at a radius such that its area defined by d_a is identical to that defined by d_p.

A range of possible values for d_p and d_a were considered by evaluating a range of values of d_p and as a fixed value of d_p of 250m (199.5m for d_a, Figure 6.5). From Figure 6.5 a value of 90% and d_p=250 and d_a = 199.5 were selected as this left a feasible number of triplets whilst maintaining a tight distance ring. Furthermore at 90% the ratio of numbers of *controls* to numbers of *cases* is around 3:2, much smaller than at some higher values of d_p. Both absolute values and proportions were used because at small values of d_c, d_a becomes very small and matching becomes unlikely.

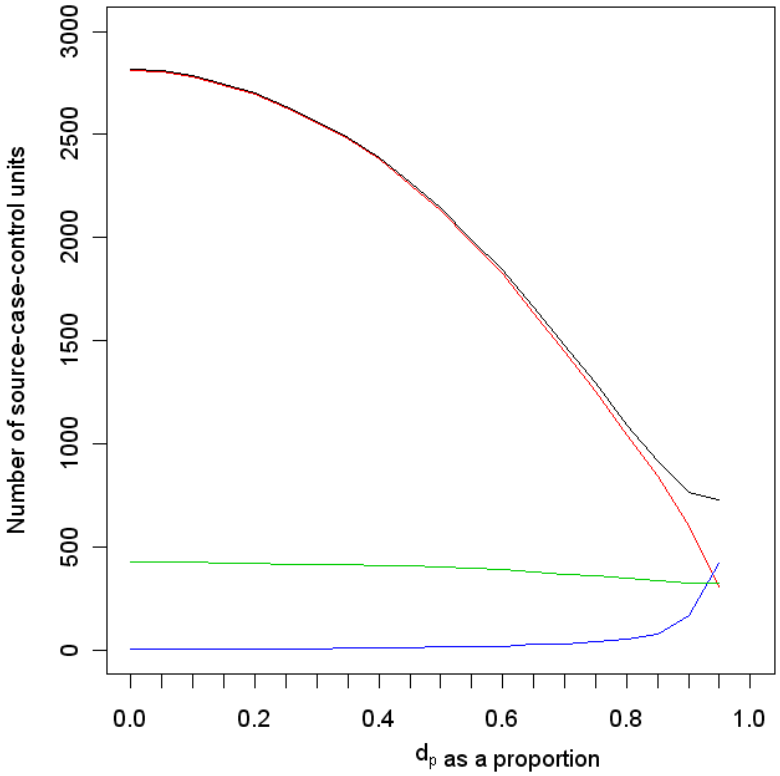


Figure 6.5: Plot of changes in numbers of source-case-control units as d_p changes and d_a changes accordingly to calculate a matching area. The red line is the number of triplets generated by calculating d_p as proportions of d_c (x-axis) and the blue line the equivalent values using exact values of d_p of 250m and 195.5m for d_a. The black line represents the total number of units and is the sum of the red and blue lines. The total number of *cases* is shown by the green line.

.

The triplet selection processes were implemented in Java^{TM}code.

6.2.4 Groups and potential biases

The program identified a total of 325 *case-source* pairings matched to 647 *controls*. Within these 325 groups between 1 and 11 *controls* were assigned to each pairing, and certain *controls* were assigned to several *cases* (Figure 6.6).

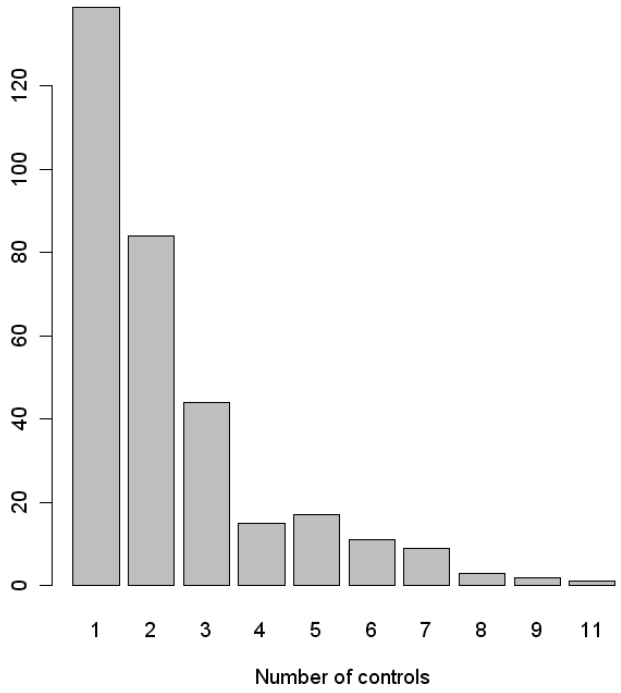


Figure 6.6: Frequency distribution of the number of *controls* for each of the 325 *cases*.

Groups were evaluated to look for bias in terms of differences in animal numbers between *case* and *control*. Animal numbers were available for all the *controls* from the agricultural census and numbers were available for all *cases* and *sources* from the DCS data. However there is not total agreement between cull and census data (Figure 6.7). Numbers of both cattle and sheep on *cases* and *controls* were compared using logistic regression; numbers were statistically significantly different both when census and cull data were compared (Table 6.1).

This analysis underlines previous analysis by other authors (Keeling et al., 2001, 2003) and analysis presented in Chapter 4 that susceptibility to infection is a function of animal numbers. Those holdings which did not become infected (*controls*) were smaller than those infected (*cases, sources*). Therefore in addition to controlling for Euclidean distance, it is essential to control for differences in susceptibility arising

			Cattle		Sheep	
Comparison	Data	n	χ	p	χ	p
Case	Census	251	98.59	<0.001	46.7	<0.001
Control	Census	647				
Case	Cull	325	116.8	<0.001	59.7	<0.001
Control	Census	647				
Case	Cull	325	1.61	0.18	20.07	<0.001
Source	Cull	325				

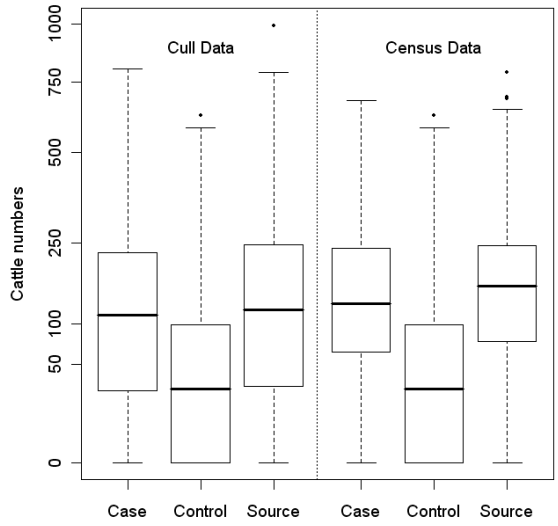
Table 6.1: Results of generalised linear model (binomial errors) analysis of animal numbers on group members.

from differences in animal numbers. To ensure that susceptibility is controlled for, the *controls* must have similar numbers or more animals of each species than the *case*. To control for this only *controls* will be selected which meet one of the following criteria:

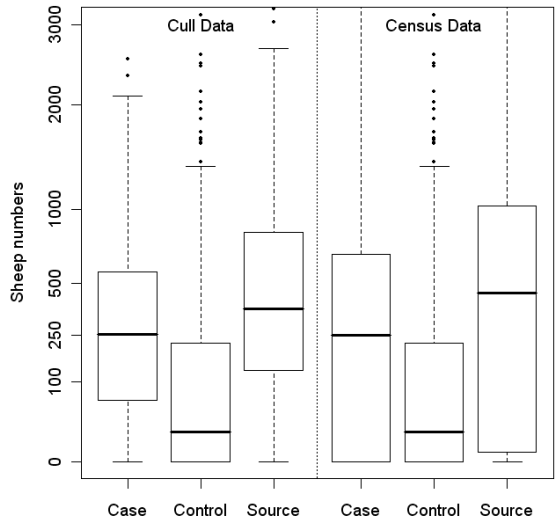
1. At least 70% of the numbers of cattle, sheep and pigs as the *case*.

2. No more than 50 fewer cattle than the *case* and no more than 100 fewer sheep and pigs than the *case*.

The latter criterion makes allowances for small holdings, for instance it allows a *case* with 50 head of cattle to be paired with a *control* with 30 head which would not be permitted under the first criterion and could introduce a size bias.



(a) Cattle



(b) Sheep

Figure 6.7: Boxplots of animals numbers on the group members broken down by data used and by cattle and sheep. Both plots are on a square root scale.

Revised dataset

After applying the animal number selection criterion the revised dataset comprised 123 groups with 208 *controls*, 8 of these *controls* were *controls* for 2 *cases*. The distribution of *controls* to *cases* is shown by Figure 6.8.

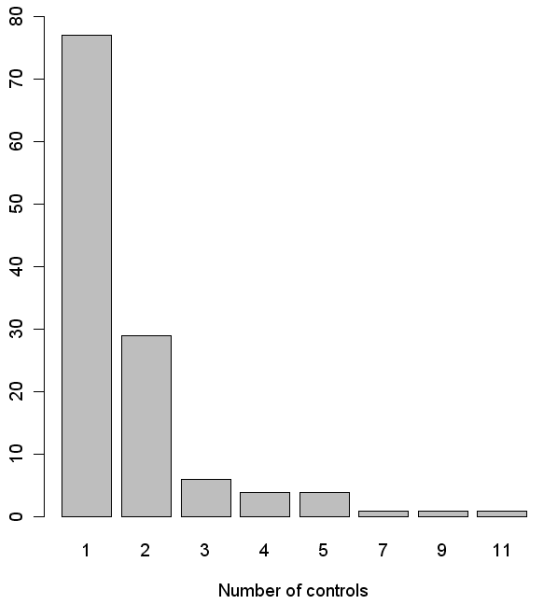


Figure 6.8: Frequency distribution of the number of *controls* for each of the 123 *cases*.

Figure 6.8 shows that the number of controls per case in the revised dataset has a greater degree of negative skew as there are many more *cases* with only one *control* compared to Figure 6.6. The difference between these distributions is significant at $p<0.05$ (Two-sample Kolmogorov-Smirnov test $D=0.198$, $p=0.002$). This set of triplets was further reduced by removing instances in which the *case* contained cattle and the *control* contained only sheep. Due to the difference in infectiousness and susceptibility between cattle and sheep (Donaldson et al., 2001) the reverse scenario is not considered to be a problem, however it could bias the analysis if there were instances of *cases* with up to 50 head of cattle being controlled by holdings with only sheep. Excluding such instances leaves 114 *cases* being controlled for by 196 *controls*. Furthermore there were 8 *controls* which were *controls* for more than one *case*, so in instances where a *control* was included in more than one group it was excluded from the group which had the most *controls*, when the groups had the same number of *controls* a randomly selected group lost its *control*. The resulting dataset had 113

cases and 189 *controls*.

The principal objective of the group selection process was to control for Euclidean distance. Analysis of the Euclidean distances between *source* and *case* and *source* and *control* shows that although *controls* are further away, there is no statistically significant difference (two sample, two tailed t-test t_{188}=1.51,p>0.1, Figure 6.9).

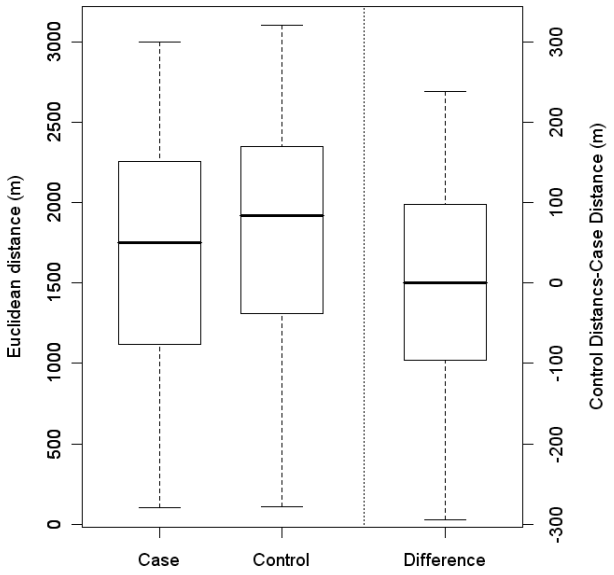


Figure 6.9: Boxplot of Euclidean distance to the *cases* compared to those to the *controls* and the difference between distances to *cases* and *controls*.

The spatial distribution of groups does not exactly match the spatial distribution of IPs during the 2001 epidemic. Comparative analysis shows that there are statistically significant differences in the numbers of groups relative to IPs for the DCCs (Figure 6.10, Fishers exact p=0.055). From Figure 6.10 it appears that it is the smaller DCCs of Leeds, Cardiff and Worcester which are overrepresented in terms of numbers of *cases*.

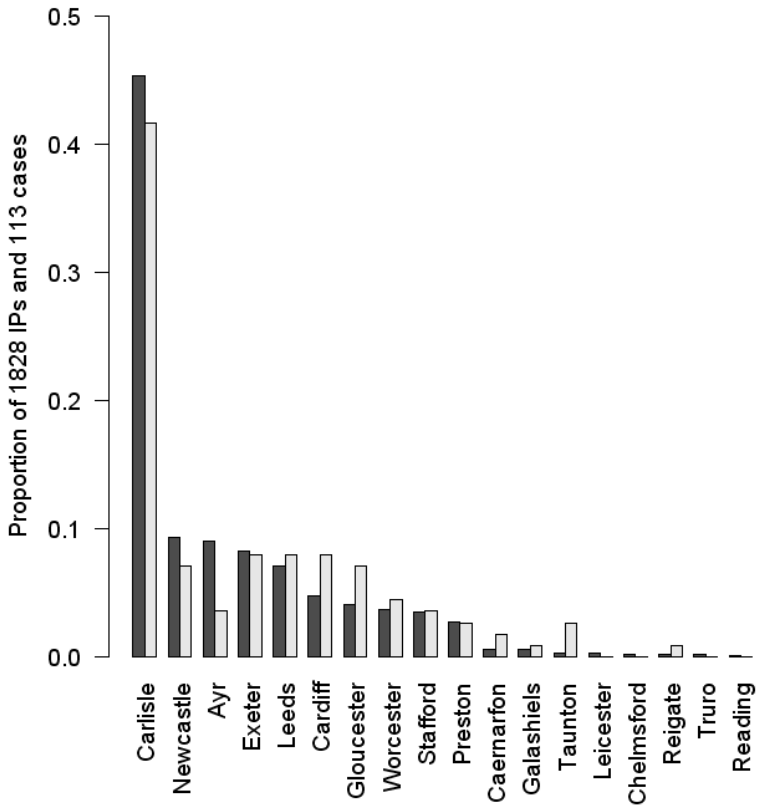
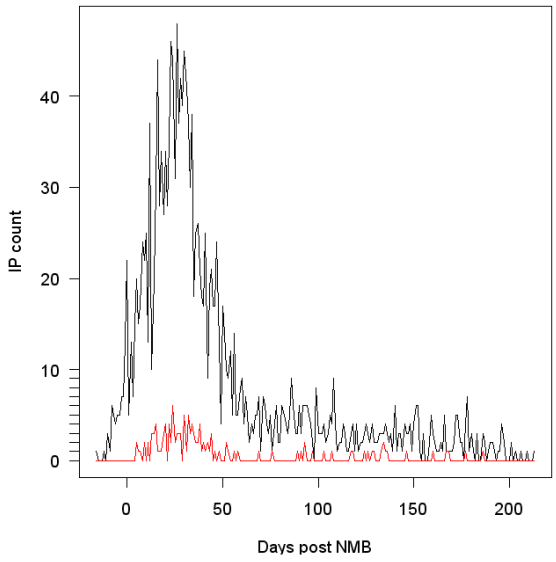
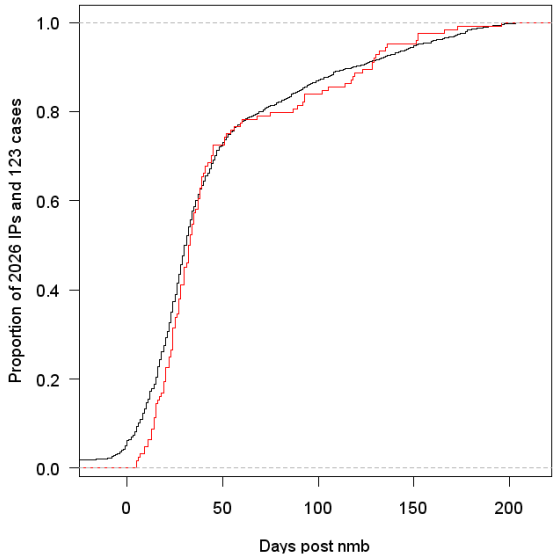


Figure 6.10: Barplot of the proportion of all IPs in each DCC (black bars) compared to the proportions of *cases* in this study in each DCC (grey bars). Bars are ordered in decreasing numbers of IPs in each DCC.

The method for selecting triplets produced a temporal distribution of *cases* which is not statistically significantly different from the temporal distribution of IPs in the actual epidemic (Figure 6.11, Two-sample Kolmogorov-Smirnov test D=0.109, p=0.126). There is a discrepancy at the beginning of the epidemic as a result of the NMB, because no triplets were selected during the pre-NMB phase of the epidemic. After this, however the curves match closely, as IPs infected before the NMB are not included this is not an issue for this analysis.



(a)



(b)

Figure 6.11: Temporal distribution of *cases* (red line) compared to the temporal distribution of all IPs (black line). This is broken down and shown as number of *cases* by day (Figure 6.11(a)) and cumulative proportion by day (Figure 6.11(b)).

In addition to accounting for distance these analyses took into account species numbers. This was to ensure that there were no species differences between *cases* and *controls* or that the *controls* were not substantially under-represented for any species. Farms were classified as cattle only, sheep only or mixed if both species were

held[2]. Figure 6.12 shows that slightly more *cases* were mixed farms however these differences are not statistically significant (χ^2_2=5.54, p=0.063).

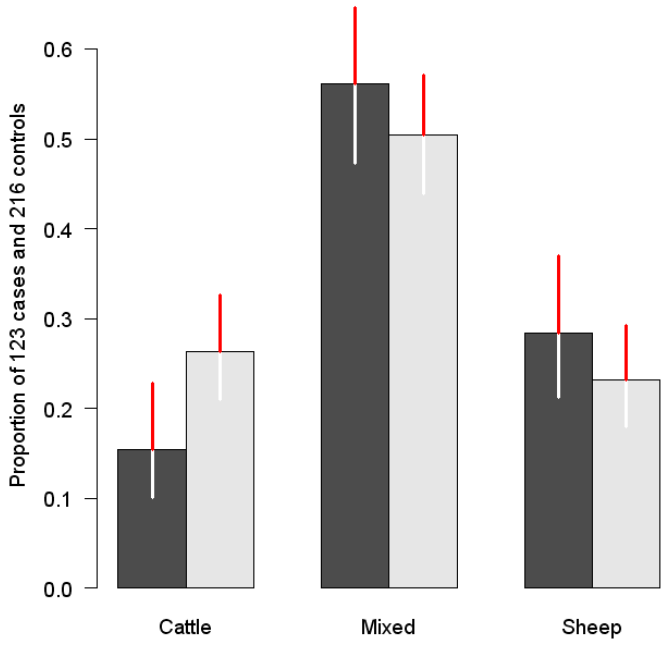


Figure 6.12: Comparison of farm type for *cases* (Black bars) and *controls* (grey bars). Red lines represent the upper limits of 95% confidence intervals and the white lines the lower limits.

Of the 113 *cases* identified in 30 of the instances the *case* tested negative for FMDV in the Institute of Animal Health Pirbright FMD reference laboratory (Chapter 2). In a further 17 instances the source tested negative but in 8 triplets both the *case* and *source* were negative, so in 39 groups either *source* or *case* tested negative to antibody or antigen. Sensitivity analysis will be carried out analysing just the laboratory positive IPs.

Having identified a set of matched *case-control* groups data on potential barriers and conduits to spread must be derived. This was undertaken in a GIS framework using data from the EDiNA Digimap service (www.edina.ac.uk/digimap).

[2]There were three instances (1 *case*, 2 *controls*) of holdings which were mixed cattle sheep and pig. There were no holdings which were pig only, therefore pig farms are not considered

Calculating predictor variables

The predictor variables analysed in this chapter can be divided into three groups described in Table 6.2. These three categories broadly describe the location of the holdings, factors which may reduce the effective distance between holdings (connectivity) and factors which may prevent the transmission of the virus between holdings (barriers).

Locational	**Connectivity**	**Barriers**
Access	Road distance	Rivers
	Number of roads	Railways
		Roads
		Elevation change
		Forest
		Urban areas
		Non-livestock land

Table 6.2: Variables used in these analyses broken into three categories.

All variables with the exception of cumulative elevation change were derived by overlaying the farm points onto the scanned OS maps and manually inspecting the *case-control* pairings. These analysis could have been performed using OS vector data in ESRI ArcView v3.2 (www.esri.com) by generating straight lines between *source* and *case* and *controls* and overlaying the lines on vector datasets. However, the vector data was found to be less accurate than the scanned raster maps. As a result the raster maps were used and the following criteria were ascertained:

1. **Road Distance between holdings.** This was calculated using the measure tool in ESRI ArcGIS 9.0. Where there were several possible routes between farms, all possibilities were measured and the shortest distance used[3].

2. **Farm accessibility.** Measured on the following scale:

   1 = Track or dead end road.

   2 = Beside a minor road

   3 = Beside a main road

   4 = At the junction of 2 main roads.

[3]This process was carried out manually using scanned maps rather than by using OS Meridian vector data in a GIS because upon inspection the Meridian data was found to be missing a number of small roads and tracks which limited the accuracy

3. **Number of intervening roads.** A count of the number of roads and footpaths crossed in a direct route between holdings.

4. **Intervening rivers.** A river is counted as a river if it appears on OS map data as more than a single line. Single line rivers or streams on OS maps denote rivers or streams that are less than 1 meter wide. Streams less than 1 meter wide are not considered to offer sufficient barrier to the movements of animals and therefore are insufficient barrier to transmission. The criteria for scoring rivers are:

   0 = No intervening river.

   1 = A partially intervening river. This is any case of a river separating two farms with a bridge for access or where the path between farms intersects the apex of a river meander.

   2 = Farms separated by a river with no direct crossing point.

5. **Intervening railways.** A similar scale is applied to scoring railways as is used for rivers:

   0 = No intervening railway.

   1 = A partially intervening railway. This is any case of a railway separating two farms with a bridge for access.

   2 = Farms separated by a railway with no direct crossing point.

6. **Intervening Forest.** A binary scale is used to denote the presence of forest.

   0 = No forest in a direct line between holdings.

   1 = Forest in a direct line between holdings.

7. **Intervening urban areas.** A binary scale is used to denote the presence of urban areas.

   0 = No urban in a direct line between holdings.

   1 = Urban in a direct line between holdings.

8. **Non-livestock holdings.** A 250m buffer was placed around farms identified in Chapter 3 as not having livestock and intersection with these buffers was evaluated using this scale:

0 = No non-livestock buffer in a direct line between holdings.

1 = Non-livestock buffer in a direct line between holdings.

The cumulative elevation change between source and case and source and control was calculated in the GIS using the OS Panorama 50m Digital Terrain Models (DTMs). In ESRI ArcView v3.2 lines connecting the source and case and source and control were constructed using the JoinTheDots Extension. The Surface tools extension (v1.6) (Jenness, 2005) was used to calculate the cumulative elevation change along the lines.

