# Edinburgh Research Explorer

## Statistical modeling of holding level susceptibility to infection during the 2001 foot and mouth disease epidemic in Great Britain

# Statistical modeling of holding level susceptibility to infection during the 2001 foot and mouth disease epidemic in Great Britain

Paul R. Bessell [a,*], Darren J. Shaw [b], Nicholas J. Savill [a], Mark E.J. Woolhouse [a]

[a] Centre for Infectious Diseases, University of Edinburgh, Kings Buildings, West Mains Road, Edinburgh EH9 3JT, UK
[b] Veterinary Clinical Sciences, R(D)SVS, The University of Edinburgh, Easter Bush Veterinary Centre, Roslin, Midlothian, Scotland

ABSTRACT

Background: An understanding of the factors that determine the risk of members of a susceptible population becoming infected is essential for estimating the potential for disease spread, as opposed to just focusing on transmission from an infected population. Furthermore, analysis of the risk factors can reveal important characteristics of an epidemic and further develop understanding of the processes operating.
Methods: This paper describes the development of a mixed effects logistic regression model of susceptibility of holdings to foot and mouth disease (FMD) during the 2001 epidemic in Great Britain following the imposition of a national ban on the movements of susceptible animals (NMB).
Results: The principal risk factors identified in the model were shorter distances to the nearest infectious seed (a holding infected before the NMB) and the county of the holding (principally Cumbria). Additional risk factors included holdings that are mixed species rather than single species, the surface area of the holding, and the number of cattle within 10 km (all $p < 0.001$), but not surrounding sheep densities ($p > 0.1$). The fit of the model was evaluated using the area under the receiver operator characteristic curve (ROC) and the Hosmer and Lemeshow Chi-squared statistic; the fit was good with both tests (area under the ROC = 0.962 and Hosmer and Lemeshow Chi-squared statistic = 49.98 ($p > 0.1$)).
Conclusions: Holdings at greatest risk of infection can be identified using simple readily available risk factors; this information could be employed in the control of future FMD epidemics.

© 2009 International Society for Infectious Diseases. Published by Elsevier Ltd. All rights reserved.

## Introduction

Foot and mouth disease (FMD) is a highly infectious viral disease affecting cloven-hoofed animals.[1] In the UK there was just one outbreak of FMD between 1968 and 2001.[2] However, in 2001 there was a major epidemic of FMD lasting seven months with infection reported on 2026 premises (known as infected premises, IPs) in Great Britain (GB - comprising mainland England, Scotland and Wales). Within GB a further 8131 dangerous contact (DC) holdings had animals culled under different policies. These premises include 'traditional DCs', contiguous premises (CPs), 3 km culls–sheep premises within 3 km of an IP and slaughter on suspicion (SOS) premises–premises on which infection could not be ruled in or out on clinical grounds.[3,4] With the exception of SOS premises, a number of these non-IPs may have had animals latently infected with FMD, because in many instances freedom from infection was never established through laboratory testing.[4]

Descriptions of the 2001 UK FMD outbreak have been extensively published elsewhere,[3,5–8] but of importance here is that the early stages of the 2001 FMD outbreak were characterized by widespread seeding of infection throughout the country by animal movements.[5,7] A national movement ban (NMB) on susceptible species was implemented on 23 February 2001, after which disease transmission became spatially localized with around 50% of new IPs within 3 km of an infectious IP and around 80% within 10 km.[9] As a result, FMD remained largely restricted to the areas in which it was seeded.[10]

During the period of local spread, the distance to a source of infection appeared to be the principal risk factor for infection.[4,10] However, despite the wealth of literature on the 2001 epidemic, potential risk factors over and above the distance to a source of infection have not been investigated explicitly. Such factors might include the species of animals on the holding, the numbers or densities of animals in the locality, and the area of the holding.

The dynamics of the epidemic during the period of local spread have been mathematically modeled using spatially explicit stochastic[10,11] and deterministic[12,13] mathematical models. The models of Keeling et al.[10] and Tildesley et al.[11] are spatially explicit

simulations in which the next generation of infections is dependant upon the spatial location of the present generation. The predictive accuracy of the model of Tildesley et al.[11] has been evaluated by comparing the predicted IPs with the actual IPs; the resulting accuracy is between 5% and 15%.[14] Given the small number of IPs relative to the number of farm holdings, this represents good accuracy. However, no equivalent statistical evaluation with the potential to directly interrogate the data has been carried out on the dynamics of the epidemic during the period of local spread. Furthermore, there have been no studies that have explicitly attempted to identify risk factors for FMD infection.