An example of a group and the intervening features is shown by Figure 6.13. This example also illustrates some of the problems which arose in evaluating the data for the groups and shows how the issues were resolved. This group is in Dumfries and Galloway, the *source* was infected on the 11th June and slaughtered on the 20th. The estimated date of infection for the *case* is the 20th June. The *case* is 1969m from the *source* and is a small sheep unit with 310 head of sheep. The road route from the *source* to the *case* is a relatively direct route. The *case* lies between a dead end track and a minor road, and it is unclear from which direction the farm is accessed. However, the farm was measured to be closer to the track than the road, so it was assigned an access score of 1 and the road distance was measured to the end of the dead end track. With this taken into account there was only one reasonable shortest route by road to the Farm. There are no intervening railways or rivers in a direct line between the *source* and *case* and this same line crosses a minor road, a track and a footpath, so the road score is 1 as no major roads are crossed and the road count is 3. One non-livestock holding (not shown in Figure 6.13) and no forest or shaded built up areas are crossed, although the latter is somewhat ambiguous as the shortest route passes close to some buildings.

The *control* is 1,789m from the *source* and is a large sheep and cattle holding containing 230 cattle and 1005 sheep. The road route between the *source* and *control* is less direct than that between *source* and *case* as the road must take a detour to a crossing point over the river. Across the river there are two possible routes to the *control* both of which were measured and the shortest was 3505m, 130m shorter than the alternative. The *control* lies on a dead end track, therefore the access score is

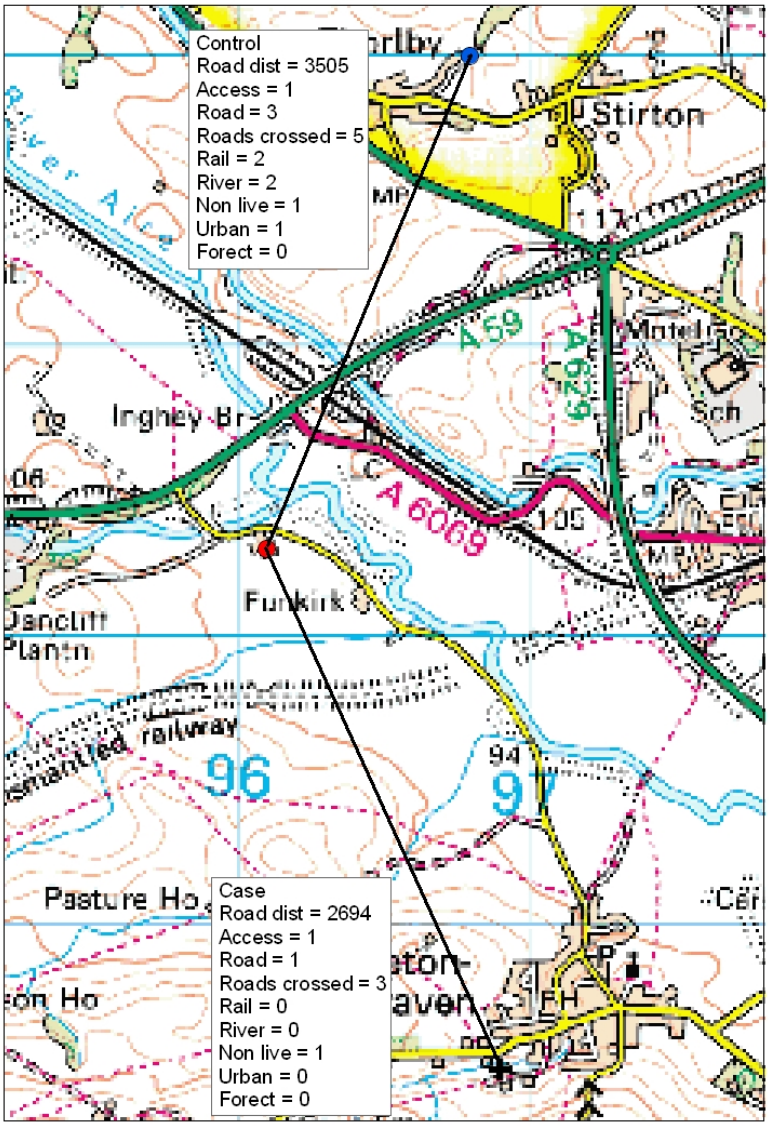


Figure 6.13: Map intervening features between the *source* (red point) and *case* (blue point) and *source* and *control* (Green point). For simplicity non-livestock holdings were excluded from this map.

1. Two rivers lie in between the *source* and *control* as well as a a railway and 3 major roads. A main road does cross the railway, however this still offers limited opportunity for close contact between animals, therefore the railway score was 2. There is no direct crossing for the first river, therefore this score is 2. There are no crossing points over the roads therefore the road score is 3. There are intervening non-livestock farms, and the direct route does intersect some of the shaded built up area, therefore these scores are both 1.

Data

Ordnance Survey (OS) data from the EDiNA®Digimap service were used for these analyses. OS 1:50,000 scanned raster maps tiles were downloaded for the relevant areas. In addition, 1:50,000 OS Land-Form Panorama Digital Terrain Model (DTM) data was downloaded for carrying out the topographic analyses. The DTM data were downloaded in NTF format and converted to ESRI Grid format in ESRI MapManager v6.2. Where appropriate, the Grid data were mosaiced in ESRI ArcMap 9 creating five separate Grids covering the relevant areas as follows as these areas cover the distribution of triplets:

1. North Cumbria and Dumfries and Galloway.
2. South Cumbria, the North East and North Yorkshire.
3. Anglesey.
4. The Midlands and Welsh borders.
5. Kent
6. Devon.

Data for each tile was compiled separately and subsequently merged to form one dataset.

6.2.5 Statistical Analysis

To undertake this analysis some method of handling the instances where there are multiple *controls* for a *case* is required. Introducing the same *case* (with the same barrier data) into the analysis several times would create duplicate values and potentially bias the analysis, but the additional *controls* will provide useful additional data for these analyses. Wherever possible the data will be analysed by comparison of *cases* and *controls* treating them as separate units of data, however this may not always be valid and there will be some replication by having multiple *controls* in a group.

Screening of continuous variable will be conducted using a Student's t-test. The problem of multiple controls is overcome by using the mean of the controls in each group.

Remaining analysis will be carried out using indirect comparison of the 113 *cases* and 188 *controls* using a logistic generalised linear mixed effects model (GLMM) with binomial errors. Whether a holding was a *case* or *control* will be treated as the binary dependent variable. To control for the effect of multiple *controls* the particular group to which the *case* and its *controls* belong will be included as a random effect. Therefore if a particular *case* has four *controls* the *case* and all its *controls* will be classified as group i. The test statistic used to evaluate risk factors within the R statistical package is the Wald statistic given by the t-value.

The variables will be analysed using both univariate and multivariate analysis. Where predictors are likely to co-vary, multivariate techniques will be employed to check for interactions where appropriate. The *control* category will form the reference level for the outcome in all logistic regression analysis.

Prior to model construction, analysis of the relationship of Euclidean distance to group size will be carried out using a GLM with Poisson errors as this could potentially confound other results.

6.3 Results

A dataset has been generated comprising 113 groups of farms with data regarding factors which may act as potential barriers or conduits for virus transmission.

6.3.1 Univariate Data Analysis

The methods used to analyse the data were discussed in Section 6.2.5.

Potential confounders

Figure 6.4 describes the increase in the search area for selecting potential *controls* as Euclidean distance from *source* to *case* increases. As a result there is a statistically significant relationship between the size of the group and the Euclidean distance

between *source* and *case* (estimate=0.242km^{-1}, p=0.038, generalised linear model with Poisson errors).

Locational

Initial analysis of the access scores assigned to *cases* and *controls* suggests that *cases* are on the whole less accessible than *controls* (Figure 6.14). However there is no statistically significant difference in the values for farm access (p>0.5).

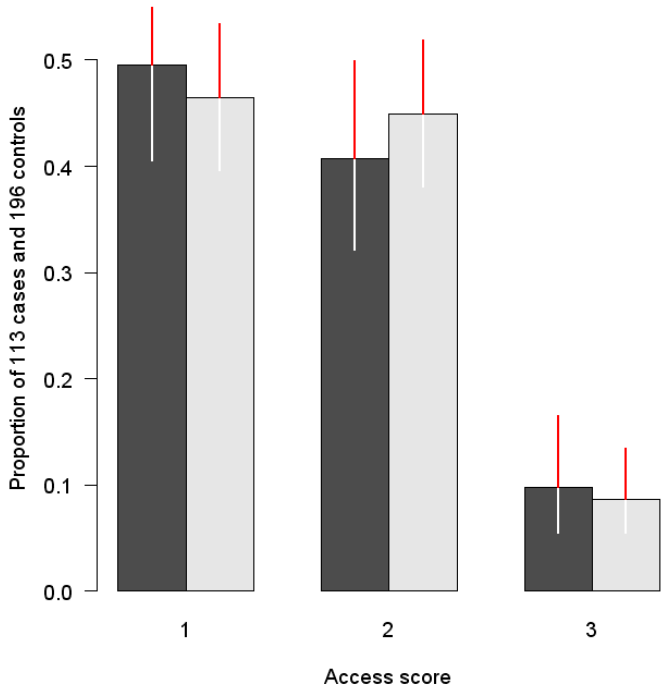


Figure 6.14: Access to farms as measured by the 1:3 scale (section 6.2.4, there were no values of 4). Dark grey denotes *cases*, light grey *controls*. White lines represent the lower half of 95% confidence intervals, red lines are the upper half.

Connectivity

Figure 6.15 shows that the median road distance is slightly greater to the *control* than the *case*. However this difference is statistically non-significant (t_{188}=1.67, p>0.1), furthermore this result remains when only laboratory positive triplets are analysed (p>0.1).

The number of roads separating holdings is not statistically significant (Figure 6.16, t_{188}=1.40, p>0.1). The statistical significance of these analyses did not

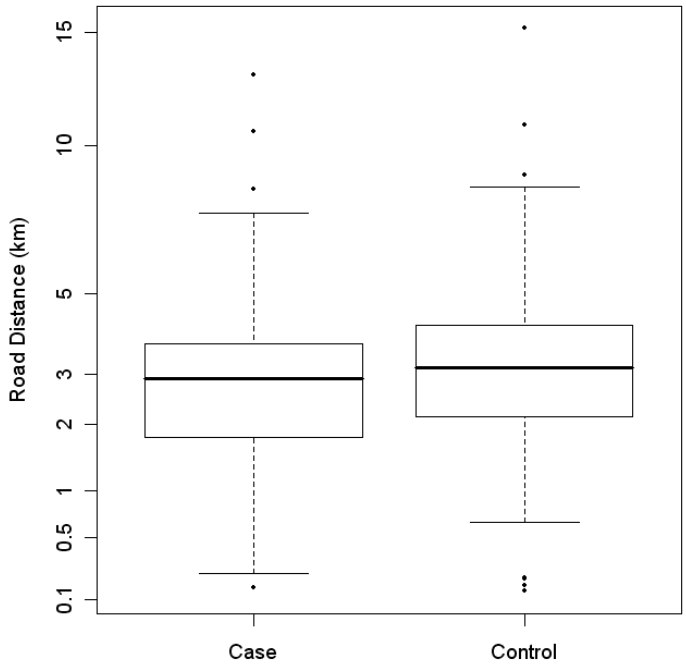


Figure 6.15: Boxplot of road distance between *source-case* pairings compared to road distance between *source-control* pairs.

change when the numbers of categories were changed by merging levels based on the sizes of roads. For instance by creating a 2 level variable of 2 or fewer roads and more than 2 roads.

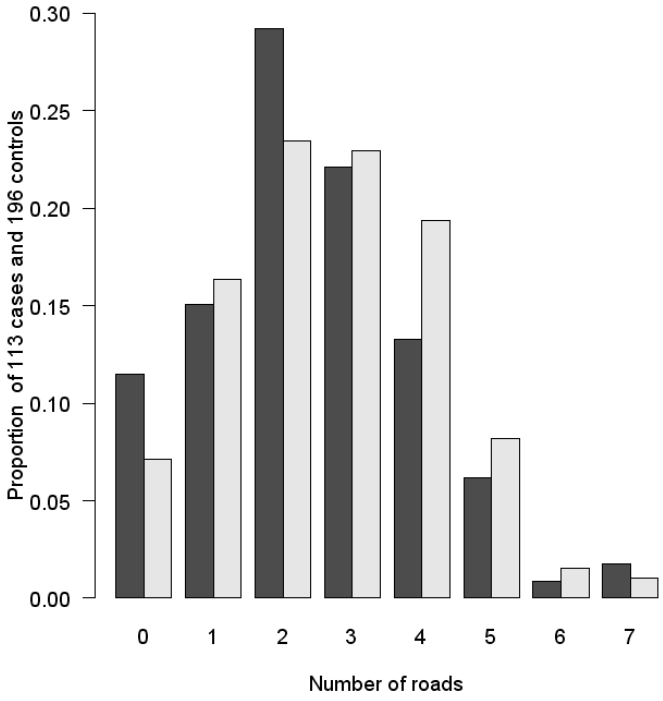


Figure 6.16: The number of roads separating *cases* (dark grey) and *controls* (light grey).

Barriers

Generalised linear mixed model analysis of the barrier variables presented in Table 6.2 is presented in Table 6.3. In the univariate analysis none of the results were statistically significant at $p<0.05$, however certain variables will be analysed in more detail.

Variable	Unit	OR (95% CIs)	t-value	p
Road	0	1	-	-
	1	0.959 (0.46, 2.02)	-1.109	0.913
	2 & 3	0.737 (0.33, 1.67)	-0.732	0.465
Rivers	Absent	1	-	-
	Present	0.594 (0.33, 1.08)	-1.72	0.087
Rail	Absent	1	-	-
	Present	0.550 (0.24, 1.28)	-1.389	0.166
Forest	Absent	1	-	-
	Present	0.990 (0.60, 1.62)	-0.041	0.967
Urban	Absent	1	-	-
	Present	0.778 (0.42, 1.45)	-0.791	0.430
Non-livestock	Absent	1	-	-
	Present	0.929 (0.57, 1.53)	-0.292	0.771
Elevation change	$\sqrt{m}$	0.965 (0.92, 1.01)	-1.448	0.149

Table 6.3: Univariate generalised linear mixed model analysis (binomial errors) analysis results of barrier variables. For rivers and railways categories 1 and 2 were merged to form the present category as there were less than 5 holdings in each of the '1' categories.

Further analysis of road barrier scores indicates very little difference between *cases* and *controls* (Figure 6.17). These categories can be further collapsed to 'major roads' (categories 2 and 3) and 'minor roads' (categories 0 and 1) but there is still no statistical significance when this variable is analysed ($p>0.2$). Groups were investigated further to establish whether the major road was multi-carriageway to investigate any possible effect but there were only five instances of holdings separated by a multi-carriageway road.

There were relatively few instances in which either *cases* or *controls* were separated by rivers or railways. Table 6.4 shows that fewer than 26% of holdings were separated by rivers and fewer than 15% by railways. However, Table 6.4 also shows that a larger proportion of *controls* are separated by rivers and railways than *cases*.

To overcome the lack of observations the river and rail variables were combined to produce a barrier variable defining whether a pair of holdings is separated by a

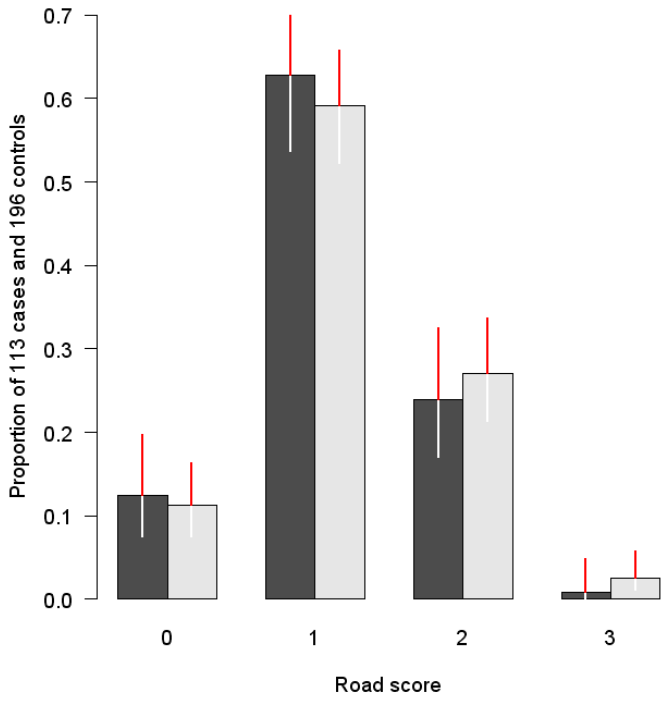


Figure 6.17: Road barriers as measured by the 0:3 scale (section 6.2.4). Dark grey denotes cases, light grey controls, red lines are 95% confidence intervals.

river and/or railway. These barriers would act in a similar way to prevent virus transmission by preventing the movements across them. Furthermore, railways are often built alongside rivers. Of the 31 holdings separated by a railway in Table 6.4, 15 were also separated by a river. This combination variable is statistically significant (t_{188}=2.38,p=0.018, OR=0.507, 95% CIs=0.29, 0.89, mixed effects logistic regression analysis). This result remains when only laboratory positive groups are considered (t_{124}=-2.12, p=0.036, OR=0.499 95% CIs=0.26,0.95, mixed effects logistic regression analysis).

		0	1	2
River	Case	94 (83.2)	3 (2.6)	16 (14.2)
	Control	141 (74.6)	3 (1.6)	45 (23.8)
Rail	Case	105 (92.9)	1 (0.9)	7 (6.2)
	Control	166 (87.8)	3 (1.6)	20 (10.6)
Rail & Rail	Case	91 (80.5)	NA	22 (19.5)
	Control	128 (67.7)	NA	61 (32.3)

Table 6.4: Counts of holdings separated by rivers and railways, numbers in brackets are the percentages based on the row total. The Rail & River category is the result of merging categories 1 and 2 for the features.

There is little difference in the elevation change between *source* and *case* and *source* and *control* (Table 6.3, Figure 6.18). This result is also not statistically significant (p=0.174) when only laboratory confirmed cases are analysed.

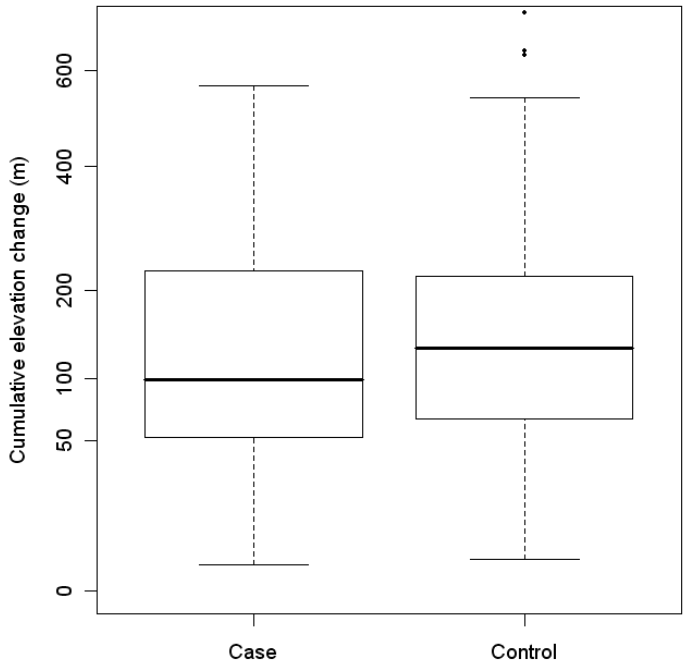


Figure 6.18: Cumulative elevation change between *source-case* pairings and *source-control* pairings. The y axis is on a square root scale.

6.3.2 Multivariate analysis

In light of the relationship between group size and Euclidean distance, the relationship between the barrier variable and Euclidean distance was tested and found to be a significant positive relationship. The odds of a barrier being present increase approximately 3 fold with every km (generalised linear mixed model with binomial errors odds= $3.096km^{-1}$, 95% CIs = 1.84, 5.20), p<0.001). As a result of the relationships between number of *controls* and the presence of barriers with Euclidean distance, the barrier predictor was analysed in a multivariate model with Euclidean distance to the holding the first term entered and whether the holding is a *case* or *control* as the outcome. The presence of an interaction between distance and barriers was checked but was not statistically significant (p>0.2) and was removed. After removal of this

interaction the presence of the barrier remained statistically significant (Table 6.5).

	Unit	odds (95% CIs)	t-value	p-value
Intercept		NA	-0.317	0.752
Dist	km	0.865 (0.61,1.22)	-0.824	0.411
Barriers	Absent	1	-	-
	Present	0.540 (0.30,0.96)	-2.087	0.038

Table 6.5: Multivariate generalised linear mixed model analysis (binomial errors) of the presence of barriers and Euclidean distance.

Further multivariate analysis was carried out between barriers and road distance and elevation change however there were no statistically significant relationships ($p>0.1$).

6.4 Discussion

The results presented in this paper suggest that rivers and railways formed semi-permeable geographic barriers to virus transmission during the 2001 GB epidemic. The results were generated by a case-control methodology generated by a complex series of matching for Euclidean distance and animal numbers.

The results presented are heavily dependent upon the process of identifying groups, particularly the process of identifying *controls*. Numbers of animals by species and Euclidean distance from the infectious source have previously been identified as the two principal risk factors for FMD transmission (Ferguson et al., 2001a,b; Keeling et al., 2001, 2003). Here a methodology was developed to identify triplets which controlled for the Euclidean distance only. Analysis of these groups underlined the importance of animal numbers in determining the susceptibility of holdings and underlined the results of previous chapters (Chapters 4, 5) and previous studies (Ferguson et al., 2001a,b; Keeling et al., 2001; Diggle, 2006). Therefore when researching inter-farm FMD transmission there must be accurate information on the demographics of the holdings in question.

In addition to the controls for animal numbers and Euclidean distance further constraints were placed on the triplet selection criteria. These are principally that the *case* had to be infected by local spread (ie. over a distance of 3km or less). However, this distance is arbitrary and as there can be no distinct cut off between

local and longer range spread, the mechanisms of spread may be distinct although there will be overlap between local and long range spread. Other studies (Savill et al., 2006) used a maximum distance of 10km, although the authors were not just examining local spread. Further constraints placed upon group identification are that the *controls* must be holdings listed on the agricultural census and those without off-fields identified in Chapter 3. Whilst this removes a number of potential groups from the analysis it is essential to ensure an accurate estimate of animal numbers on the holding, which are not available for off-fields. The numbers of animals on each holding relate to June of the previous year, and the numbers for the period in question will be those values for June with some error either side. However the section on demography (Chapter 3) demonstrated that, particularly for cattle this error is relatively small and there are no large variations in animal numbers between June 2000 and the FMD epidemic.

Having identified a set of *case-control* groups, predictor variable data for each group was gathered manually (with the exception of elevation change) using scanned paper maps as this was found to be more accurate than using less complete digital data. The data were analysed using linear mixed effects modelling. In univariate analysis the one factor which came out as statistically significant was that more *controls* were separated by rivers and/or railways than *cases*. This result remained when Euclidean distance was included as a confounder. This finding suggests that these linear features are acting as semi-permeable barriers to the transmission of FMD and that a holding on the far side of such a feature from an infectious source is less likely to become infected. This result is partially in agreement with the findings of Savill et al. (2006) however the authors conclude that the role of these barriers is to increase travel times and effective distance between holdings. Multivariate analysis in this study could find no interaction between barriers and road distance, although statistical power was limited. The analysis conducted by Savill et al. (2006) used all possible transmission events rather than transmission events inferred on the ground and therefore the actual effects may be masked. This study uses matched farms and transmission data gathered on the ground, therefore results are more likely to reflect what was involved in transmission between individual farms.

Therefore, rivers and railways act as a barrier to virus transmission, probably by preventing the crossing of people, animals and vehicles. However major roads were not statistically significant when included as a barrier, possibly suggesting that they are more permeable than rivers and railways.

Other potential influences on transmission of FMD were investigated, roads and topography do not appear to have an influence on the transmission of the virus by acting as conduits. It is possible the 3km scale is too small for differences in topography to be seen, however it may be more of a factor in longer range spread.

In conclusion, case-control pairs have been derived using a robust methodology and using OS data layers a dataset of potential barriers and conduits for FMD transmission has been derived. Results show that rivers and railways act as statistically significant barriers to FMD transmission. More data on sources of infection and more robust demographic data particularly with regard to numbers of animals on off-fields would be valuable to increase the number of groups and generate sufficient statistical power to analyse.

Chapter 7

Risk Mapping of FMD

7.1 Introduction

Retrospective analysis of FMD epidemics could be aided by maps of the distribution of cases in relation to risk. Such maps could be used to understand why a disease spread widely in some areas whilst the epidemic ended earlier in other areas. In the event of an epidemic of FMD both effective targeting of resources and evaluation of the control effort required could be facilitated by maps of the spatial distribution of risk of local FMD spread. Such maps could be used to forecast and understand the size of FMD outbreaks.

Risk mapping has been applied to infectious diseases using a variety of techniques to first generate a measure of risk and then map these values. For FMD Keeling et al. (2001) generated values of R_0 for each holding in GB and aggregated these values to create a 10km raster surface of R_0. However, this is a very coarse scale of analysis and Chapter 4 demonstrated that the model of susceptibility developed in this thesis is more accurate at predicting IPs than stochastic mathematical models of disease transmission. Lawson and Zhou (2005) developed maps of relative risk for FMD in Cumbria during the 2001 epidemic using Standardised Incidence Ratio (SIR) calculations. These data are fortnightly counts of IPs aggregated to the parish level and spatial dependency is modelled by including terms in a Bayesian binomial model for the x and y ordinates of the parish centroid. The SIR implicitly accounts for the population at risk in terms of the number of farms. However, this thesis

has demonstrated that numbers of animals in the locality and farm size is more important in determining risk of spread than the number of susceptible holdings (Chapters 4 and 5). Furthermore, by aggregating to the parish level a degree of accuracy is lost as the median number of farms on the census per parish is 7 but the 95 percentiles lie in the range from 1 to 50. As FMD was in areas of greatest livestock density there are likely to be greater numbers of farms in parishes with greater animal density.

A model based upon a distance kernel was used to map the risk of H7N7 Highly Pathogenic Avian Influenza (HPAI) transmission in the Netherlands (Boender et al., 2007). The authors used a transmission kernel to demonstrate that there were two poultry dense foci in the Netherlands which were at elevated risk of HPAI. The transmission kernel was a function of interfarm distance and was therefore strongly correlated with poultry density, as a result the risk map closely reflected the poultry distribution in the Netherlands. In a similar study applied to Classical Swine Fever (CSF) Boender et al. (2008) use data from previous CSF epidemics to parameterise the kernel and derive a risk map at the herd level using these data.

Risk mapping of H5N1 HPAI in Vietnam has been conducted by Pfeiffer et al. (2007). The authors have records of reports of HPAI H5N1 in 2,771 of 10,067 communes in Vietnam. HPAI appeared in three distinct temporal waves in Vietnam during the period 2004-2006 and the outcome variable for the spatial multivariate model is whether a commune reported a case during a particular wave. Predictors included the Normalised Difference Vegetation Index (NDVI), area used for paddy and domestic poultry density as well as epidemic period and region. Spatial autocorrelation in the predictors was accounted for by including the district to which the commune belongs as a random effect. As HPAI was clustered in different regions of the country in different epidemic periods, the interaction between epidemic phase and region is a very strong risk factor in the model, however other predictors such as distance to an area with high population density and poultry density were also important in the model. The predicted probability of infection for each commune for each of the three epidemic periods are plotted to generate choropleth risk maps of HPAI.

On risk maps areas at elevated risk, known as 'hotspots' are of interest. During the 2001 FMD epidemic the majority of disease transmission was over relatively short distances, so the shape and size of hotspots is important. This is because if the virus is introduced into a relatively small hotspot the local epidemic is likely to burn itself out relatively quickly. However, if the hotspot is large there is much greater potential to generate a substantial epidemic.

This chapter will describe the development of risk mapping based upon the risk factors for holding level susceptibility (Chapters 4) and the risk of an IP transmitting disease given geographical barriers (Chapters 5 and 6). Maps of retrospective risk for the 2001 epidemic will be created, as well as general non-epidemic risk maps. Furthermore, the techniques will be applied prospectively to an outbreak of FMD in Surrey, England in 2007. Additionally risk mapping will be carried out for Scotland which was an uninfected area but could have become infected given a set of potential introduction events via animal movements from Surrey.

7.2 Methodology

Developing the risk surface requires the combination of two models developed elsewhere in this thesis:

1. The probability of a holding becoming infected (Chapter 4).

2. The probability of a holding already infected with FMD transmitting infection to further holdings (Chapter 5) given the potential of landscape barriers to inhibit disease transmission (Chapter 6).

These models must be combined to produce the map of risk. As the potential for a holding to transmit infection is dependent upon that holding already being infected the risk ascribed to each holding (R_h) can be described by equation 7.1.

$$R_h \approx Susceptibility \times Transmission|Barriers \tag{7.1}$$

Some method of transforming the models described is Chapters 4, 5 and 6 is required in order to generate inputs for equation 7.1 which can be applied to all

holdings. The process for doing this will be described for each of the susceptibility, transmission and barrier models.

7.2.1 Susceptibility

A logistic regression model of holding level susceptibility was developed in Chapter 4. This model can be used to calculate modelled values for different datasets which represent the probability of being an IP given another 2001 type of epidemic. As the model was generated in the R statistical environment these values are generated using the predict.glm function. The model contains 6 main effects (two - country and species are factors with three levels) all of which can be calculated for each holding. The only value which cannot be generated directly from agricultural census data is the distance to the infectious seed which is dependent upon the distribution of seeds at the start of the epidemic. There are four possible solutions to the problem of identifying seeds which will be used in this chapter:

1. **Distribution of seeds from epidemic data:** This formed the variable in the initial model and therefore the modelled values can be used to generate a risk map for the 2001 or 2007 GB epidemics.

2. **Each holding is a seed:** Calculate an average modelled value for each holding based on every other holding in turn being a seed. Given the population of holdings (n) the mean probability of holding h becoming infected given that each other holding is a seed will be calculated (S_h). Therefore, this is the mean value of P_{hi} where P_{hi} is the probability of holding h becoming infected given that i is a seed where i is each farm in n and $i \neq h$. This is summarised as:

$$\begin{aligned} S_h &= \frac{\sum_{i=1}^{n} P_{hi}}{(n-1)} \quad i \neq h \\ Seed &= i \end{aligned} \tag{7.2}$$

3. **A farm identified at high risk of being a seed:** Farms receiving animals from holdings known to have FMD or from areas in which FMD is present are treated as seeds. This is of relevance at the start of an epidemic if there have been movements of potentially infected animals.

4. **No seeds:** The model described in Table 4.9 which does not have a distance to seed parameter is used.

The distribution of risk over the full course of the epidemic is very dependent upon the control policies which were implemented. During the 2001 epidemic this is particularly pertinent to phases 2 and 3 during which extended culling was implemented. In future epidemics extended culling on the scale of 2001 may not be employed. To overcome this a model in which the outcome was only IPs from Phase 1 will be used (Table 4.11). During phase 1, control by culling was moderate and relatively consistent across space and time so is more likely to resemble the control approach that would be implemented in future outbreaks.

7.2.2 Transmission

The transmission model (Chapter 5) contains 7 main effects. Some of these are static variables derived from demographic data which can be applied to all holdings. Others are pertinent only to holdings which were declared IPs. The use of these variables is described below:

1. **Infected period:** Pertinent to IPs only. A value of 10 days will be applied to all holdings as this is equivalent to the median infected period for the 2001 FMD data.

2. **Epidemic phase:** Pertinent to IPs only. As the risk mapping is only being applied to phase 1 this will be set to phase 1.

3. **Laboratory result:** Pertinent to IPs only. This will be treated as positive in all instances thus assuming that diagnosis is perfect in every instance.

4. **Cumbria:** It will be assumed that the features of Cumbria which result in IPs in Cumbria being more predisposed to transmitting infection would apply to all epidemics. Furthermore, it is assumed that similar processes do not operate in other counties in which the virus may not have been seeded prior to the NMB.

5. **Species on the holding:** Mixed (reference level), cattle only and sheep only. Can be derived for all holdings from the census.

6. **Cattle density in 3km:** Can be calculated for all holdings from agricultural census data

7. **Sheep densities in 3km:** Can be calculated for all holdings from agricultural census data

7.2.3 Barriers

The transmission model describes the probability of transmission within 3km of an IP. Likewise the barrier model (Chapter 6) describes the protective effect of river and rail barriers on transmission within 3km of an IP. Therefore, the modelled transmission value should be transformed to reflect the presence of barriers. Thus, the proportion of farms in the 3km ring which are hidden by a barrier will be calculated as described by Figure 7.1.

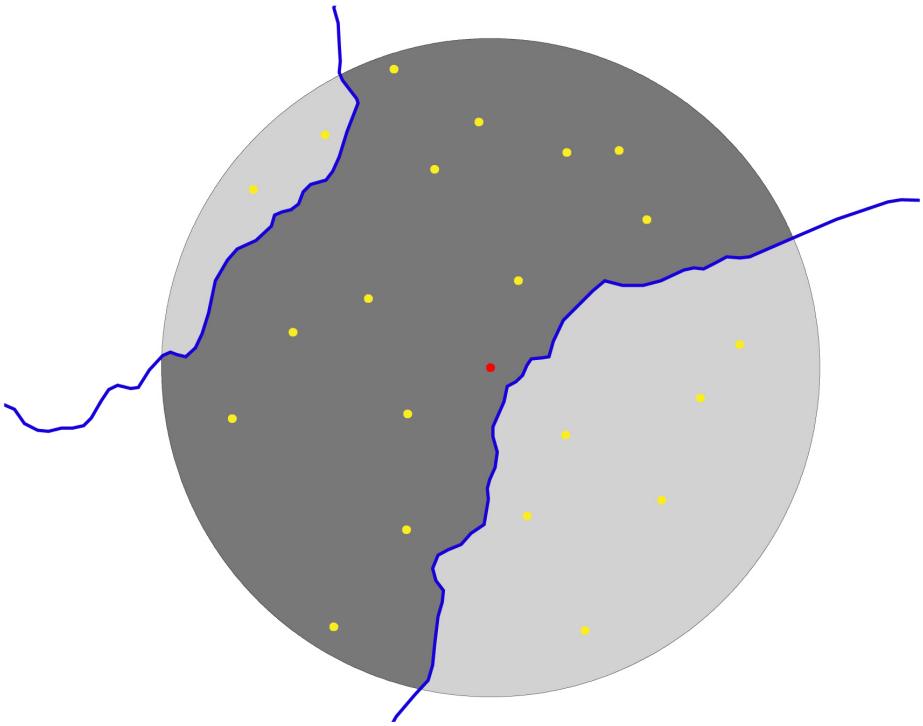


Figure 7.1: The 3km ring around the IP (red point) along with semi-permeable river/rail barriers (blue lines) and the distribution of farms (yellow points) within that ring.

The modeled values of transmission (T_h) will be transformed to T'_h based on the proportion of farms in the 3km ring which are hidden by barriers using the formula

in equation 7.3. In equation 7.3 P_a is the proportion of farms in the 3km ring which are not hidden by barriers and P_b the proportion which are so $P_a + P_b = 1$. O_+ is the odds ratio quantifying the decreased likelihood of infection associated with the presence of a barrier, from Chapter 6 we know that O_+=0.54. However, the total risk of transmission should remain the same and therefore $\sum_{n=1}^{n} T_h = \sum_{n=1}^{n} T'_h$. To ensure this, those holdings not separated by a barrier should be multiplied by a function O_-

$$T'_h = T_h \times P_a \times O_- + T_h \times P_b \times O_+ \tag{7.3}$$

where O_- is

$$O_- = \frac{\sum_{i=1}^{n}(T_{hi}) - O_+ \sum_{i=1}^{n}(T_{hi} \times P_a)}{\sum_{i=1}^{n}(T_{hi} \times P_b)} \tag{7.4}$$

where n is the population of farms. The resulting value of O_- is 1.12 given that O_+ is 0.54 (Table 6.5).

The proportion of farms inside the barrier was calculated using an ESRI Arc Macro Language (AML) script in ArcInfo 9.0. The code takes each holding in turn and creates a 3km buffer around the holding. An OS Strategi vector dataset supplied by EDINA Digimap of the combined river and rail network is clipped with this buffer and the clipped layer itself is buffered to 1m to simulate its width. Using the buffered river/rail network the 3km farm buffer is broken into several polygons using the Erase function and the polygon containing the farm is identified. The number of holdings in this polygon and the number in the entire 3km buffer are written to a file along with the CPH of that holding.

7.2.4 Mapping values

Taking R_h as the product of S_h and T'_h generates a value of risk for each holding. These values will be mapped to describe spatial variations in risk. There are a total of 129,226 holdings for which R_h was calculated, therefore, mapping individual points would generate a very unclear picture as there is too much data to detect spatial patterns. Instead of mapping discrete data points a continuous gridded raster surface

will be created by aggregating values where the aggregated value is some summary statistic within the neighbourhood of each raster cell. Interpolation techniques, in particular Kriging (Bailey and Gatrell, 1995) are commonly used for such mapping but were inappropriate for these analyses as the data should meet the following criteria in order to interpolate which these data do not meet (derived from Bailey and Gatrell (1995, p.143)):

1. The data are a spatially continuous surface. This applies to variables such as elevation or rainfall for which every point has some value and the data are implicitly spatially autocorrelated. These data are not spatially continuous, they are a number of discrete events (in this case the events are the farms). Furthermore the data are not implicitly autocorrelated in the same way as a variable such as rainfall is. For instance a small sheep farm (low risk) can be the neighbour of a large mixed farm (high risk).

2. The data must be a sample of the surface. For instance it is not currently possible to measure rainfall across the entire surface. Interpolation uses sampled spatially autocorrelated datapoints to impute datapoints which were not sampled, however in these analyses the farm data is the population and not a sample.

3. Kriging is not a method applicable to finite populations and therefore does not allow for differences in the density of the population. For instance consider two areas of the same size, one is populated with 5 farms each with a modelled value of 0.2, the second area with 10 farms with a modelled value of 0.1. In both areas the expected number of cases in each area is 1 (5×0.2 and 10×0.1). However, as Kriging extrapolates from the given values to create a surface the interpolated value across the first area would be 0.2, and 0.1 in the second area.

Rather than interpolation, data have been aggregated by taking the summary statistic (sum) of all points (farm holdings) within some radius of each raster cell, this method is described by Figure 7.2. Two parameters must be set, the cell size (c) and the radius of the search circle (s).

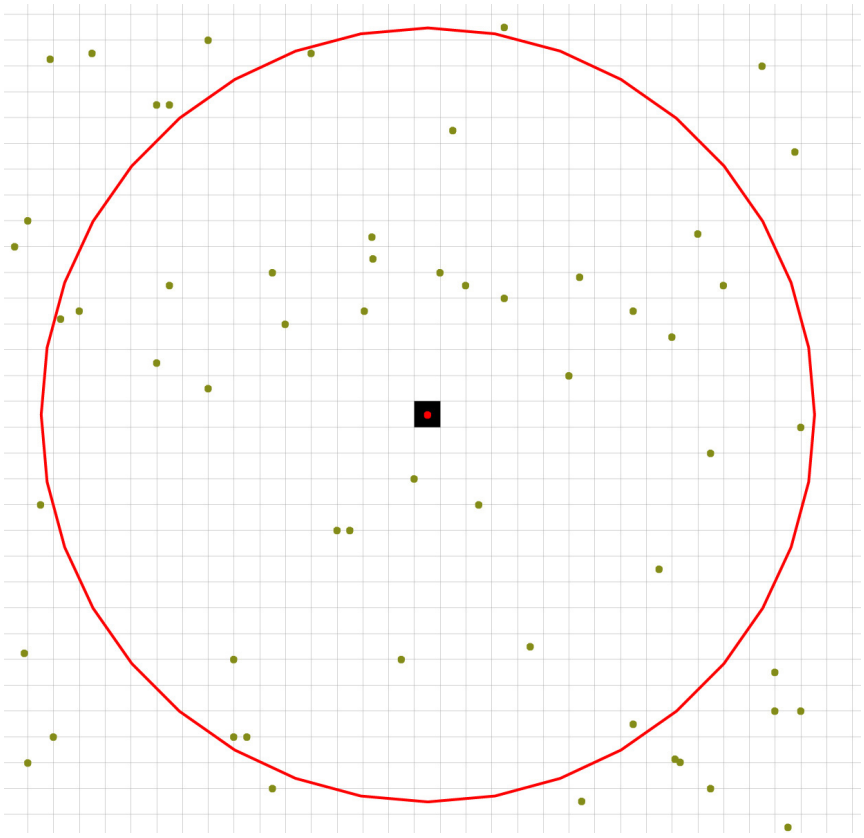
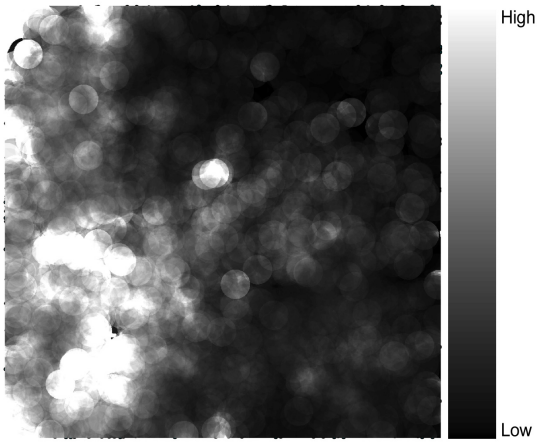


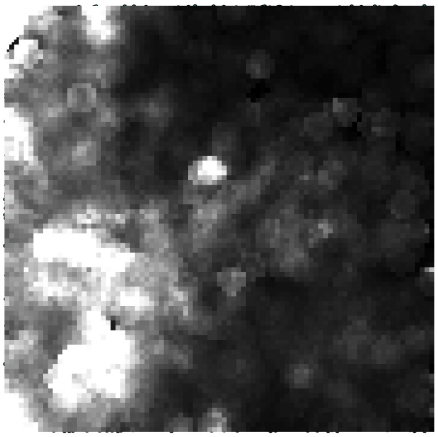
Figure 7.2: A 200m grid is created (grey outline represents the cells of the grid and the black cell represents the grid cell for which data is being calculated. The data for the black cell is some summary statistic of all the points (green points) within a 3km ring (red ring) of the centroid of the cell (red point) in question.

The effect of altering s and c can be seen in Figure 7.3. Comparing Plot 7.3(b) with Plot 7.3(a) highlights the effect of large values of c which increases the granularity of the output map. The chosen value of c is a trade off between the granularity of the output and computational power. The result of this trade off should create a non-granular output whilst still allowing efficient browsing of the data. Altering the values of s affects the actual values in the output data. At smaller values of s individual holdings with large values have a greater influence than at greater values at which the output is much smoother. However, extending s too far may compromise the local nature of the analysis. This is shown by comparing Plot 7.3(c) with Plot 7.3(d). In Plot 7.3(c) the value of s is relatively small and individual holdings have a great effect with pockets of higher values. Plot 7.3(d) has a large value of s and the effect is a much smoother surface. Values of s of 3km and c of 100m corresponding to Plot 7.3(a) were selected for these analyses because this represents a

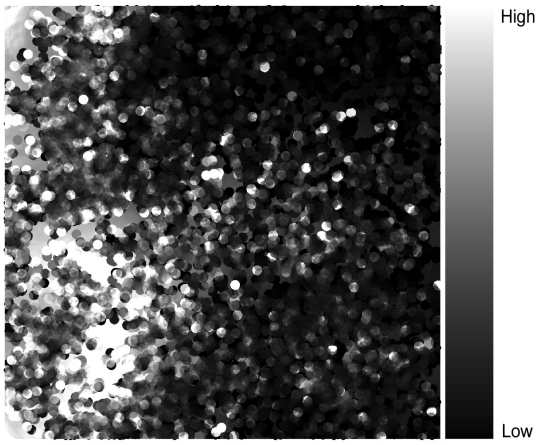
reasonable trade off between clarity and computational efficiency. Furthermore, the coordinates are from OS grid references which have a resolution of 100m.



(a) c=100m, s=3km



(b) c=1000m, s=3km



(c) c=100m, s=1km

(d) c=100m, s=10km

Figure 7.3: Examples of the effects of altering cell size and search radius on the same dataset. The data used to generate these maps are sheep numbers on holdings.

One drawback with this method of aggregation is the generation of slivers at the intersection of two points (Figure 7.4). This is principally a problem in instances where data are sparse, in denser areas such features are smoothed out.

As a result the maps will show the following depending upon the statistic being summarised:

1. **Susceptibility maps:** The sum of modelled values for susceptibility within

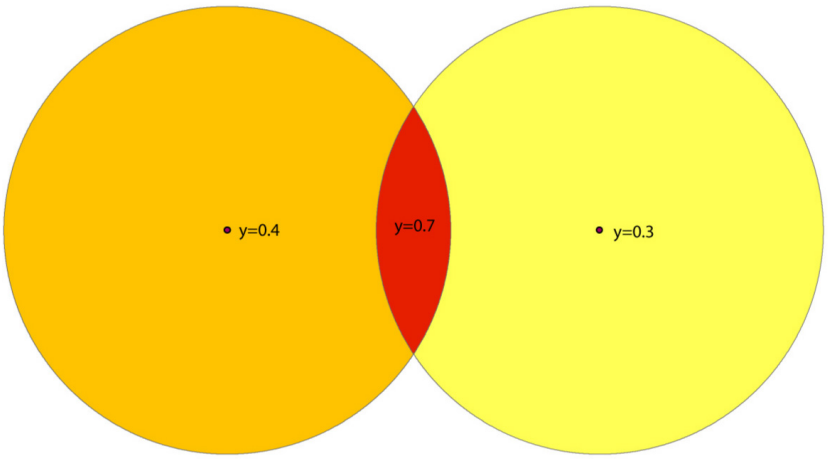


Figure 7.4: The generation of a high risk sliver (red area) at the intersection of two 3km search areas. Such slivers are an artefact of sparse data and will be visible in areas with a low density of farms.

3km of each raster cell. Therefore the maps show the expected number of IPs within 3km of each cell. Methods 1 and 3 outlined in section 7.5 will generally generate larger values than method 2, unless the seeds are in areas of extremely sparse livestock density. However, the magnitude of the values of risk will be dependent upon the number and distribution of seeds. More seeds dispersed more widely will ensure that the distance to seed parameter is low for most holdings and therefore the risk of holdings becoming infected is higher. In method 2 the effect of taking one seed at a time is averaged over the total number of farms, as only a single seed at a time is treated the averaged values are very small.

2. **Transmission maps:** The mean of modelled values for transmission within 3km of each raster cell. Given that IP i is infected and that all n holdings within 3km of i are susceptible the probability that i will infect at least one other holding. To generate the risk map this is repeated for every holding and the sums of the probabilities mapped. Therefore, the map represents the mean expected number of IPs spreading FMD within a 3km radius of each raster cell.

3. **Combined susceptibility and transmission maps:** The product of the modelled transformed transmission and susceptibility values. Therefore, this represents the probability of a farm transmitting infection given the probability

that the farm becomes infected. These values are also summed such that each raster cell is the total for all holdings within 3km. Therefore the map represents the expected number of IPs and IPs transmitting infection within 3km and is therefore the most complete measure of virus spread.

The Orkney and Shetland Isles were not included in the mapping as they were a long way to the north of the remainder of GB and their inclusion would reduce the clarity of the remainder of the country.

7.2.5 Risk mapping the 2007 FMD outbreak in Surrey

Section 1.2 described the outbreak of FMD in Surrey in 2007. In this outbreak there was a long temporal lag between the first and second clusters during which time many considered the area disease free, restrictions were eased and on farm vigilance was reduced (DEFRA, 2007c). As a reflection of the period with eased restrictions between outbreaks risk maps will be generated for the two clusters in isolation. The seeds for the first cluster are IPs 1 and 2 as IP2 is estimated to have been infected before IP1 was confirmed and restrictions were put in place (DEFRA, 2007d). In the second cluster IPs 3 (B and C), 4 and 5 were infected before IP3 was declared, however IPs 6B, 7 and 8B are all secondaries so IPs 3 (B and C), 4 and 5 are treated as the three seeds (Figure 1.4).

FMD risk for the affected areas of Surrey will be mapped using agricultural census data from June 2006 supplied by the National Epidemiology Emergency Group (NEEG). Holdings with cattle and/or sheep were extracted from this data and risk maps generated for the two clusters using the seeds described above. The susceptibility model alone will be used to generate these risk maps as this is substantially more accurate and therefore the outputs more reliable than the transmission model. Furthermore, the transmission model is a dynamic model contingent upon FMD being present. Therefore, the susceptibility model describes the risk of IPs in the subsequent generation, whilst the transmission model will describe the risk of subsequent generations. Evaluation of the use of susceptibility data alone to predict transmission risk will be conducted by plotting the modelled values of susceptibility against transmission. The relationship between transmission and susceptibility will be evaluated

by assessing the fit of linear models with increasing number of polynomial terms until the addition of further terms leads to an increase in AIC.

The relative likelihood of an individual holding becoming infected will be evaluated by assessing the ranks of all actual secondary IPs (IPs 6, 7 and 8 for outbreak 2) against the ranks of all holdings within the surveillance zone (SZ) for outbreak 2. There were no secondary IPs in outbreak 1. The total expected number of cases will be calculated by taking the sum of predicted values for the two outbreaks.

7.2.6 Risk mapping FMD in Scotland during the 2007 outbreak

The potential for spread following introduction of FMD into Scotland via animal movements was evaluated using animal movement data. Both direct and indirect movements of cattle, sheep and pigs into Scotland from farms within a 20km radius of the Pirbright laboratory were identified from the Cattle Tracing System (CTS), Animal Movement Licensing System (AMLS) and Scottish Animal Movement Licensing System (SAMS) databases. All movement data was analysed using an algorithm written and implemented by Dr Nick Savill (University of Edinburgh). The algorithm searched for both direct (single movement from Surrey into Scotland) and indirect (a series of movements connected by animals sharing premises en route) movements. This identified 15 indirect movements into Scotland and 30 indirect movements into northern England from which the disease could spread locally across the border into Scotland. The locations of holdings receiving these animals were treated as infectious seeds for the susceptibility model. These seeds could be handled in two ways:

1. The model is run n times where n is the number of potential introductions identified. A modelled value is calculated for each seed in turn and the mean is taken as the overall measure of risk.

2. All n introductions are taken as seeds simultaneously, so there are n seeds.

The second method was used to generate the risk maps. This is because an individual introduction may have the potential to cause a large outbreak but this could be masked if its effect was averaged over n. Risk maps were generated using the phase 1 susceptibility model and populated with Scottish agricultural census data from 2006

supplied by the Scottish Government. The map shows the likely number of secondary cases arising from individual introductions in the same way as the risk map based upon the 2001 seeds. However, as the seeds are not actual IPs, the map shows the potential risk if a seed becomes a case and should be used for targeting resources for inspecting possible introductions.

7.3 Results

7.3.1 Susceptibility

Figure 7.5 is the risk map parameterised with 2000 census data and the mean distance to every other holding (section 7.5 - solution 2) as the seeds. The high risk areas are broadly in the same areas as the FMD cases although the following observations can be made about the distribution of high risk areas:

1. **The Outer Hebrides (inset a).** This high risk area stretches along the north coast. The apparent high risk in this area is likely to be an artefact of the structure of farm holdings and the way farms are modeled rather than actual high risk of susceptibility to FMD. Figure 7.6 illustrates the clustering of farms in lowland coastal areas. These locations are likely to be the addresses of farm holdings which farm the higher land inland. As a result of the apparent proximity of the holdings to each other the value for distance to seed is very low which results in a high density of high modelled values.

2. **Northern England (inset b).** This area is characterised by an area of higher risk stretching through the FMD belt of Cumbria and areas of high risk in the high intensity dairy farming areas of Cheshire.

3. **Wales and the Welsh borders and south west England (inset c).** There is a large area of high risk in the south west of Wales where the disease was not seeded. However, areas in the south west of England and the Welsh border areas where the disease was seeded are at higher risk, albeit a patchy high risk in the Welsh border area.

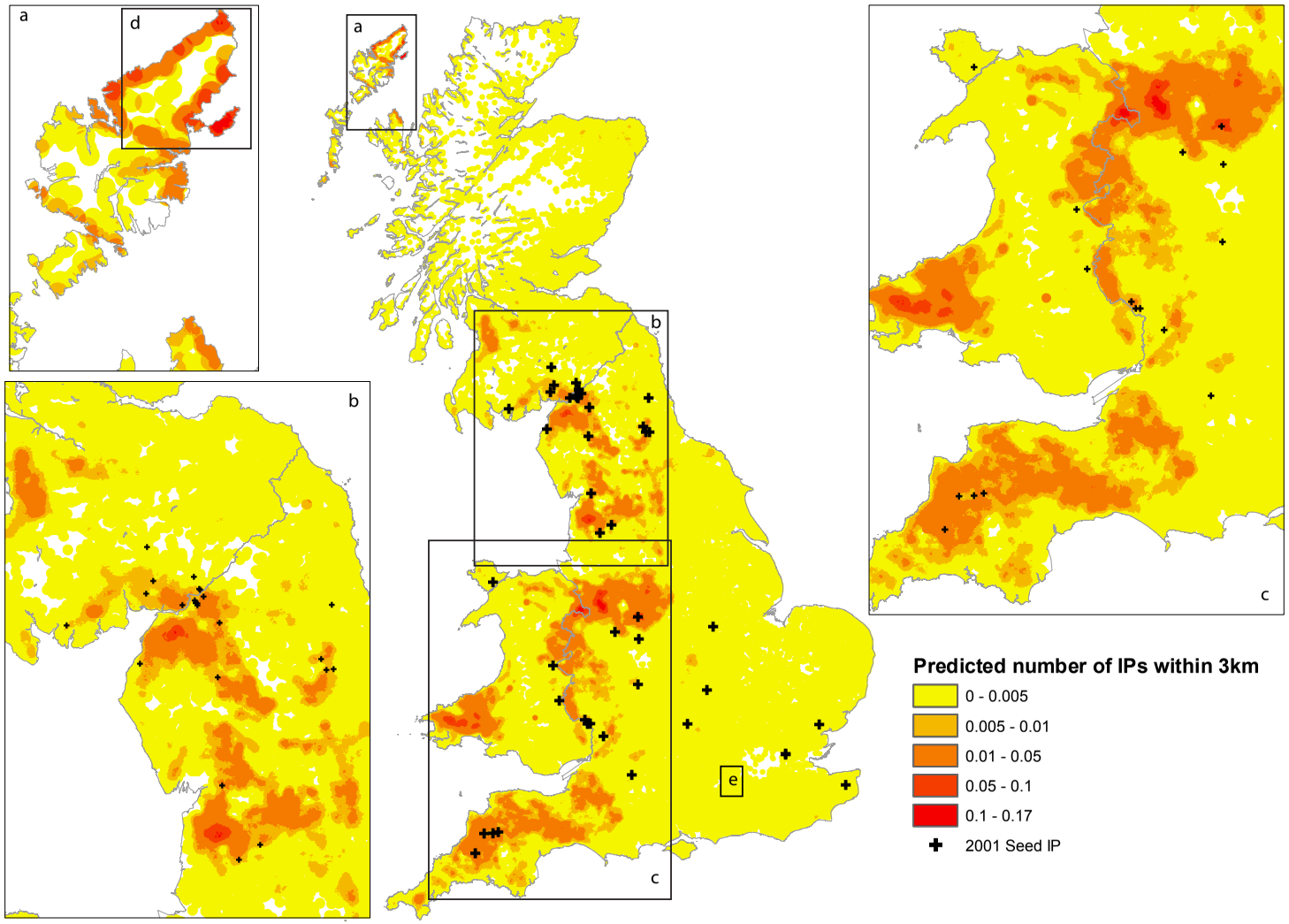


Figure 7.5: Susceptibility risk based upon distance to each other holding and is independent of the seeds in 2001. The relative risk values correspond to the sums of sums of mean modeled values using the method described by section - method 2, where c=100m and s=3km. Inset a is the northern Outer Hebrides, inset b the north of England and Southern Scotland and inset c Wales, the Welsh borders and south west England. The box labelled d in inset a corresponds to the area of Figure 7.6. Inset e corresponds to the area covered by Figure 1.4. The seed IPs from the 2001 epidemic are added to illustrate their spatial distribution in relation to FMD risk.

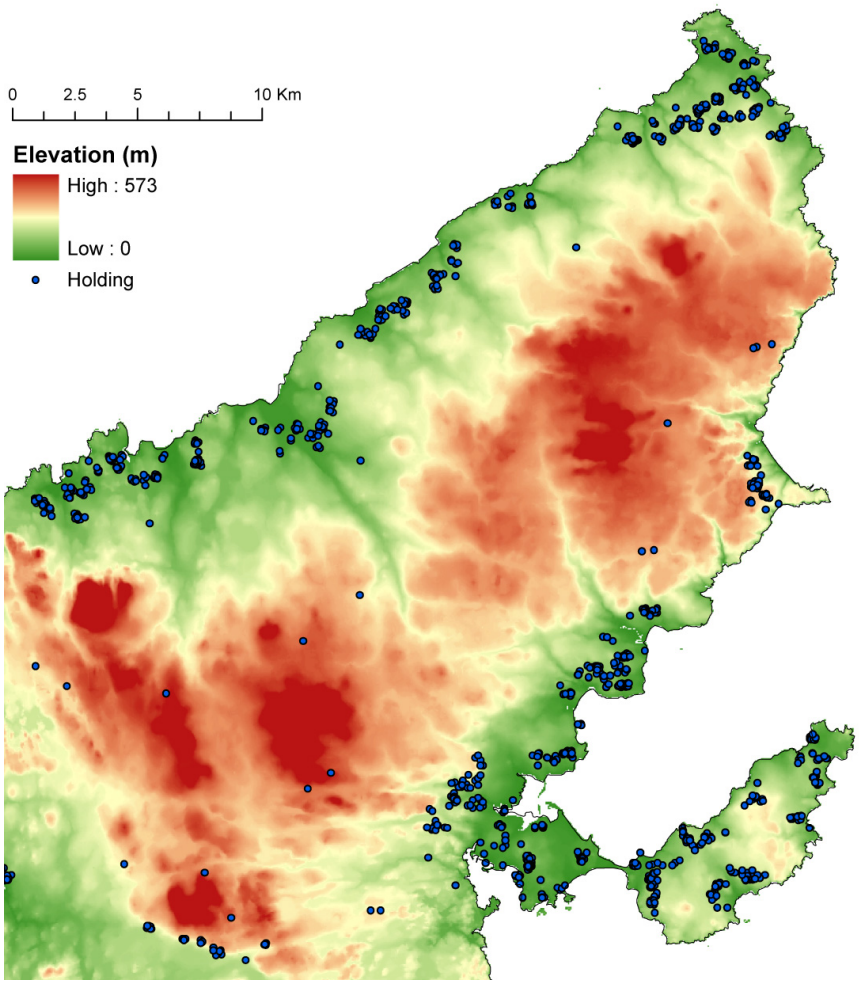


Figure 7.6: The distribution of farm holdings in relation to elevation in the Outer Hebrides. The area shown corresponds to box d in inset a on Figure 7.5. Elevation data is derived from a OS Panorama 50m Digital Elevation Model (supplied by EDiNA digimap (www.edina.ac.uk)).

The static risk map presented in Figure 7.5, should be compared to the map of the output of the susceptibility model without reference to the distance to seed variable (Figure 7.7, section 7.5 - solution 4). In this map the Outer Hebrides are no longer a high risk area, instead risk is distributed throughout south west and western England. This demonstrates that these higher risk areas in the Outer Hebrides are an artefact of the data which records a high density of farms in this area.

Further influence of holding density in the susceptibility model can be seen by comparing the higher risk in southern Scotland in Figure 7.7 with Figure 7.5, in particular Dumfriesshire. This area has larger areas of higher risk in Figure 7.7 compared to Figure 7.5. This is because this area is characterised by more sparsely

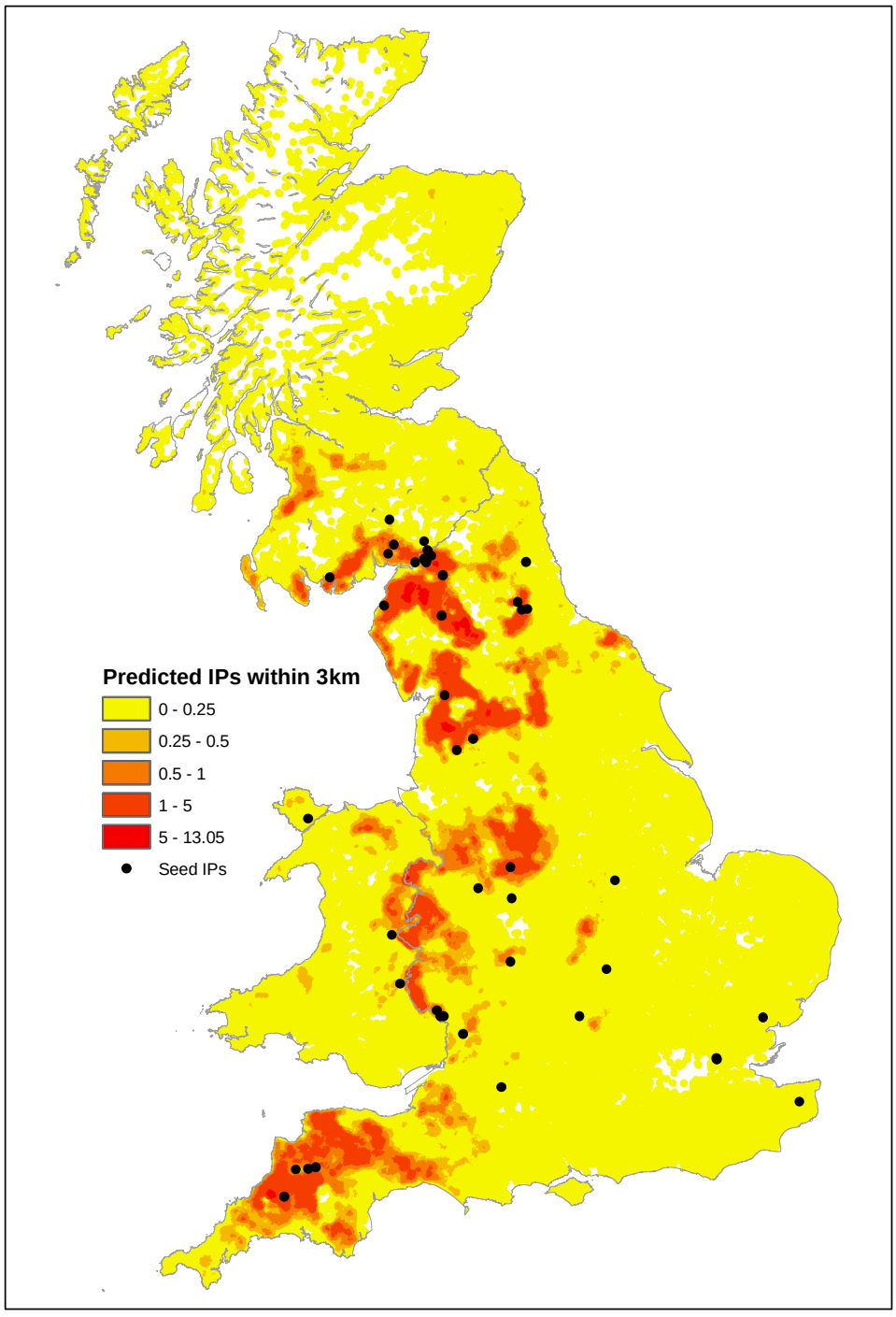


Figure 7.7: Susceptibility relative risk for the susceptibility model without the distance to seed variable, described in section 7.5 - solution 4.

distributed but very large holdings (Table 7.1). Both mean numbers of cattle and sheep on each holding in Dumfriesshire are greater than in Cumbria, but the density

of livestock holdings is almost 50% lower. As a result of the difference in the density of farms there will be fewer close farms in Dumfriesshire and as a result when calculating susceptibility risk based on distances to all other holdings, this area appears at lower risk than areas with greater farm density. The reverse is true for Dyfed where holding density is much higher than Cumbria, yet holdings have around 40% fewer animals (Table 7.1). This reflects this differences in the Dyfed area of south west Wales when Figure 7.5 is compared with Figure 7.7 the former shows high risk for this area and the latter does not.

County	Area (km^2)	Holdings	Hold/ km^2	Cattle	Cattle/ holding	Sheep	Sheep/ holding
Cumbria	6,823	5,114	0.74	510,266	99.8	2,625,583	513
Dumfries'	2,791	1,085	0.39	173,297	159	757,382	698
Dyfed	5,576	7,978	1.43	509,791	63.9	2,281,385	285

Table 7.1: Numbers of holding, cattle and sheep and in Cumbria, Dumfriesshire and Dyfed.

The risk map generated by the model run with the 78 seed IPs from the 2001 epidemic using methods described in section 7.5 - solution 1 is displayed in Figure 7.8. The Cumbria and Dumfries & Galloway (inset a) cluster and to an extent the Devon cluster (inset c) were in large areas of high risk and there was extensive spread in these areas. Although there were areas of high risk in the Welsh Borders area (inset b), these areas are more fragmented than those in south west or north west England. This may explain why the epidemic in this area was more sporadic with fewer, less-clustered cases. In the north east of England there are patches of high risk but a number of cases spilled over into areas of relatively low risk. In south east England however, the risk map accurately identifies those areas which reported the greatest number of IPs.

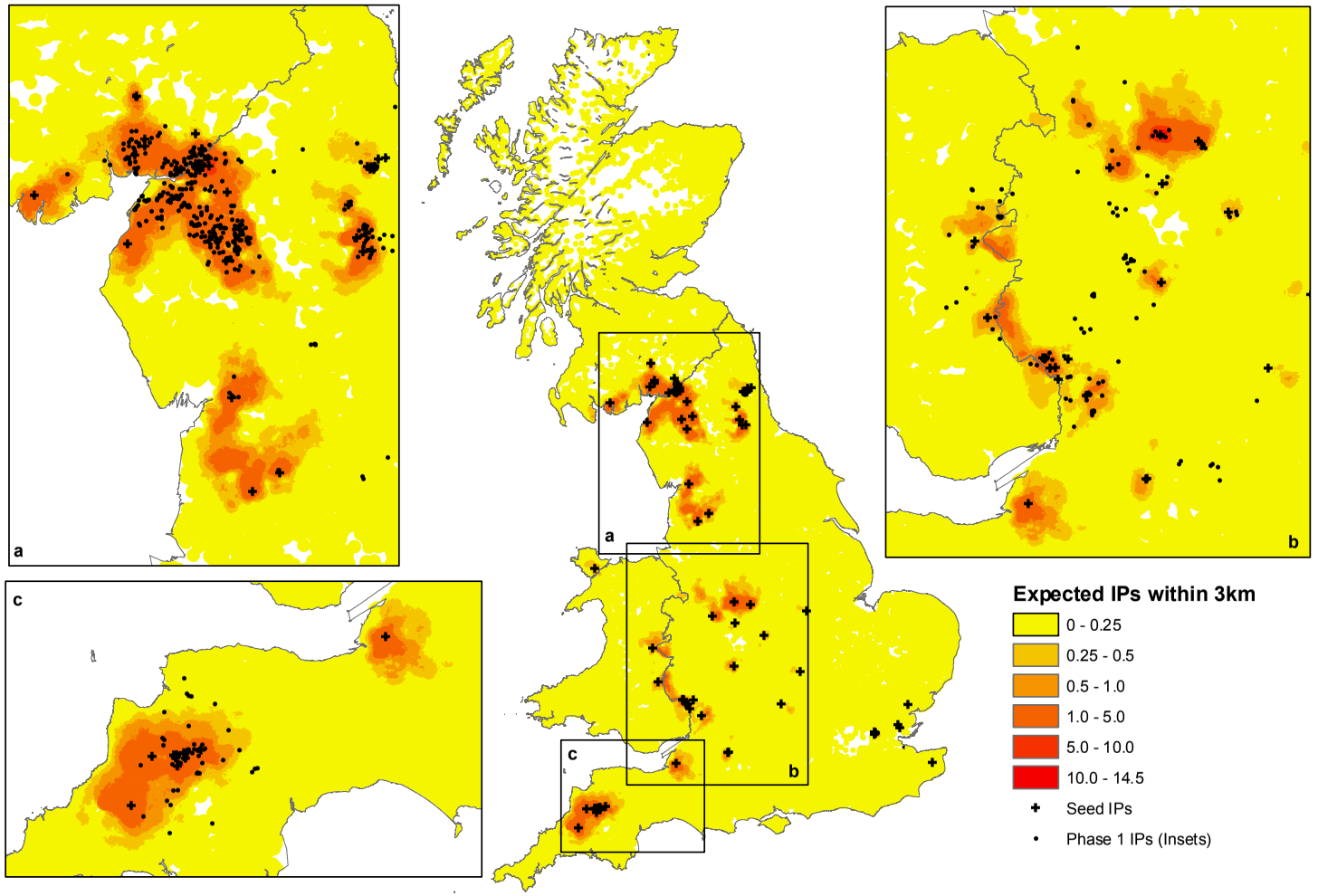


Figure 7.8: Susceptibility risk based upon the distribution of the 78 seeds from the 2001 epidemic. The relative risk corresponds to the sums of modelled values where c=100m and s=3km. Principal epidemic areas are expanded as insets and the distribution of IPs overlain, inset a representing southern Scotland and northern England, inset b the Welsh border area and inset c the south west of England. The distribution of IPs during phase 1 is shown for illustration.

7.3.2 Transmission

Risk of transmission (Figure 7.9) does not show the same distinctive high risk hot spots as risk of susceptibility. There are higher risk areas in Cumbria, Dumfries & Galloway, Wales, the Welsh borders and the south west of England which essentially follows areas of high cattle density illustrated by comparison of Figure 7.10 and Figure 7.9. The effect of cattle densities upon Figure 7.9 can be seen by setting all values for cattle density to the national mean (Figure 7.11). The resulting map is rather homogenous in colour with few differences in risk between regions. The importance of cattle densities reflects the fact that most variables in the transmission model are dependent upon the holding actually being infected and were all set to the same value. Cattle densities were the only major variable remaining in the model for which the values varied between holdings.

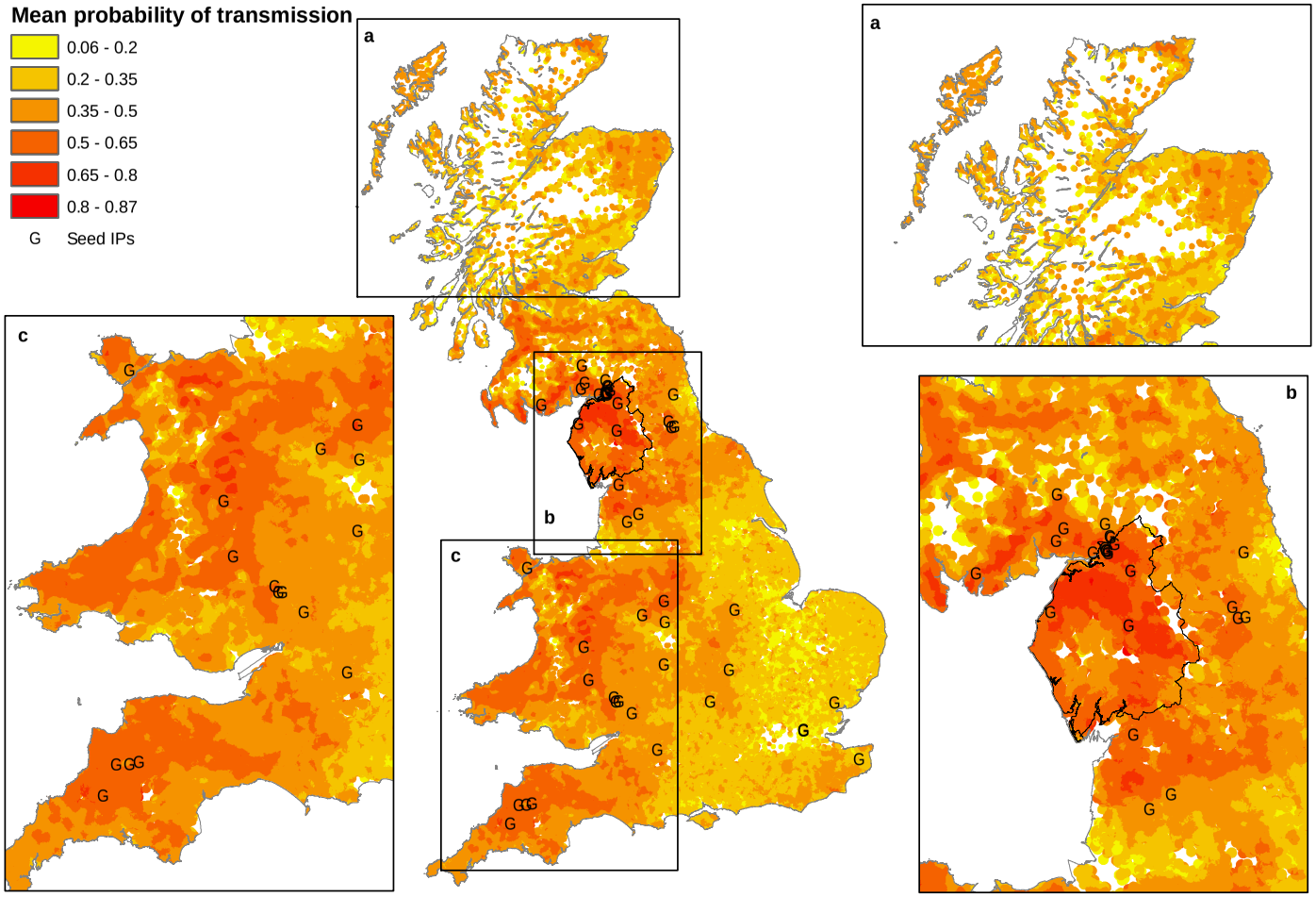


Figure 7.9: Mean transmission risk based upon the model presented in Table 5.6. The values correspond to the sums of modelled transmission risk. The raster is parameterised with c=100m and s=3km. The insets represent regions containing high risk areas of interest.

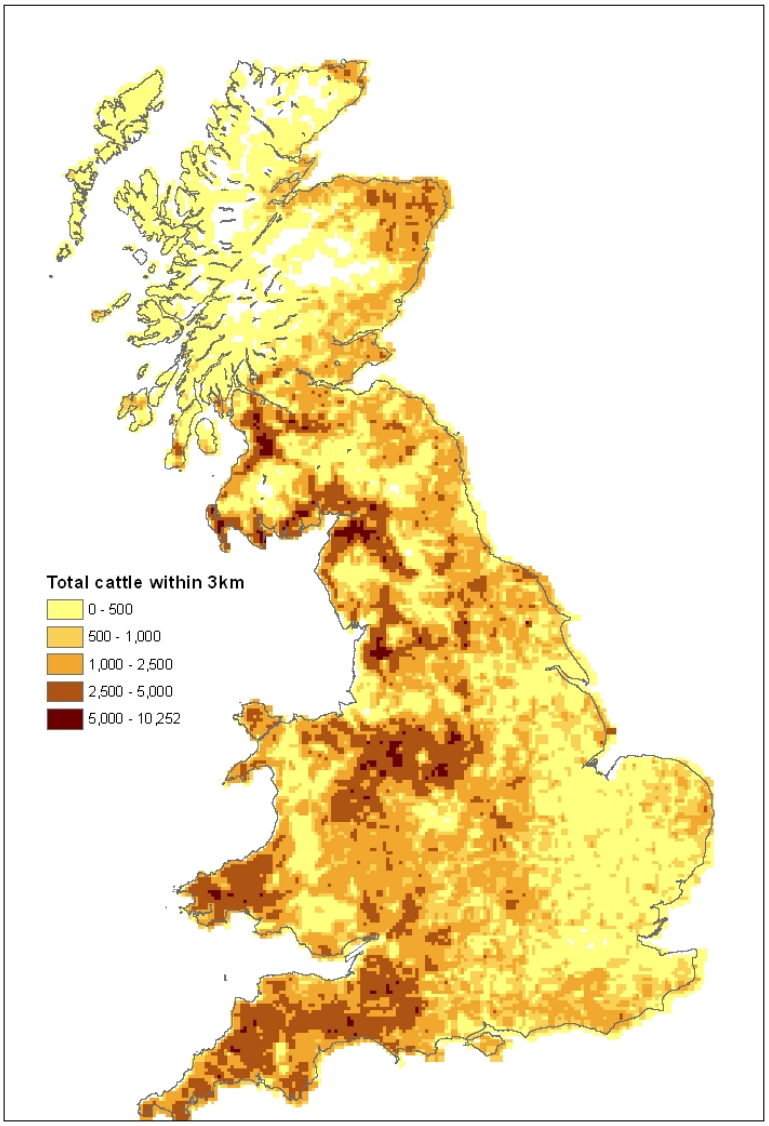


Figure 7.10: Cattle densities in GB. The raster is parameterised with c=100m and s=3km. The values represent the total number of cattle within 3km of each pixel.

7.3.3 Combined risk map

The result of combining the barrier transformed transmission risk with the susceptibility risk model in which every farm is a seed (section 7.5 solution 2) is presented in Figure 7.12. The main areas of risk have not changed extensively from the susceptibility map. However, areas in Cumbria and Dumfries and Galloway (inset a) are at higher risk than in the susceptibility map (Figure 7.7) as a result of the higher

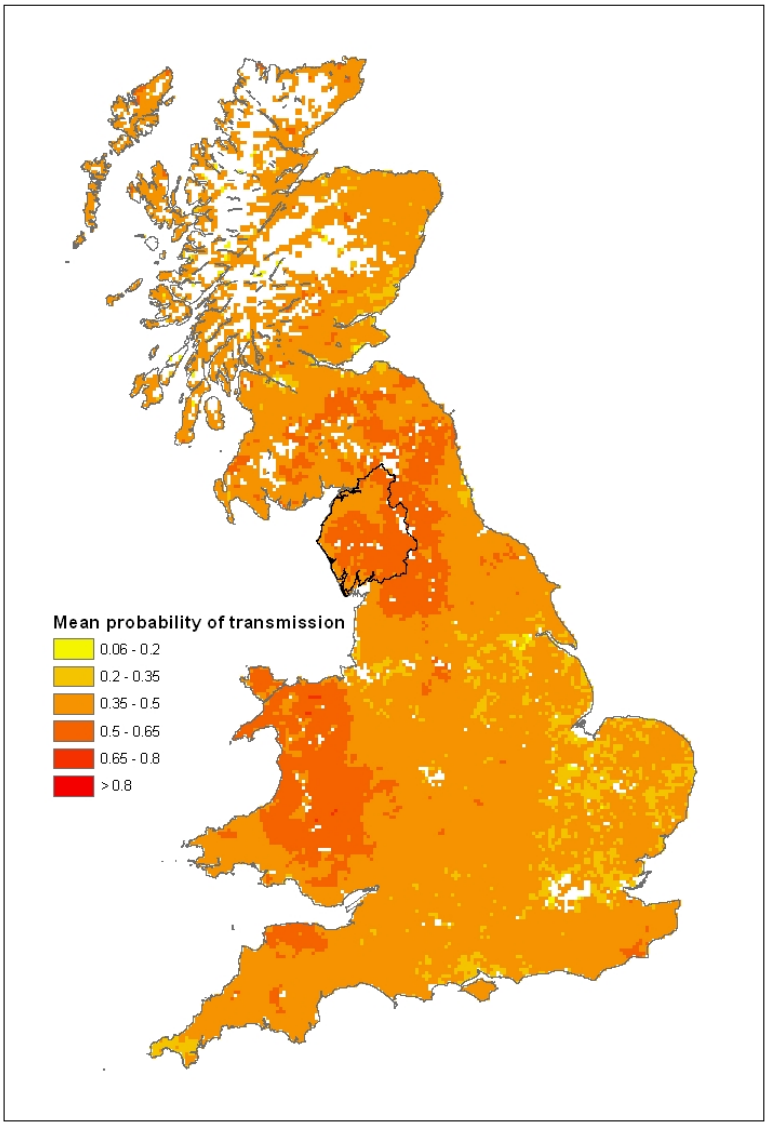


Figure 7.11: Transmission risk based upon the model presented in Table 5.6 but with cattle density set to the national mean value. The values correspond to the sums of modelled transmission risk. The raster is parameterised with c=100m and s=3km.

transmission values in these areas. In contrast, some of the patches of higher risk in the Welsh Border region (inset b) are lower risk compared to the susceptibility map because of a drop in the values for transmission in these areas. In south west England (inset c) it can be seen that the disease was seeded roughly centrally in a high risk area.

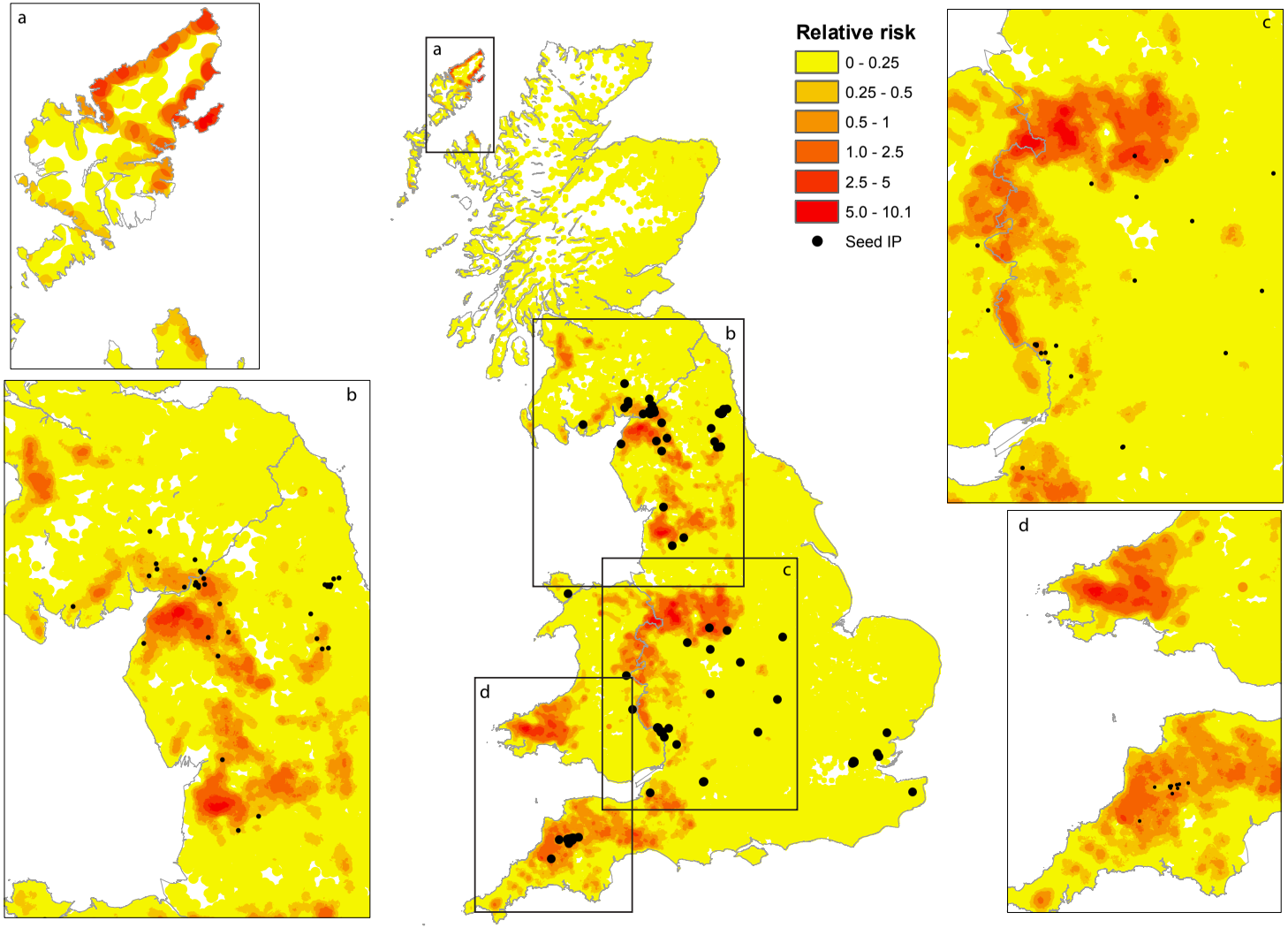


Figure 7.12: Overall risk of FMD based upon the combination of the susceptibility, transmission and barrier models. The raster is parameterised with c=100m and s=3Km. Insets are areas of higher risk which are of interest with a representing the Outer Hebrides, b the north of England and southern Scotland, c the Welsh borders and d the south west England and south west Wales.

There is a highly significant relationship between modelled transmission risk and modelled susceptibility risk (Figure 7.13). The relationship between the covariates is modelled most parsimoniously by a third order polynomial, which returned an AIC of -170484 lower than either first order (-168597), second order (-170460) or fourth order (-170482) linear models. r^2 for the third order model is 0.322 and $F_{3,129222}$=15400, p<0.001. This suggests that there is not a direct linear relationship between transmission and susceptibility. The gap at the bottom of the y-axis transmission data in Figure 7.13 is a result of the transmission model being parameterised for all IPs rather than all farms, as a result there are several IPs specific predictors which were all set to the same value, so all farms already have some base level of risk. The discriminatory power of the model was poorer than the susceptibility model and as a result of the data points being IPs it does not distinguish IPs which have no probability of transmission because they were in areas with no other holdings.

7.3.4 Mapping the 2007 FMD outbreak

Risk mapping of the Surrey outbreak using the susceptibility model suggests that there was a very low likelihood of spread in the vicinity of the epidemic (Figure 7.14). The scale of risk measured as the total number of expected cases within 3km is over two orders of magnitude lower than for the 2001 epidemic, furthermore comparison with Figure 7.5 (box e) shows that this area is very low risk. This is a result of very low livestock densities in the area, Surrey has 25 cows and 48 sheep per km^2 compared to 75 and 386 in Cumbria.

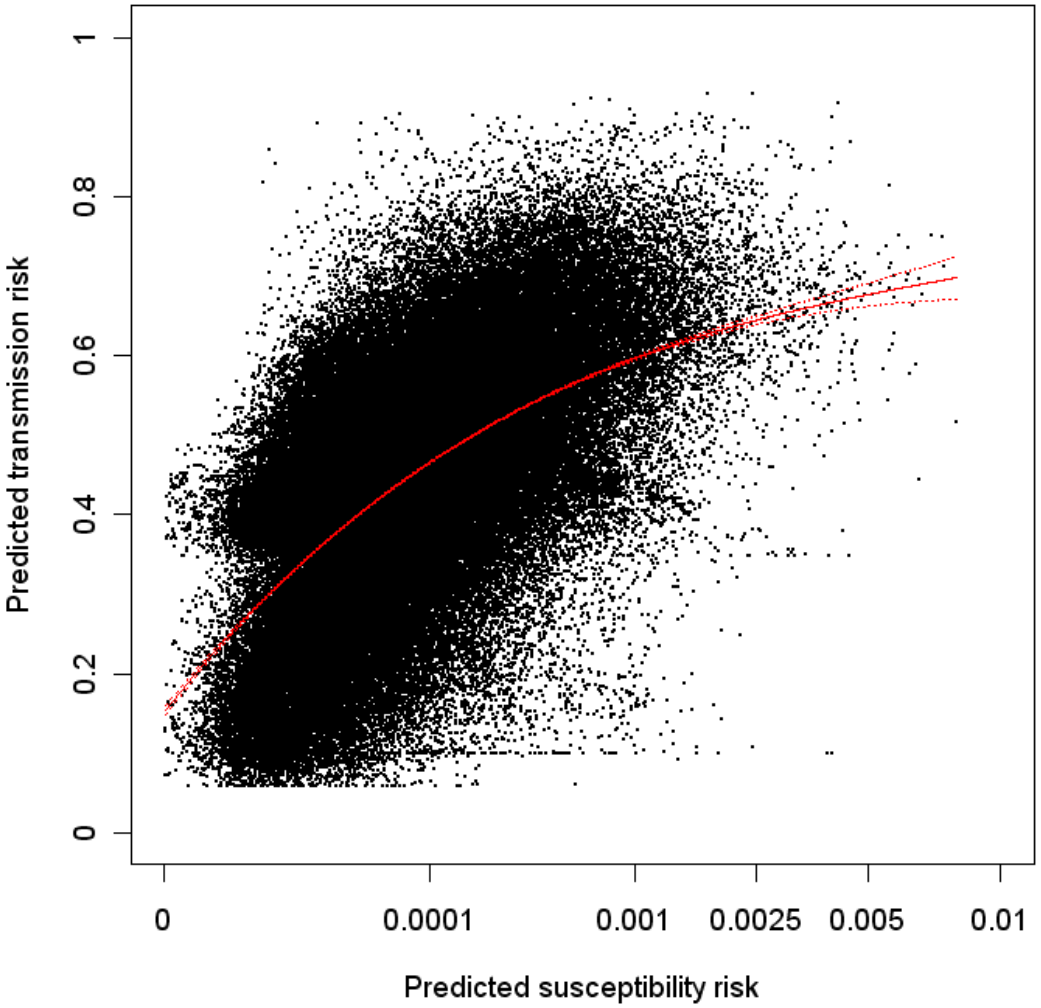


Figure 7.13: Scatterplot of transmission risk against susceptibility risk. Susceptibility risk has been double square root transformed to ease viewing. The solid red line is the regression line from a third order polynomial linear model and the broken red lines are the corresponding 95% CIs.

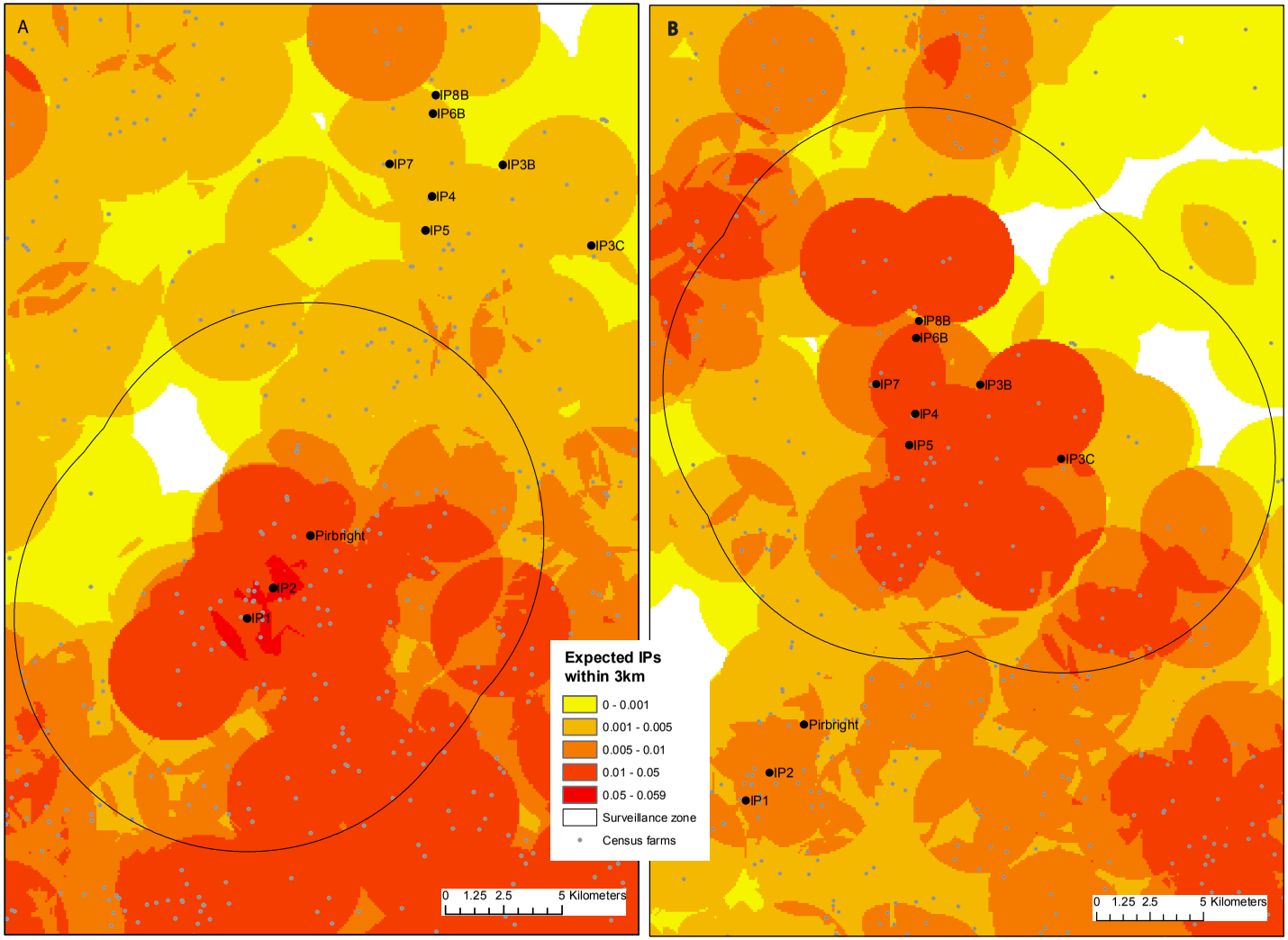


Figure 7.14: FMD susceptibility risk in Surrey in 2007 based upon agricultural census data from 2006 and the first 2 cases being seeds (map a) and IPs 3,4 and 5 being seeds (map b). The area covered by these maps is shown by box e on Figure 7.5. The raster is parameterised with c=100m and s=3Km.

However, these do not appear to predict the risk of infection associated with individual holdings. Analysing the ranks of the modelled values for the 107 farms within the SZ for outbreak 2 shows that the three subsequent IPs in the second outbreak were ranked relatively low; IP6A was ranked 42, IP7 64 and IP8B 54. Although the model was rather more accurate in predicting the total epidemic sizes, the sums of predicted values for outbreak 1 where there were no secondaries was 1.99, the sums for outbreak 2 where there were three secondaries was 1.83.

Potential FMD spread in Scotland was mapped based upon the 45 potential introductions into Scotland and northern England (Figure 7.15). The majority of potential introductions and subsequently risk is focussed in the south of the country. However, there are additional high risk areas associated with introductions further north in the central lowlands. The scale of risk in these maps is similar to that for the 2001 epidemic suggesting that had the virus been introduced the risk of a sizeable epidemic was real.

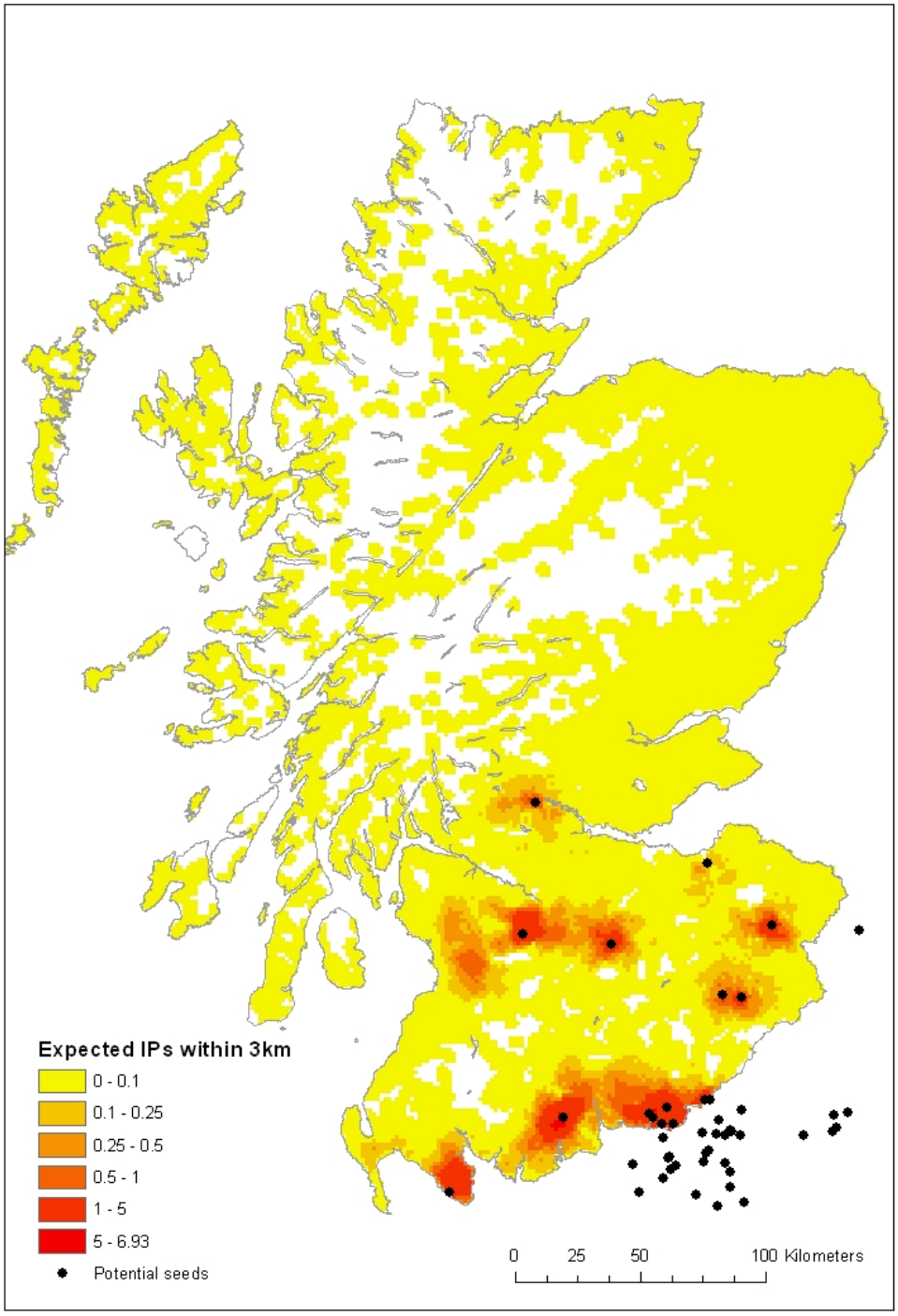


Figure 7.15: FMD susceptibility risk in Scotland in 2007 based upon agricultural census data from 2006 and potential disease introductions being movements of batches of infected animals from the areas affected by FMD in Surrey. The raster is parameterised with c=100m and s=3Km.

7.4 Discussion

A method of producing maps of FMD risk has been developed based upon both the susceptibility of each holding and the likelihood that each infected holding could infect other holdings locally. These values have been derived from three models described elsewhere in this thesis (Chapters 4, 5, 6); as a result the risk map is dependent on the model upon which it is based; furthermore, the epidemic should be similar in nature in terms of species affected, virus strain and control effort. However, the latter point is handelled to an extent by using a model parameterised for only phase 1 of the 2001 epidemic.

As the mapping is derived from the results of modelled risk of infection and transmission during the course of an epidemic the mapping does not display smoothed R_0 values as was done by Boender et al. (2007) and Keeling et al. (2001). Instead the mapping is a combination of the predicted numbers of holdings acquiring and transmitting infection within 3km. However, more information can be gleaned by interpreting maps of susceptibility and transmission in isolation. This is due to differences in the nature of the models, the transmission model being a dynamic model parameterised for IPs only and the susceptibility model is a static model defined at the start of the epidemic. Furthermore, Figure 7.13 demonstrated that there is a positive relationship between susceptibility and transmission, this suggests that using the susceptibility model alone can be used to map both risk of susceptibility and transmission with some accuracy.

The susceptibility risk map based upon the actual distribution of the 2001 seeds (Figure 7.5) shows that the IPs were distributed throughout the high risk areas in northern England (inset a, Figure 7.8), this is with the exception of the high risk hotspot in the southern area of inset a. Likewise the areas of Devon (inset c) are relatively high risk and experienced a number of IPs. In Wales (inset b) despite the extensive seeding and large numbers of animals which created high risk areas, the numbers of IPs in this area were relatively few. The disease was seeded in some areas, particularly in the South East of England that are low risk areas and subsequently did not experience a major epidemic. This shows that this tool can be used as a measure

of risk of FMD 'take-off', given a seeding it shows the likelihood of an epidemic.

The overall map of risk based upon all holdings being potential seeds (Figure 7.5) shows that the virus was seeded in many of the areas of highest risk. However, there are areas of northern and western Wales which were relatively high risk but which do not contain seed IPs. The issues discussed concerning the Outer Hebrides underlines one of the problems with applying this model outside the context of the 2001 epidemic. The model assumes that farming practices are the same in areas which experienced IPs and those which did not. As shown by Figure 7.6, the pattern of farming in this area is very different, with farm holdings clustered around low coastal areas, whilst the areas actually farmed are the higher areas inland. If the disease had been seeded to these areas this would be accounted for in the model; however, as there was no seeding in the Outer Hebrides and this area is over 350km from a seed, the data points in the susceptibility model have little influence in the model. The effect of distance in the Hebrides can be seen in the map of the model without a distance component (Figure 7.7) in which the Hebrides are no longer recorded as being at elevated risk. The problems of heterogeneity in farming practices is not just restricted to outlying areas such as the Outer Hebrides. Table 7.1 demonstrated how the farm population can be distributed very differently in two neighbouring counties both of which experienced the greatest FMD epidemic in their respective countries. The differences in farm densities which are discussed have a strong effect on the pattern of risk.

The risk map for disease transmission is strongly driven by cattle densities. As the model included many IP specific parameters these had to be simplified to create a map of disease transmission applicable to all farms. This smoothed risk map suffers because the dependent variable for the model is whether an IP transmitted infection. Therefore, the model is parameterised upon a subset of farms described by the model in Chapter 4. As a result the IPs which form the data upon which the transmission model is parameterised cannot deal with areas of low risk of disease introduction. Such areas include areas with very few holdings such as northern Scotland still have some risk of disease transmission.

The susceptibility map showed strong spatial differences which were not evident

in the transmission maps. As a result, when results from these are combined and spatially aggregated to form a map it closely resembles the susceptibility map.

The risk maps were applied to analyse the 2007 FMD outbreaks in the UK. In this respect the risk mapping was applied in two different ways:

1. As a start of epidemic tool in Surrey to evaluate the potential risk of spread and identify potential hotspots.

2. To identify the potential for spread in Scotland given a distribution of potential seeds.

The analysis showed that in the two areas of Surrey which experienced outbreaks of FMD the risk of local spread was very low, with just under two secondaries predicted in both outbreaks. This reflects the low densities of animals in Surrey compared to the areas on which the model is parameterised for, principally Cumbria.

In Scotland, had the virus been introduced onto one of the farms identified as receiving animals indirectly from Surrey the likelihood of subsequent spread was high. Application of these methods to other outbreaks are very dependent upon the characteristics of the outbreak and were contingent upon the assumption that the virus strain and the epidemic would behave similarly to the 2001 epidemic. In applying these models to the 2007 outbreak it must be born in mind that although both epidemics were caused by the same serotype (type O) of FMDV, the isolate was different but from the 1967/68 outbreak it appears that the 2007 isolate spreads in a similar manner to that of 2001 (Northumberland, 1968). However, under different circumstances, such as a different serotype or an epidemic in a pig population, modifications to the model may be necessary. Based upon these maps, surveillance could be targeted to those areas with the greatest risk of localised spread. Little is known about the infectiousness of the virus strain, but the epidemic started in and was principally in cattle which, as has been discussed elsewhere (Chapters 4, 5), were the principal driver of the 2001 epidemic.

These maps of FMD risk can be compared with others, principally that of Keeling et al. (2001) which is based upon a stochastic mathematical model (Figure 7.16). Keeling et al. (2001) calculate the risk map using the formula in equation 4.1 to

calculate an R_0 value for each farm. The value plotted is the mean R_0 in the 10km grid squares. Figure 7.16 is directly comparable with Figure 7.12 and Figure 7.5. There are some distinct differences namely that the risk as shown outside the south west of Scotland is distinctly higher in Figure 7.16 compared to Figures 7.12 and Figure 7.5. There are several reasons for this:

1. The susceptibility model has a country level effect in which being in Scotland is a protective effect.

2. The cells in Figure 7.16 are populated using mean rather than the sum of values in each cell. The mean value may exaggerate the risk in cells which are sparsely populated if that cell contains one or a few high risk holdings. These areas of Scotland are sparsely populated with large (and therefore high R_0 of around 5) holdings (Table 7.1). Areas of northern England and the Welsh borders are densely populated with smaller (medium risk holdings, R_0 around 1 or 2) however the overall risk of these combined values is much higher than the mean. Taking the sum of the values allows for the number of holdings and as a result gives a more realistic measure of risk.

3. Local animal densities are much more important predictors in the transmission and susceptibility models than species on the holding. This relates to the previous point that large farms in otherwise low density areas will have a higher estimate of risk in equation 4.1 which places similar importance on holding size and animal density.

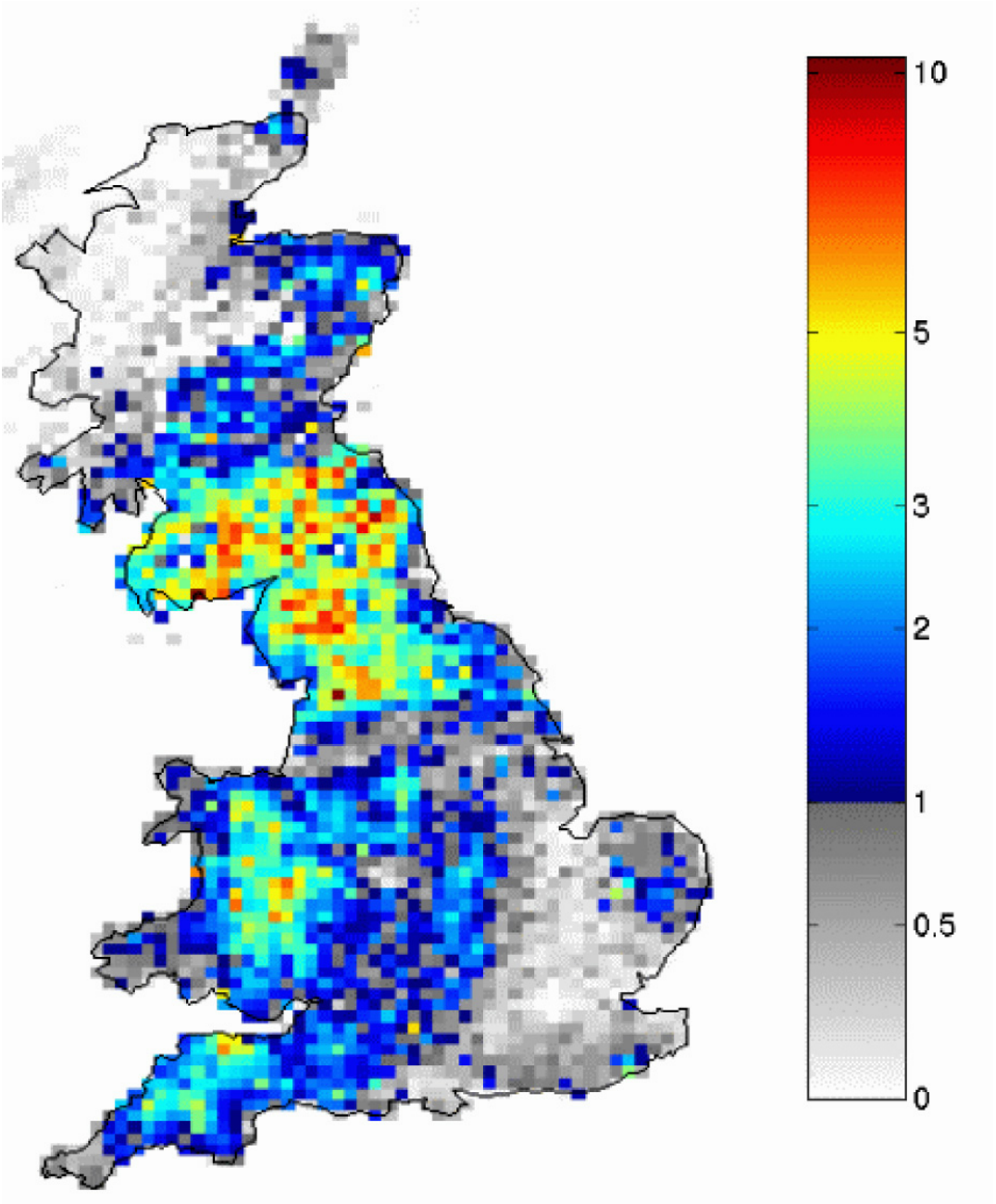


Figure 7.16: Risk map of FMD from Keeling et al. (2001). The scale is R_0 and is based upon the stochastic model of FMD in Keeling et al. (2001). The model calculates an R_0 value for each holding and the average of these values for 10km squares is shown.

The risk maps developed in this paper are comparable to those of Pfeiffer et al. (2007) in that they use predicted values from a logistic regression model to generate the surface. However, the model of Pfeiffer et al. (2007) of HPAI risk in Vietnam is based upon predictors which are much more spatially correlated than the holding specific predictors which form several of the predictors for the FMD model. Furthermore, the data are on the commune level rather than individual holdings, so there is no need to generate a continuous raster surface as was done in this chapter.

In conclusion, a method of mapping FMD risk has been developed and applied in different contexts; retrospectively to the 2001 epidemic, in a static non-epidemic context, to a start of epidemic situation (Surrey 2007) and to guide resource allocation for surveillance given the potential for introduction of the virus into Scotland during the 2007 FMD outbreak. This demonstrates that these methods are valuable and easily applicable in the event of future FMD outbreaks and can be used to target resources and evaluate the potential size of an epidemic.

Chapter 8

General Discussion

This thesis has explored the epidemiology of the 2001 FMD epidemic with particular reference to the underlying distributions of farms and livestock. The risk factors identified have been used to develop models of FMD susceptibility, transmission, and an understanding of how landscape features influence transmission. The three models were combined to create risk surfaces of FMD for the UK in order to understand the spatial variations in FMD risk and incidence. These risk surfaces show that many of the most at-risk areas with the greatest densities of susceptible animals in the UK were areas which experienced large numbers of cases of FMD during the 2001 epidemic. By contrast, the 2007 FMD outbreak was introduced into and contained in a low risk area with low probabilities of local spread. A third way in which these risk surfaces have been employed is through investigation of potential virus introduction into an apparently FMD free area, in this case Scotland in 2007. This demonstrated that had the virus been introduced from one of 45 indirect animal movements from the affected area in Surrey there was great potential for a large epidemic. The maps were valuable for prioritising surveillance in Scotland because those potential introductions which presented the greatest likelihood of local spread could be investigated first.

The thesis was a three stage process involving collating, cleaning and conducting accuracy assessment of the data, model construction and finally combining these to map risk. These three stages will be discussed in turn before considering how this research can be used to inform decision makers in the event of a future epidemic as well as how the research might be taken forward.

8.1 Data

The 2001 FMD dataset represents one of the most comprehensive veterinary epidemiological datasets gathered to date, particularly for an epidemic of this size. The FMD dataset includes the results of detailed epidemiological evaluation of each infected premises. Knowledge of each case, information on the timing and source of infection are usually only available from controlled experiments, rarely has such data been produced in the field.

However, as was discussed in Chapter 2 and elsewhere (McLaws et al., 2006) there remain limitations with these data. This is because thorough scientific investigation should always seek to push the limits of what the data can tell us, and the FMD data is no exception. It is essential to understand and investigate the limitations of the data in order to be able to guide the research and interpret the results.

One of the principal limitations of these data is that there is ambiguity into what is actually a case farm (IP). The 'gold standard' definition of an IP would be a farm holding on which one or more animal(s) have FMDV antigen or antibodies to FMDV. As discussed in Chapter 2 the presence of antibody and antigen can be identified in two ways:

1. Identifying animals which are showing clinical signs of disease.

2. Laboratory identification of antibody or antigen.

This definition of a case is a major simplification. A single case could comprise a holding with one hundred animals on which all the stock has FMD, a holding of a hundred animals on which one animal is infected or a holding of one animal which is infected. The disease burden for the first example is markedly greater than for either the second or the third farm, but all three holdings represent one case each. However, the farm with one hundred infected animals has greater potential to transmit infection and in this respect farm size and the number of infected animals could be regarded in the same way as super shedding cattle for *Escherichia coli* are considered (Matthews et al., 2006). Although animal numbers on the holding are often considered when evaluating the potential of a holding to transmit infection, this still does not capture

the within-farm dynamics and therefore does not allow for the number of infected animals on the farm.

In spite of these limitations, modeling a case as any holding with one or more animals with FMD is sensible. A farm holding is a meaningful epidemiological unit. Although there may be epidemiological sub-units on that holding other animals on that holding are at greater risk of being infected than animals on different holdings. Consider one hundred infected cattle, if these cattle are on one holding that is one case, if they are spread over 100 holdings that represents 100 cases. Although the difference may not be 100-fold, the epidemiological burden and potential to spread FMD is undoubtedly greater in the latter instance. For epidemiological analyses what is required is detailed on-farm epidemiology describing the numbers and locations of infected stock. This is unfortunately lacking in the FMD data.

Further issues arise as a result of the way in which cases are identified. Both methods- clinical and laboratory diagnosis have problems. Firstly, not all animals display distinctive signs of FMD, or animals may have recovered. A particular problem in this respect are sheep as sheep display at best mild clinical signs which are easily confused with other common ailments such as foot rot and orf (Kitching and Hughes, 2002). Laboratory diagnosis is dependent upon the timely and adequate sampling of animals (National Audit Office, 2002), which may be difficult if animals are not showing signs of disease. Furthermore, not all IPs were tested in the laboratory (Ferris et al., 2006): 1,320 were positive in the laboratory, 396 negative and 310 untested.

In addition to overreporting there was underreporting of the number of infected holdings as a result of the extensive culling of apparently uninfected farms (National Audit Office, 2002; Woolhouse et al., 2001). Animals on a proportion of these holdings will have been infected and not detected because they were pre-clinical or because there were not the veterinary resources to inspect all animals prior to culling. However, there have been no empirical studies which attempt to identify the likely numbers of culled premises which were infected. As a result, the number of farms actually infected with FMDV is the 1,320 IPs which tested positive in the laboratory, plus a proportion of the 706 IPs which tested negative or were untested, plus

some proportion of the 8,820 non-IP culls. As both proportions in this equation are unknown the outcome in this thesis and indeed all papers which use data on IPs is whether a holding is declared an IP rather than whether it is confirmed as being infected with FMDV. There should be no seropositive animals on farms which were not culled as such farms would have been detected in post-epidemic serosurveillance.

Therefore, the final pattern of IPs is heavily contingent upon the policies and decisions regarding resources which were implemented at the time. The final pattern (spatial and temporal) of cases is entirely dependent upon these decisions, particularly decisions regarding epidemic control and where the seeds are distributed. If the epidemic were being run as a laboratory experiment the experiment would be set up with a control (an epidemic in which no controls were implemented) and several parallel epidemics incorporating different control interventions which would be implemented uniformly in space in time. As a result of this the risk maps were based on just phase 1 of the epidemic as this represents the period of the epidemic in which control by culling was at its minimum and was subject to the least spatial variation. However, all analyses must take these control policies into account when interpreting their results.

Analysis was conducted on the variables which are estimated for IPs; namely the date of infection and the most probable source of infection (Chapter 2). Both are based upon veterinary judgement and experience although it has been subsequently demonstrated that the latter can be established with certainty using sequence data (Cottam et al., 2006), although this was only conducted for a subset of 22 IPs. Instead, Chapter 2 evaluated the source of infection data with respect to the estimated dates of infection and dates of slaughter of source and daughter farm. There were a large number of inconsistencies which had been introduced, either in the estimated infection dates or the sources, removing these inconsistent sources reduced the number of holdings with an identified source of infection from 1,425 to 953. This suggests that there is potential for improvement in the identification of the infection date and infection source in the event of future epidemics.

Chapter 3 described the farm inventory of the UK which is one of the most complete in the world and is regularly updated. For this study the data was mainly

derived from the UK agricultural census but also from the UK farm list. The farm list comprises a full inventory of farm holdings in the UK with address data and an accurate (usually within a few metres) georeference for the main farm building. From this list a census of the numbers of animals on holdings is taken and in 2000 (the year of interest) a full census was taken. The UK is one of few countries which can distribute such complete farm level data to researchers.

The data gathered on holdings culled during the 2001 FMD epidemic is in a similar format to the agricultural census. However, there is one major difference which is the definition of a farm holding. The census defines a farm holding as a single farm enterprise irrespective of how its livestock may be distributed. As animals on fields more than 1km away from the main farm holding form a separate epidemiological unit this was defined as a separate farm holding during the FMD epidemic (Mansley et al., 2003). Although these holdings could be identified in the farm list data they could not be reliably populated with animals. The only way these problems could be overcome would be to maintain a full register of all land parcels and ensure that farmers update the register every time animals are moved between parcels of land, however implementation of this would be difficult.

Further differences arise as a result of the timing of the FMD epidemic with respect to the collection of the census data. This is particularly a problem with respect to the national sheep flock whose numbers are subject to an annual cycle with a peak in spring immediately after lambing which is followed by a gradual decline and a trough immediately prior to lambing in early March (section 3.2.3). Therefore the June 2000 census represents a time when the sheep flock was near its peak and the FMD epidemic started at a time when the national flock was near its minimum. In this respect the representativeness of the census data should not be considered in terms of how recently it was taken, rather, whether it was taken at the same time of the year as the epidemic. This problem was observed when animal numbers on the FMD cull data were compared to numbers on the census (section 3.5, particularly Table 3.6). Sheep numbers were substantially lower when farms were culled than they reported on the census, this is particularly the case in Wales, something also observed by Keeling et al. (2001) and born out in the holding

susceptibility model (chapter 4) where Wales is a highly significant protective effect. Wales is very distinctive due to the amount of hill sheep farming practiced in Wales. Hill sheep farming is even more cyclical (section 3.2.3) as it is only practiced in the summer. Farms rear lambs on the hills in the summer and in September or October they are moved to slaughter or other lowland premises and only the adult animals are retained in lowland fields. Given that the average number of lambs born per pregnant ewe is 1.1 (section 3.2.3) this explains why the sheep population in Wales was 41% lower during the FMD epidemic than in June 2000 (Table 3.6). The same would be true for other areas of the country, notably the Highlands of Scotland, however this is not seen in the data as there was no FMD in these areas. As a result of this the models of Keeling et al. (2001) over estimate the risk of disease spread in central Wales which the susceptibility model in this thesis does not as it has the country parameter to allow for this. However, other areas of hill sheep farming where there was no FMD are not allowed for. Furthermore, if the model was applied to an epidemic in Wales in June it would underestimate the risk as then the sheep numbers on the census are accurate.

The spatial locations of farm holdings was derived from the address of the main farm holding. Generally, this is a representative point for the farm holding and in a study from Cornwall this was found to be the most representative location to georeference (Durr and Froggatt, 2002). However, there is an exception to this which was identified in the risk mapping chapter. This was the Outer Hebrides in which the farms were registered to addresses around the coast whilst the farms actually kept animals on higher ground inland. Therefore, whilst the farm holding may be the most optimal location it should be born in mind that this is not always the case and there can be exceptions.

In spite of these limitations this thesis has compiled the most complete demographic and disease datasets that have been compiled for any non-human epidemic to date. It has built on spatial epidemiological analyses of FMD from other countries in which the data is aggregated to some areal unit such as the county (Rivas et al., 2003; Ward and Perez, 2004). Full knowledge of the individual locations and size of the entire population at risk coupled with similar data for disease incidence as well as

detailed epidemiology of the incidences is largely unprecedented. This demographic dataset coupled with the FMD data generated by this epidemic provided the basis for further developments in spatial epidemiological tools described in this thesis.

8.2 FMD spread

The extended period between the introduction of the disease into the country and the confirmation of the first case coupled with the extensive movements of infected animals during this period resulted in the extensive seeding of the disease throughout the UK (Anderson, 2002; Gibbens et al., 2001; Haydon et al., 2003). The final distribution of IPs shows that some of these seedings resulted in extensive epidemics in the locality, whilst others did not (Figure 1.7). IPs are clustered around these seeds and therefore spread is determined (in part) by Euclidean distance from a seed (chapter 4). A large part of the reason that the 2001 epidemic was so large was that the disease was seeded so widely throughout the country, creating a large number of foci from which the disease could spread.

After Euclidean distance to a seed has been taken into account it is still possible to accurately derive a measure of holding level susceptibility using farm level factors and in particular the density of animals in the locality of the holding (chapter 4). Animal densities are of particular importance as they represent the potential infectious challenge to a farm. In these analyses cattle densities are a more important risk factor than sheep densities. This is likely to represent the greater susceptibility and ability of cattle to transmit FMD (Donaldson et al., 2001).

The model of transmission also shows that cattle are more important transmitters than sheep (Chapter 5). Many sheep holdings were infected during the epidemic and are therefore susceptible to infection, but sheep holdings do not play a major part in virus transmission (Chapter 5). This suggests that for post-NMB inter-holding spread, sheep may have been largely a dead end host, becoming infected but not passing on infection. Therefore, it could be argued that the control effort should have focused upon cattle farms rather than sheep holdings. The main apparent danger sheep represent in prolonging the epidemic is of acquiring infection and passing the

disease to cattle on the same holding.

There are regional differences in the FMD epidemic, in both the susceptibility model in which the country of the farm was a significant risk factor and the transmission model in which farms in Cumbria behaved differently to the rest of the UK. These regional effects are likely to be acting as a proxy for some variable which has not been measured explicitly. Country level differences primarily reflect differences in the seeding of the virus and in differences concerning the farm demographics already discussed. In England the virus was seeded in areas with large numbers of animals and therefore high risk (Cumbria, Devon, Cheshire, see Figure 7.7) as well as lower risk areas (Kent, Essex, Northamptonshire), whilst in Scotland and Wales the virus was seeded only in areas at greater risk. The differences noted between Cumbria and the remainder of GB are likely to result from epidemiological factors with the Cumbrian epidemic spanning the entire epidemic which is at least in part due to the extent of seeding in Cumbria as a result of trade through the infected Longtown market. However, this may also result from Cumbria comprising a large continuous area of high FMD risk through which the virus was able to spread during the course of the epidemic. Other areas tended to be smaller areas of high risk and experienced an epidemic of shorter duration. Differences may also be a result of the intensity of the epidemic in Cumbria. At its peak there may not have been the personnel resources available to manage the epidemic effectively (National Audit Office, 2002).

Fine scale analysis of FMD transmission with respect to barriers (Chapter 6) shows that at the level of the IP farm size is a very significant predictor of FMD transmission over and above Euclidean distance. This is an important result because the models of transmission and susceptibility developed in this thesis included animal density as the principal measure of animal numbers. The result at the holding level shows that numbers of animals on the individual farm does actually matter and is a risk factor rather than solely animal densities. However, incorporating these predictors into the statistical models used in this thesis is problematic because numbers of animals on a holding can not easily be treated as a continuous variable (Section). As a result they were treated as species presence/absence and are unable to take size into account above this simple variable.

These fine scale analyses go on to show that geographical features, namely rivers and railways can act to inhibit virus spread. Furthermore, the findings support the conclusions of Savill et al. (2006) by finding that road distance is no more of a risk factor than Euclidean distance (Chapter 6). This is an important result as future FMD epidemics in the UK are likely to be spatially dependent and therefore the correct metric to measure the spatial dependence must be known. These analyses took a rather novel approach to spatial analyses in applying a case-control methodology has rarely been applied to spatial epidemiology. Analysis of barriers to transmission has been identified as an area of spatial epidemiology to be further investigated (Ostfeld et al., 2005).

This thesis has investigated FMD spread from the point of both the 'sending' farm (transmission) and the 'receiving' farm (susceptibility) as well as investigating intervening geographical features. Farm size and animal densities are important determinants for both acquiring and transmitting infection and within this cattle holdings are more important for both. There is also some residual spatial level effect with national level differences in susceptibility and regional level differences concerning Cumbria for transmission (Chapter 5).

8.3 Risk mapping

The data on the locations of holdings andthe models of susceptibility and transmission were combined to generate risk maps for FMD. The risk maps give an accurate description both of the 2001 outbreak, but also applied prospectively to the 2007 epidemic and can accurately predict the likely distribution of IP clusters. These maps are generated at a very fine scale (100m grid cells) which is made possible by the high spatial accuracy of the input data.

Risk maps will always be dependent upon accuracy and assumptions of the underlying models. For these maps, the susceptibility model is a very good fit (area under the ROC=0.91), whilst the transmission model suffers from a rather poorer model fit (area under the ROC=0.71). The transmission model is more limited than the susceptibility model in that it is a subset of farms that became infected with

FMD and therefore are already likely to exhibit the characteristics described in the susceptibility model. It may be the case that within this narrow subset of farms those which will go on to transmit infection are highly stochastic and the risk factors used do not offer the scope to fully identify differences. A further issue with the use of the transmission model to develop risk maps was that it includes a number of IP specific parameters (infected period and laboratory result) which had to be set to the same value for every holding. As a result the map of transmission risk is much more of a homogenous smear than the susceptibility risk maps which show distinct hotspots of risk. The variation which is present on the transmission map is largely the result of differences in cattle densities.

The map of susceptibility risk showed that there were extensive areas of high risk in many areas which experienced FMD epidemics during the 2001 epidemic. Indeed, analysis of the non-epidemic risk map shows that the virus was seeded in many of the high risk areas. This is because the virus is seeded by animal movements and animals typically move to areas with the most animals, which in turn defines high risk. The 2001 risk maps suggest that in many areas but in particular in Cumbria and Dumfries & Galloway the epidemic appears to have exhausted itself. The cases cover the areas at high risk and therefore the epidemic effectively burned itself out.

These methods were used to evaluate the likelihood of extensive spread during the 2007 FMD outbreaks in Surrey and subsequently used to guide surveillance efforts within Scotland (Savill et al., 2007b). This demonstrated that these are valuable tools which can be applied rapidly in the early stages of an epidemic to understand the risk and allocate resources accordingly. Surrey represents a very different epidemiological setting to Cumbria in that Surrey is an area of very low animal densities with typically small farms. As Cumbria experienced the majority of the 2001 epidemic the model was parameterised for this area. However, the model was able to predict the size of the epidemic in the Surrey area albeit it did not predict the overall risk of infection in this area.

These risk maps represent a development of previous models (Keeling et al., 2001). Keeling et al. (2001) use model output to map mean values of R_0 and Lawson and Zhou (2005) map standardised incidence ratios at the parish level. The former is

aggregated at a coarse scale (5km grid cells) and the problems with using mean values of R_0 are discussed in chapter 7. The model of Lawson and Zhou (2005) is constructed at the parish level and the data is aggregated to fortnightly counts, so a large amount of information is immediately lost. This study uses a very fine raster surface (100m cells) and uses the predicted number of IPs within 3km of the cells as the metric which due to heterogeneities in risk is a more robust statistic than the mean number of cases (chapter 7). Such tools should always be used with consideration of the key limitation that they were developed for a unique epidemic which involved a particular strain of the virus. Use of these tools in different epidemics should be with reference to virus differences and local differences in the nature of farming, and the model may require modification under different circumstances. For example if an epidemic was seeded in areas of East Anglia with large pig populations and the potential for considerable viral plumes, these tools would require alteration to allow for the characteristics of FMD in pig farms and the potential for longer range transmission events. Likewise, differences in the nature of farming compared to GB should be considered when applying these techniques to FMD epidemics in different countries.

8.4 Future directions

The datasets resulting from the 2001 FMD epidemic generated an enormous amount of research. This is largely as a result of the fact that veterinary epidemiology was a field reaching maturity and found itself presented with the largest and most detailed disease and demographic datasets ever available. The opportunity to work on such a dataset has generated a large amount of research using a variety of techniques. However, this epidemic is a one-off and was the result of a large amount of apparently stochastic spread. What is unclear is if the 2001 epidemic was seeded in exactly the same way again how it would behave; however, some conclusions such as the species level factors should be robust. In the epidemic from 2001 it appears that in Cumbria and Dumfries & Galloway the epidemic was not brought under control. Rather, through a process of spread and culling that it burned itself out and the density of

the susceptible animal population became sufficiently low that the epidemic could not sustain itself. In light of this and the observation that sheep appear to be a near dead-end host, control strategies could be reevaluated such that the cattle population is vaccinated promptly using the methodology outlined by Tildesley et al. (2006), the sheep population should be monitored for infection and to ensure that they are not transmitting infection in future epidemics.

From the FMD epidemic of 2001 and the outbreak in 2007 the importance of virus seeding in determining the final size of the epidemic can be seen. When the virus is seeded by animal movements the areas at greatest risk appear to become infected, this is as a result of animals moving to areas with a lot of animals. This relationship should be investigated more closely particularly to identify whether there are any high risk areas which appear to receive excessively high numbers of animal movements and are therefore at elevated risk. The 2007 outbreak was contained because it was introduced into an area of low risk and there were no out movements of infected stock.

Given a future epidemic more detailed information on the on-farm epidemiology of the disease such as gene sequencing to reconstruct the pattern of transmission would allow more accurate analysis of transmission. A second major further development would be detailed analysis of the pattern of on-farm spread. Understanding of within-farm epidemiology would also assist in modeling transmission as the actual number of infectious animals could be established and the potential for FMD super-shedders further investigated.

Whilst data on animal demographics in the UK is very good there is still room for improvement. This is particularly important given that scares surrounding epidemics of livestock are becoming increasingly frequent, with introductions of FMD, bluetongue (DEFRA, 2007b) and HPAI into the UK during the first 10 months of 2007. To facilitate management of these epidemics an accurate and up to date knowledge of precisely where animals are with accurate and accessible records of movements is essential to understanding the threat these diseases can pose to the country. The agricultural census is at best adequate, however it has limited the scope of this and previous studies (Keeling et al., 2001; Matthews et al., 2003; Morris et al., 2001; Savill

et al., 2006) and both a more regularly updated and higher resolution dataset would be of enormous assistance to scientists and policy makers.

Bibliography

D. Abrial, D. Calavas, N. Lauvergne, E. Morignat, and C. Ducrot. Descriptive spatial analysis of bse in western france. *Veterinary Research*, 34(6):749–760, 2003.

N. Aggarwal, Z. Zhang, S. Cox, R. Statham, S. Alexandersen, R. P. Kitching, and P. V. Barnett. Experimental studies with foot-and-mouth disease virus, strain O, responsible for the 2001 epidemic in the United Kingdom. *Vaccine*, 20(19-20): 2508–2515, Jun 2002.

S. Alexandersen, Z. D. Zhang, and A. I. Donaldson. Aspects of the persistence of foot-and-mouth disease virus in animals - the carrier problem. *Microbes and Infection*, 4(10):1099–1110, 2002.

S. Alexandersen, R. P. Kitching, L. M. Mansley, and A. I. Donaldson. Clinical and laboratory investigations of five outbreaks of foot-and-mouth disease during the 2001 epidemic in the United Kingdom. *Vet Rec*, 152(16):489–496, Apr 2003a.

S. Alexandersen, Z. Zhang, A. I. Donaldson, and A. J. M. Garland. The pathogenesis and diagnosis of foot-and-mouth disease. *Journal of Comparative Pathology*, 129 (1):1–36, 2003b.

B.F. Allan, F. Keesing, and R.S. Ostfeld. Effect of forest fragmentation on lyme disease risk. *Conservation Biology*, 17(1):267–272, 2003.

D Allen. *Planned Beef Production and marketing*. NSP Professional Books; Oxford, 1990.

I. Anderson. *Foot and Mouth Disease 2001: Lessons to be Learned Inquiry Report*. The Stationary Office: London, 2002.

Anon. Animal welfare problems arising as a result of fmd: short-term solutions. *Veterinary Record*, 148(14):428–429, 2001a.

Anon. Movement and disposal schemes for alleviating welfare problems on farms caused by fmd-related movement restrictions. *Veterinary Record*, 148(14):426–427, 2001b.

Anon. Foot-and-mouth disease: waiting for the peak. *Veterinary Record*, 148(11): 322–+, 2001c.

Anon. Foot-and-mouth disease: the government moves towards limited vaccination. *Veterinary Record*, 148(16):490–490, 2001d.

Anon. Foot-and-mouth disease: the government holds fire on vaccination. *Veterinary Record*, 148(17):518–518, 2001e.

P. Armitage and G. Berry. *Statistical Methods in Medical Research*. Blackwell Scientific, Oxford, 1987.

T. C. Bailey and A. C. Gatrell. *Interactive spatial data analysis*. Longman Scientific & Technical, Harlow, 1995.

P. Barnett, A. J M Garland, R. P. Kitching, and C. G. Schermbrucker. Aspects of emergency vaccination against foot-and-mouth disease. *Comp Immunol Microbiol Infect Dis*, 25(5-6):345–364, Oct 2002.

P. V. Barnett and H. Carabin. A review of emergency foot-and-mouth disease (FMD) vaccines. *Vaccine*, 20(11-12):1505–1514, Feb 2002.

L. M. Bartley, C. A. Donnelly, and R. M. Anderson. Review of foot-and-mouth disease virus survival in animal excretions and on fomites. *Veterinary Record*, 151 (22):667–669, 2002.

G. J. Boender, T. J. Hagenaars, A. Bouma, G. Nodelijk, A. R. W. Elbers, M. C. M. de Jong, and M. van Boven. Risk maps for the spread of highly pathogenic avian influenza in poultry. *PLoS Comput Biol*, 3(4):e71, Apr 2007. doi: 10.1371/journal.pcbi.0030071.

G.J. Boender, G. Nodelijk, T Hagenaars, A Elbers, and M de Jong. Local spread of classical swine fever upon virus introduction into The Netherlands: mapping of areas at high risk. *BMC Vet Res*, 4(1):9, Feb 2008. doi: 10.1186/1746-6148-4-9. URL `http://dx.doi.org/10.1186/1746-6148-4-9`.

A. Bouma, A. R. W. Elbers, A. Dekker, A. de Koeijer, C. Bartels, P. Vellema, P. van der Wal, E. M. A. van Rooij, F. H. Pluimers, and M. C. M. de Jong. The foot-and-mouth disease epidemic in the netherlands in 2001. *Preventive Veterinary Medicine*, 57(3):155–166, 2003.

A. Bouma, A. Dekker, and M. C. M. de Jong. No foot-and-mouth disease virus transmission between individually housed calves. *Veterinary Microbiology*, 98(1): 29–36, 2004.

J. Brownlie. Strategic decisions to evaluate before implementing a vaccine programme in the face of a foot-and-mouth disease (fmd) outbreak. *Veterinary Record*, 148 (12):358–360, 2001.

M.E. Chase-Topping, I.J. McKendrick, M.C. Pearce, P. MacDonald, L. Matthews, J.E.B. Halliday, L. Allison, J. Fenlon, J.C. Low, G. Gunn, and M.E.J. Woolhouse. Risk factors for the presence of high-level shedders of Escherichia coli O157 on Scottish farms. *J Clin Microbiol*, 45(5):1594–1603, May 2007. doi: 10.1128/JCM.01690-06.

I. Chmitelin and F. Moutou. Foot and mouth disease: lessons to be learned from the experience of France. *Rev Sci Tech*, 21(3):731–7, 723–30, Dec 2002.

A. C. A. Clements, D. U. Pfeiffer, M. J. Otte, K. Morteo, and L. Chen. A global livestock production and health atlas (glipha) for interactive presentation, integration and analysis of livestock data. *Preventive Veterinary Medicine*, 56(1):19–32, 2002.

A.C.A. Clements, D.U. Pfeiffer, V. Martin, and M.J. Otte. A Rift Valley fever atlas for Africa. *Prev Vet Med*, 82(1-2):72–82, Nov 2007. doi: 10.1016/j.prevetmed.2007.05.006.

J. A. Costelloe, M. C. Gaynor, S. Gaynor, W. J. McAteer, and P. J. O'Reilly. Control of foot and mouth disease: lessons from the experience of Ireland. *Rev Sci Tech*, 21(3):739–750, Dec 2002.

E.M. Cottam, D.T. Haydon, D.J. Paton, J. Gloster, J.W. Wilesmith, N.P. Ferris, G.H. Hutchings, and D.P. King. Molecular epidemiology of the foot-and-mouth disease virus outbreak in the United Kingdom in 2001. *J Virol*, 80(22):11274–11282, Nov 2006. doi: 10.1128/JVI.01236-06.

E.M. Cottam, J. Wadsworth, A.E. Shaw, R.J. Rowlands, L. Goatley, S. Maan, N.S. Maan, P.P.C. Mertens, K. Ebert, Y. Li, E.D. Ryan, N. Juleff, N.P. Ferris, J.W. Wilesmith, D.T. Haydon, D.P. King, D.J. Paton, and N.J. Knowles. Transmission pathways of foot-and-mouth disease virus in the United Kingdom in 2007. *PLoS Pathog*, 4(4):e1000050, Apr 2008. doi: 10.1371/journal.ppat.1000050. URL `http://dx.doi.org/10.1371/journal.ppat.1000050`.

M. Crawley. *Statistical computing. An introduction to data analysis using S-Plus*. John Wiley & sons, Chichester, 2002.

G. Davies. Foot-and-mouth disease control measures. *Veterinary Record*, 148(12): 386–387, 2001.

G. Davies. The foot and mouth disease (fmd) epidemic in the united kingdom 2001. *Comparative Immunology Microbiology and Infectious Diseases*, 25(5-6):331–343, 2002a.

G. Davies. Foot and mouth disease. *Research in Veterinary Science*, 73(3):195–199, 2002b.

H. C. Davison, R. P. Smith, A. R. Sayers, and S. J. Evans. Dairy farm characteristics, including biosecurity, obtained during a cohort study in england and wales. *Cattle Practice*, 11:299–310, 2003.

B. M de C Bronsvoort, V. N. Tanya, R. P. Kitching, C. Nfon, S. M. Hamman, and K. L. Morgan. Foot and mouth disease and livestock husbandry practices in the Adamawa Province of Cameroon. *Trop Anim Health Prod*, 35(6):491–507, Dec 2003.

B. M de C Bronsvoort, C. Nfon, S. M. Hamman, V. N. Tanya, R. P. Kitching, and K. L. Morgan. Risk factors for herdsman-reported foot-and-mouth disease in the Adamawa Province of Cameroon. *Prev Vet Med*, 66(1-4):127–139, Dec 2004. doi: 10.1016/j.prevetmed.2004.09.010.

DEFRA. Animal health and welfare: Fmd data archive, 2007a.

DEFRA. Initial epidemiological report on the outbreak of bluetongue in east anglia and south east england from investigations completed to 19 october 2007. Technical report, DEFRA, 2007b.

DEFRA. Fmd 2007 epidemiology report. situation at 12.00 sunday 30 september 2007. day 58. Technical report, DEFRA, 2007c.

DEFRA. Fmd 2007. summary epidemiology report. situation at 16:00 monday 13th august. day 10 of the outbreak. Technical report, DEFRA, 2007d.

DEFRA. The geo referencing of holdings on the agricultural census, 2008. URL `http://www.defra.gov.uk/esg/work.htm`.

P.J. Diggle. Spatio-temporal point processes, partial likelihood, foot and mouth disease. *Stat Methods Med Res*, 15(4):325–336, Aug 2006.

M. G. Doherr, A. R. Hett, J. Rufenacht, A. Zurbriggen, and D. Heim. Geographical clustering of cases of bovine spongiform encephalopathy (bse) born in switzerland after the feed ban. *Veterinary Record*, 151(16):467–472, 2002.

I. Dohoo, W. Martin, and H. Stryhn. *Veterinary epidemiologic research*. AVC Inc.: Charlottetown, P.E.I., 2003.

E. Domingo, E. Baranowski, C. Escarmis, and F. Sobrino. Foot-and-mouth disease virus. *Comparative Immunology Microbiology and Infectious Diseases*, 25(5-6): 297–308, 2002.

A. Donaldson. Clinical manifestations of fmd. In Esteban Domingo and Fransisco Sobrino, editors, *Foot and Mouth Disease: Current Perspectives*. Horizon Bioscience, Wymondham, 2004.

A. I. Donaldson and S. Alexandersen. Relative resistance of pigs to infection by natural aerosols of FMD virus. *Vet Rec*, 148(19):600–602, May 2001.

A. I. Donaldson and S. Alexandersen. Predicting the spread of foot and mouth disease by airborne virus. *Revue Scientifique Et Technique De L Office International Des Epizooties*, 21(3):569–575, 2002.

A. I. Donaldson, S. Alexandersen, J. H. Sørensen, and T. Mikkelsen. Relative risks of the uncontrollable (airborne) spread of FMD by different species. *Vet Rec*, 148 (19):602–604, May 2001.

P. A. Durr and A. E A Froggatt. How best to geo-reference farms? A case study from Cornwall, England. *Prev Vet Med*, 56(1):51–62, Nov 2002.

EDiNA. Digimap service. URL `www.edina.ac.uk`.

FAO. Programme for the world census of agriculture. Technical report, FAO, 1995.

FAO. The statistics division: Census of agriculture. URL, March 2008. URL `http://www.fao.org/es/ess/census`.

N. M. Ferguson, C. A. Donnelly, and R. M. Anderson. Transmission intensity and impact of control policies on the foot and mouth epidemic in Great Britain. *Nature*, 413(6855):542–548, Oct 2001a. doi: 10.1038/35097116.

N. M. Ferguson, C. A. Donnelly, and R. M. Anderson. The foot-and-mouth epidemic in Great Britain: pattern of spread and impact of interventions. *Science*, 292 (5519):1155–1160, May 2001b. doi: 10.1126/science.1061020.

C. Ferrer-Orta and I. Fita. *Foot and Mouth Disease: Current Perspectives*, chapter Structure of Foot-and-Mouth Disease Virus Particles. Horizon Bioscience, Wymondham, 2004.

N. P. Ferris, D. P. King, S. M. Reid, A. E. Shaw, and G. H. Hutchings. Comparisons of original laboratory results and retrospective analysis by real-time reverse transcriptase-PCR of virological samples collected from confirmed cases of foot-and-mouth disease in the UK in 2001. *Vet Rec*, 159(12):373–378, Sep 2006.

A. Fraser. *Sheep husbandry and Diseases*. Colline; London, 1987.

N. P. French, T. D. Parkin, and K. L. Morgan. A case control study of blowfly strike in lambs. *Vet Rec*, 139(16):384–388, Oct 1996.

A.J. Garland and A. I. Donaldson. Foot and mouth disease. *Surevillance*, 17(4):6–8, 1990.

J. C. Gibbens and J. W. Wilesmith. Temporal and geographical distribution of cases of foot-and-mouth disease during the early weeks of the 2001 epidemic in Great Britain. *Vet Rec*, 151(14):407–412, Oct 2002.

J. C. Gibbens, C. E. Sharpe, J. W. Wilesmith, L. M. Mansley, E. Michalopoulou, J. B. Ryan, and M. Hudson. Descriptive epidemiology of the 2001 foot-and-mouth disease epidemic in Great Britain: the first five months. *Vet Rec*, 149(24):729–743, Dec 2001.

J. Gloster, H. Hewson, D. Mackay, T. Garland, A. Donaldson, I. Mason, and R. Brown. Spread of foot-and-mouth disease from the burning of animal carcases on open pyres. *Vet Rec*, 148(19):585–586, May 2001.

J. Gloster, H. J. Champion, J. H. Srensen, T. Mikkelsen, D. B. Ryall, P. Astrup, S. Alexandersen, and A. I. Donaldson. Airborne transmission of foot-and-mouth disease virus from Burnside Farm, Heddon-on-the-Wall, Northumberland, during the 2001 epidemic in the United Kingdom. *Vet Rec*, 152(17):525–533, Apr 2003.

J. Gloster, H. J. Champion, L. M. Mansley, P. Romero, T. Brough, and A. Ramirez. The 2001 epidemic of foot-and-mouth disease in the United Kingdom: epidemiological and meteorological case studies. *Vet Rec*, 156(25):793–803, Jun 2005.

M. J. Grubman and B. Baxt. Foot-and-mouth disease. *Clinical Microbiology Reviews*, 17(2):465–+, 2004.

D. Haine, F. Boelaert, D. U. Pfeiffer, C. Saegerman, J-F. Lonneux, B. Losson, and K. Mintiens. Herd-level seroprevalence and risk-mapping of bovine hypodermosis in Belgian cattle herds. *Prev Vet Med*, 65(1-2):93–104, Aug 2004. doi: 10.1016/j.prevetmed.2004.06.005.

J.E.B. Halliday, M.E. Chase-Topping, M.C. Pearce, I.J. McKendrick, L Allison, D Fenlon, C. Low, D.J. Mellor, G.J. Gunn, and M.E.J. Woolhouse. Herd-level risk factors associated with the presence of Phage type 21/28 E. coli O157 on Scottish cattle farms. *BMC Microbiol*, 6:99, 2006. doi: 10.1186/1471-2180-6-99.

C. Hamblin, I. T. R. Barnett, and J. R. Crowther. A new enzyme-linked-immunosorbent-assay (elisa) for the detection of antibodies against foot-and-mouth-disease virus .2. application. *Journal of Immunological Methods*, 93(1): 123–129, 1986a.

C. Hamblin, I. T. R. Barnett, and R. S. Hedger. A new enzyme-linked-immunosorbent-assay (elisa) for the detection of antibodies against foot-and-mouth-disease virus .1. development and method of elisa. *Journal of Immunological Methods*, 93(1):115–121, 1986b.

F.E. Harrell. *Hmisc: Harrell Miscellaneous*, 2007. R package version 3.4-1.

S.I. Hay and R.W. Snow. The malaria Atlas Project: developing global maps of malaria risk. *PLoS Med*, 3(12):e473, Dec 2006. doi: 10.1371/journal.pmed.0030473.

D. T. Haydon, M. E. Woolhouse, and R. P. Kitching. An analysis of foot-and-mouth-disease epidemics in the UK. *IMA J Math Appl Med Biol*, 14(1):1–9, Mar 1997.

D. T. Haydon, M. Chase-Topping, D. J. Shaw, L. Matthews, J. K. Friar, J. Wilesmith, and M. E J Woolhouse. The construction and analysis of epidemic trees with reference to the 2001 UK foot-and-mouth outbreak. *Proc Biol Sci*, 270(1511): 121–127, Jan 2003. doi: 10.1098/rspb.2002.2191.

D.T. Haydon, R.R. Kao, and R.P. Kitching. The UK foot-and-mouth disease outbreak - the aftermath. *Nat Rev Microbiol*, 2(8):675–681, Aug 2004. doi: 10.1038/nrmicro960.

N. Honhold and N.M. Taylor. Data quality assessment: comparison of recorded and contemporary data for farm premises and stock numbers in cumbria, 2001. In *Proceedings of the Society for Veterinary Epidemiology and Preventive Medicine*, 2006.

N. Honhold, N. M. Taylor, L. M. Mansley, and A. D. Paterson. Relationship of speed of slaughter on infected premises and intensity of culling of other premises to the rate of spread of the foot-and-mouth disease epidemic in Great Britain, 2001. *Vet Rec*, 155(10):287–294, Sep 2004a.

N. Honhold, N. M. Taylor, A. Wingfield, P. Einshoj, C. Middlemiss, L. Eppink, R. Wroth, and L. M. Mansley. 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*, 155(12):349–355, Sep 2004b.

D. W. Hosmer and S. Lemeshow. *Applied Logistic Regression*. New York ; Chicheter : Wiley c2000., 2000.

T. Hothorn, F. Bretz, and P with contributions by Richard M. Heiberger Westfall. *multcomp: Simultaneous Inference for General Linear Hypotheses*, 2007.

S. C. Howard and C. A. Donnelly. The importance of immediate destruction in epidemics of foot and mouth disease. *Research in Veterinary Science*, 69(2):189–196, 2000.

G. J. Hughes, V. Mioulet, R. P. Kitching, M. E J Woolhouse, S. Alexandersen, and A. I. Donaldson. Foot-and-mouth disease virus infection of sheep: implications for diagnosis and control. *Vet Rec*, 150(23):724–727, Jun 2002.

J. Jenness. *Surface Tools (surf_tools.avx) extension for ArcView 3.x, v. 1.6*. Jenness Enterprises, 2005.

R. Jones, L. Kelly, N. French, T. England, C. Livesey, and M. Wooldridge. Quantitative estimates of the risk of new outbreaks of foot- and-mouth disease as a result of burning pyres (vol 154, pg 161, 2004). *Veterinary Record*, 154(10):304–304, 2004.

R. R. Kao. The role of mathematical modelling in the control of the 2001 fmd epidemic in the uk. *Trends in Microbiology*, 10(6):279–286, 2002.

R. R. Kao. The impact of local heterogeneity on alternative control strategies for foot-and-mouth disease. *Proceedings of the Royal Society of London Series B-Biological Sciences*, 270(1533):2557–2564, 2003.

M. J. Keeling. Models of foot-and-mouth disease. *Proc Biol Sci*, 272(1569):1195–1202, Jun 2005. doi: 10.1098/rspb.2004.3046.

M. J. Keeling, M. E. Woolhouse, D. J. Shaw, L. Matthews, M. Chase-Topping, D. T. Haydon, S. J. Cornell, J. Kappey, J. Wilesmith, and B. T. Grenfell. Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science*, 294(5543):813–817, Oct 2001. doi: 10.1126/science.1065973.

M. J. Keeling, M. E J Woolhouse, R. M. May, G. Davies, and B. T. Grenfell. Modelling vaccination strategies against foot-and-mouth disease. *Nature*, 421(6919): 136–142, Jan 2003. doi: 10.1038/nature01343.

R. P. Kitching. Clinical variation in foot and mouth disease: cattle. *Rev Sci Tech*, 21(3):499–504, Dec 2002.

R. P. Kitching and S. Alexandersen. Clinical variation in foot and mouth disease: pigs. *Rev Sci Tech*, 21(3):513–518, Dec 2002.

R. P. Kitching and G. J. Hughes. Clinical variation in foot and mouth disease: sheep and goats. *Rev Sci Tech*, 21(3):505–512, Dec 2002.

R.P. Kitching, J. Hammond, M. Jeggo, B Charleston, D. Paton, L. Rodriguez, and R. Heckert. Global FMD control–is it an option? *Vaccine*, 25(30):5660–5664, Jul 2007. doi: 10.1016/j.vaccine.2006.10.052.

N. J. Knowles and A. R. Samuel. Molecular epidemiology of foot-and-mouth disease virus. *Virus Res*, 91(1):65–80, Jan 2003.

V. Latora, A. Nyamba, J Simpore, B Sylvette, S. Diane, B. Sylvre, and S. Musumeci. Network of sexual contacts and sexually transmitted HIV infection in Burkina Faso. *J Med Virol*, 78(6):724–729, Jun 2006. doi: 10.1002/jmv.20614.

A. Lawson, A. Biggeri, and E. Viel J-F. Bohning, D. Lesaffre. *Disease Mapping and Risk Assessment for Public Health*, chapter Disease Mapping and Its Uses, pages 3–14. John Wiley and Sons, Chichester, 1999.

A. B. Lawson and H. Zhou. Spatial statistical modeling of disease outbreaks with particular reference to the UK foot and mouth disease (FMD) epidemic of 2001. *Prev Vet Med*, 71(3-4):141–156, Oct 2005. doi: 10.1016/j.prevetmed.2005.07.002.

A.B. Lawson, W.J. Browne, and C.J. Vidal Rodeiro. *Disease Mapping with WinBUGS and MLwiN*. John Wiley /& Sons, Chichester, 2003.

I. Lowles, R. Hill, V. Auld, H. Stewart, and C. Colhoun. Monitoring the pollution from a pyre used to destroy animal carcasses during the outbreak of foot and mouth disease in cumbria, united kingdom. *Atmospheric Environment*, 36(17):2901–2905, 2002.

Z. Lu, Y. Cao, J Guo, S. Qi, D. Li, Q. Zhang, J. Ma, H. Chang, Z. Liu, X. Liu, and Q. Xie. Development and validation of a 3ABC indirect ELISA for differentiation of foot-and-mouth disease virus infected from vaccinated animals. *Vet Microbiol*, May 2007. doi: 10.1016/j.vetmic.2007.05.017.

D. K. J. Mackay, A. N. Bulut, T. Rendle, F. Davidson, and N. P. Ferris. A solid-phase competition elisa for measuring antibody to foot- and-mouth disease virus. *Journal of Virological Methods*, 97(1-2):33–48, 2001.

MAFF. *The BSE enquiry.* HMSO, London, 2000.

L. M. Mansley, P. J. Dunlop, S. M. Whiteside, and R. G H Smith. Early dissemination of foot-and-mouth disease virus through sheep marketing in February 2001. *Vet Rec*, 153(2):43–50, Jul 2003.

L. Matthews, D. T. Haydon, D. J. Shaw, M. E. Chase-Topping, M. J. Keeling, and M. E J Woolhouse. Neighbourhood control policies and the spread of infectious diseases. *Proc Biol Sci*, 270(1525):1659–1666, Aug 2003. doi: 10.1098/rspb.2003.2429.

L. Matthews, I. J. McKendrick, H. Ternent, G. J. Gunn, B. Synge, and M. E J Woolhouse. Super-shedding cattle and the transmission dynamics of Escherichia coli O157. *Epidemiol Infect*, 134(1):131–142, Feb 2006. doi: 10.1017/S0950268805004590.

M. McLaws, C. Ribble, W. Martin, and C. Stephen. Factors associated with the clinical diagnosis of foot and mouth disease during the 2001 epidemic in the UK. *Prev Vet Med*, 77(1-2):65–81, Nov 2006. doi: 10.1016/j.prevetmed.2006.06.003.

E. Morignat, C. Ducrot, R. Roy, T. Baron, J. L. Vinard, A. G. Biacabe, J. Y. Madec, A. Bencsik, S. Debeer, M. Eliazsewicz, and D. Calavas. Targeted surveillance to assess the prevalence of bse in high- risk populations in western france and the associated risk factors. *Veterinary Record*, 151(3):73–77, 2002.

R. S. Morris, J. W. Wilesmith, M. W. Stern, R. L. Sanson, and M. A. Stevenson. Predictive spatial modelling of alternative control strategies for the foot-and-mouth disease epidemic in great britain, 2001. *Veterinary Record*, 149(5):137–+, 2001.

R. S. Morris, R. L. Sanson, M. W. Stern, M. Stevenson, and J. W. Wilesmith. Decision-support tools for foot and mouth disease control. *Revue Scientifique Et Technique De L Office International Des Epizooties*, 21(3):557–567, 2002.

National Audit Office. The 2001 oubreak of foot and mouth disease. Technical report, National Audit Office, 2002.

Northumberland. *Report of the Committee of Inquiry on Foot and Mouth Disease, 1968: part two.* HMSO, 1968.

M. Odiit, P. R. Bessell, E. M. Fevre, T. Robinson, J. Kinoti, P. G. Coleman, S. C. Welburn, J. McDermott, and M. E. Woolhouse. Using remote sensing and geographic information systems to identify villages at high risk for rhodesiense sleeping sickness in uganda. *Trans R Soc Trop Med Hyg*, 100(4):354–62, 2006.

OIE. *Manual of Standards for Diagnostic Tests and Vaccines.* Office Internalionaldes Epizootes, Paris, 2000.

S. Openshaw. Methods for investigating localized clustering of disease. tests of clustering based on pattern-recognition procedures. *IARC Sci Publ*, (135):203–206, 1996.

R. S. Ostfeld, G. E. Glass, and F. Keesing. Spatial epidemiology: an emerging (or re-emerging) discipline. *Trends in Ecology & Evolution*, 20(6):328–336, 2005.

D. O'Sullivan and D. J. Unwin. *Geographic information analysis.* J. Wiley: Hoboken, N.J., 2003.

G. A. Paiba, J. Anderson, D. J. Paton, A. W. Soldan, S. Alexandersen, M. Corteyn, G. Wilsden, P. Hamblin, D. K J MacKay, and A. I. Donaldson. Validation of a foot-and-mouth disease antibody screening solid-phase competition ELISA (SPCE). *J Virol Methods*, 115(2):145–158, Feb 2004.

E.N. Pavlovsky. *Natural nidality of transmissible diseases : with special reference to the landscape epidemiology of zooanthroponoses.* University of Illinois Press, Urbana, 1966.

A. M. Perez, M. P. Ward, P. Torres, and V. Ritacco. Use of spatial statistics and monitoring data to identify clustering of bovine tuberculosis in argentina. *Preventive Veterinary Medicine*, 56(1):63–74, 2002.

A.M. Perez, M.C. Thurmond, and T.E. Carpenter. Control of a foot-and-mouth disease epidemic in Argentina. *Prev Vet Med*, 65(3-4):217–226, Oct 2004. doi: 10.1016/j.prevetmed.2004.08.002.

A.M. Perez, M.C. Thurmond, and T.E. Carpenter. Spatial distribution of foot-and-mouth disease in Pakistan estimated using imperfect data. *Prev Vet Med*, 76(3-4): 280–289, Oct 2006. doi: 10.1016/j.prevetmed.2006.05.013.

D.U. Pfeiffer, P.Q. Minh, V. Martin, M. Epprecht, and M.J. Otte. An analysis of the spatial and temporal patterns of highly pathogenic avian influenza occurrence in Vietnam using national surveillance data. *Vet J*, 174(2):302–309, Sep 2007. doi: 10.1016/j.tvjl.2007.05.010.

A. Picado, F. J. Guitian, and D. U. Pfeiffer. Space-time interaction as an indicator of local spread during the 2001 FMD outbreak in the UK. *Prev Vet Med*, 79(1): 3–19, Apr 2007. doi: 10.1016/j.prevetmed.2006.11.009.

F. H. Pluimers, A. M. Akkerman, P. van der Wal, A. Dekker, and A. Bianchi. Lessons from the foot and mouth disease outbreak in The Netherlands in 2001. *Rev Sci Tech*, 21(3):711–721, Dec 2002.

R Development Core Team. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria, 2004.

A. L. Rivas, S. D. Smith, P. J. Sullivan, B. Gardner, J. P. Aparicio, A. L. Hoogesteijn, and C. Castillo-Chavez. Identification of geographic factors associated with early spread of foot-and-mouth disease. *American Journal of Veterinary Research*, 64 (12):1519–1527, 2003.

A. L. Rivas, S. J. Schwager, S. Smith, and A. Magri. Early and cost-effective identification of high risk/priority control areas in foot-and-mouth disease epidemics. *Journal of Veterinary Medicine Series B-Infectious Diseases and Veterinary Public Health*, 51(6):263–271, 2004.

E. Ryan, J. Gloster, S. M. Reid, Y. Li, N. P. Ferris, R. Waters, N. Juleff, B. Charleston, B. Bankowski, S. Gubbins, J. W. Wilesmith, D. P. King, and D. J. Paton. Clinical and laboratory investigations of the outbreaks of foot-and-mouth disease in southern england in 2007. *Vet Rec*, 163(5):139–147, Aug 2008.

R. Sanson and A. Pearson. Agribase - a national spatial farm database. In *Eighth IVSEE*, volume 2, pages 12.12.1–12.16.3, Paris, 1997.

R. L. Sanson, R. S. Morris, and M. W. Stern. Epiman-fmd: a decision support system for managing epidemics of vesicular disease. *Revue Scientifique Et Technique De L Office International Des Epizooties*, 18(3):593–605, 1999.

N. J. Savill, D. J. Shaw, R. Deardon, M. J. Tildesley, M. J. Keeling, M. E. J Woolhouse, S. P. Brooks, and B. T. Grenfell. Topographic determinants of foot and mouth disease transmission in the UK 2001 epidemic. *BMC Vet Res*, 2:3, 2006. doi: 10.1186/1746-6148-2-3.

N. J. Savill, D. J. Shaw, R. Deardon, M. J. Tildesley, M. J. Keeling, M. E. J. Woolhouse, S. P. Brooks, and B. T. Grenfell. Effect of data quality on estimates of farm infectiousness trends in the UK 2001 foot-and-mouth disease epidemic. *J R Soc Interface*, 4(13):235–241, Apr 2007a. doi: 10.1098/rsif.2006.0178.

N.J. Savill, V. Volkova, and P.R. Bessell. Report on risk assessment of introduction and maintenance on scottish agricultural holdings of foot and mouth disease (fmd) virus responsible for the fmd outbreak in england declared on 3 august 2007. Technical report, Centre of Excellence in Epidemiology, Population Health & Infectious Disease Control (EPIC) University of Edinburgh, 2007b.

J. M. Scudamore, G. M. Trevelyan, M. V. Tas, E. M. Varley, and G. A. W. Hickman. Carcass disposal: lessons from great britain following the foot and mouth disease outbreaks of 2001. *Revue Scientifique Et Technique De L Office International Des Epizooties*, 21(3):775–787, 2002.

D.L. Smith, B. Lucey, L.A. Waller, J.E. Childs, and L.A. Real. Predicting the spatial dynamics of rabies epidemics on heterogeneous landscapes. *Proc Natl Acad Sci U S A*, 99(6):3668–3672, Mar 2002. doi: 10.1073/pnas.042400799.

J Snow. *Snow on cholera*. Haffner, New York, 1965.

J. H. Sørensen, D. K. Mackay, C. O. Jensen, and A. I. Donaldson. An integrated model to predict the atmospheric spread of foot-and-mouth disease virus. *Epidemiol Infect*, 124(3):577–590, Jun 2000.

M. A. Stevenson, J. W. Wilesmith, J. B. M. Ryan, R. S. Morris, A. B. Lawson, D. U. Pfeiffer, and D. Lin. Descriptive spatial analysis of the epidemic of bovine spongiform encephalopathy in great britain to june 1997. *Veterinary Record*, 147 (14):379–384, 2000.

L Stloukal. Using agricultural census data for demographic purposes: some pros and cons. Technical report, FAO, 1999.

J. Swinton, M.E.J. Woolhouse, M.E. Begon, A.P. Dobson, E. Ferroglio, B.T. Grenfell, V. Guberti, R.S. Hails, J.A.P. Heesterbeek, A. Lavazza, M.G. Roberts, P.J. White, and K. Wilson. *The Ecology of Wildlife Diseases*, chapter Microparasite transmission and persistence, pages 83–101. Oxford Biology, 2002.

A.J. Tatem, A.M. Noor, C. von Hagen, A. Di Gregorio, and S.I. Hay. High resolution population maps for low income nations: combining land cover and census in East Africa. *PLoS ONE*, 2(12):e1298, 2007. doi: 10.1371/journal.pone.0001298.

N. M. Taylor, N. Honhold, A. D. Paterson, and L. M. Mansley. Risk of foot-and-mouth disease associated with proximity in space and time to infected premises and the implications for control policy during the 2001 epidemic in cumbria. *Veterinary Record*, 154(20):617–626, 2004.

The International Statistics Institute. *The Oxford Dictionary of Statistical Terms*. Oxford University Press, Oxford, 2006.

The Royal Society. *Infectious diseases in livestock*. The Royal Society, 2002.

M. Thrusfield, L. Mansley, P. Dunlop, A. Pawson, and J. Taylor. The foot-and-mouth disease epidemic in Dumfries and Galloway, 2001. 2: Serosurveillance, and efficiency and effectiveness of control procedures after the national ban on animal movements. *Vet Rec*, 156(9):269–278, Feb 2005a.

M Thrusfield, L. M. Mansley, P. J. Dunlop, J Taylor, A Pawson, and L Stringer. The foot-and-mouth disease epidemic in dumfries and galloway, 2001. 1: Characteristics and control. *Veterinary Record*, 156:229–252, 2005b.

M.J. Tildesley, N.J. Savill, D.J. Shaw, R. Deardon, S.P. Brooks, M.E.J. Woolhouse, B.T. Grenfell, and M.J. Keeling. Optimal reactive vaccination strategies for a foot-and-mouth outbreak in the UK. *Nature*, 440(7080):83–86, Mar 2006. doi: 10.1038/nature04324.

M.J. Tildesley, R. Deardon, N.J. Savill, P.R. Bessell, S.P. Brookes, M.E.J. Woolhouse, B.T. Grenfell, and M.J. Kelling. Accuracy of models for the 2001 foot-and-mouth epidemic. *Proceedings of the Royal Scoiety Series B*, In press, 2008.

R. Tinline. Lee wave hypothesis for initial pattern of spread during 1967- 68 foot and mouth epizootic. *Nature*, 227(5260):860–&, 1970.

B. Toma, F. Moutou, B. Dufour, and B. Durand. Ring vaccination against foot-and-mouth disease. *Comparative Immunology Microbiology and Infectious Diseases*, 25 (5-6):365–372, 2002.

D.M Tompkins, A.P. Dobson, P. Arneberg, M.E. Begon, I.M. Cattadori, J.V. Greenman, J.A.P. Heesterbeek, P.J. Hudson, D. Newborn, A. Pugliese, A.P. Rizzoli, R. Rosa, F. Rosso, and K. Wilson. Parasites and host population dynamics. In P.J. Hudson, A. Rizzoli, B.T. Grenfell, H. Heesterbeek, and A.P. Dobson, editors, *The Ecology of Wildlife Diseases*. Oxford University Press, New York, 2002.

University of Illinois at Chicago. Snow on cholera: Part 1, 2008. URL `http://www.uic.edu/sph/prepare/courses/chsc400/resources/snowcase1.htm`.

W. Vosloo, A. D S Bastos, O. Sangare, S. K. Hargreaves, and G. R. Thomson. Review of the status and control of foot and mouth disease in sub-Saharan Africa. *Rev Sci Tech*, 21(3):437–449, Dec 2002.

M.P. Ward and A.M. Perez. Herd demographics correlated with the spatial distribution of a foot-and-mouth disease epidemic in Buenos Aires province, Argentina. *Prev Vet Med*, 65(3-4):227–237, Oct 2004. doi: 10.1016/j.prevetmed.2004.08.003.

P. C. L. White and J. K. A. Benhin. Factors influencing the incidence and scale of bovine tuberculosis in cattle in southwest england. *Preventive Veterinary Medicine*, 63(1-2):1–7, 2004.

C. Whittemore. *Whittemore's Science and Practice of Pig Production*, chapter Introduction, pages 1–3. Blackwell; Oxford, 2006.

J. W. Wilesmith, M. A. Stevenson, C. B. King, and R. S. Morris. Spatio-temporal epidemiology of foot-and-mouth disease in two counties of great britain in 2001. *Preventive Veterinary Medicine*, 61(3):157–170, 2003.

M. Woolhouse, M. Chase-Topping, D. Haydon, J. Friar, L. Matthews, G. Hughes, D. Shaw, J. Wilesmith, A. Donaldson, S. Cornell, M. Keeling, and B. Grenfell. Epidemiology. Foot-and-mouth disease under control in the UK. *Nature*, 411(6835): 258–259, May 2001. doi: 10.1038/35077149.

M.F. Worboys. *GIS : a computing perspective*. Taylor & Francis: London, 1995.