This paper, therefore, will develop a statistical model of holding level risk of susceptibility to FMD. By incorporating a range of risk factors in a multivariate model, more subtleties of the epidemic can be determined. The ability of the model to identify IPs will then be compared to the model of Tildesley et al.[11] This will be a means of quantifying the ability of this framework to discriminate between IPs and non-IPs in comparison to previously employed frameworks.

## Materials and methods

### Data

Farm-level data for the 2001 epidemic were taken from the June 2000 agricultural censuses conducted by the Ministry of Agriculture, Fisheries and Food (MAFF–Department for Environment Food and Rural Affairs (DEFRA) from June 2001) in England and Wales and the Scottish Executive Environment and Rural Affairs Department (SEERAD) in Scotland . This is a register of all farm holdings and records the coordinates, farm area, and numbers of animals by species for all farm enterprises under a County-Parish-Holding (CPH) identifier. The agricultural census records the main farm unit (buildings and fields) and all associated off-fields (fields that are not part of the main farm unit) under a single CPH number. A full census was carried out in June of 2000 and has been used in many previous studies of the 2001 FMD epidemic.[9,10,15,16]

Data relating to the 2001 FMD epidemic were recorded by MAFF/DEFRA in the Disease Control System (DCS) database. The DCS records data on the epidemiology of the disease on each IP, including estimated dates of infection and dates of reporting and slaughter. The period of predominantly local spread is analyzed here, so only IPs estimated to have been infected after the NMB were used as the outcome variable; this reduced the number of IPs to 1948. IPs with only pigs were excluded from the model as pig premises played little part in the epidemic after the index case; after the NMB, four pig-only premises (excluded) and 58 premises with pigs and sheep or cattle (included) became IPs.

The DCS data comprised point locations and a CPH number for all culled premises. In the DCS data, fields containing animals culled as part of the FMD control policy were recorded under a separate CPH number if the field was more than 1 km from the main farm holding;[5] these parcels of land were therefore treated as separate holdings for the purposes of epidemic management.[17] However, these parcels of land do not appear in the agricultural census, which means that for any analysis of risk of susceptibility to infection (which looks at the demographic details of all farms–not just those involved in the 2001 outbreak), 444 IPs (22%) are not in the agricultural census and therefore the denominator. Therefore, only the remaining 1527 premises that appear in both the FMD data as IPs and the agricultural census under the same CPH number were used as the outcome variable. Out of 8820 holdings containing animals culled as non-IPs, 4099 were on the census, the remaining 4721 were parcels of land recorded under a separate CPH number to its main holding.

There are 131 663 premises listed on the June 2000 agricultural census that had sheep or cattle registered on them. Not all of these holdings were exposed to infection during the epidemic, therefore only the 83 124 holdings that fall in one of the 37 counties in which there was an IP were included in the analysis. The county of the holding is identified by the CPH number.

### The model

The model is a logistic regression model in which the data points are all premises on the agricultural census with cattle or sheep recorded. The outcome variable was whether a holding was one of 1527 IPs (cases) or one of 80 070 non-IPs (controls, and the reference level). The model construction follows the methodology laid out in Hosmer and Lemeshow.[18]

The predictor variables fall into the four categories below:

(1) Distance to a seed IP: The Euclidean distance in km to the nearest of 78 seed IPs, which were estimated to have been infected before the NMB.[19] These seed IPs are identified using the DCS data on estimated dates of infection.
(2) Regional data–county of the farm: There were 37 counties involved in the study, of which 27 had fewer than 50 IPs. These 27 counties were merged to form a 'rest of GB' predictor, which was the reference level for the 11 level factor (Figure 1).
(3) Holding data: Livestock species on the holding (mixed, cattle only, or sheep only, with mixed the reference level) and area (in Ha) of the holding. Species presence rather than animal numbers on the holding was used because there are a large number of holdings with sheep and no cattle (24.6%) or cattle and no sheep (36.1%), therefore, this variable could not be treated as a continuous distribution.
(4) The neighborhood of the farm: Total numbers of cattle and total numbers of sheep within 10 km of the holding.

Control effort factors including biosecurity, surveillance, and pre-emptive culling have not been explicitly included in the model. However, the control effort varied between counties, which are accounted for here.

Predictors were analyzed in univariate models to screen for risk factors to be entered into subsequent multivariate modeling. Distance to seed and total cattle and sheep in the locality were square root transformed and holding area was $\log_{10} + 1$ transformed to linearize the logits. Predictors from the univariate models that were significant at $p < 0.25$ were included in the multivariate model. Variables were analyzed in the multivariate model and those that were significant at $p < 0.05$ were retained in the model and biologically plausible interactions were tested. Changes to the estimates of the variables were monitored as new variables were added as this could indicate intercorrelation.[18] This was further tested by dropping terms and inspecting the influence of remaining variables.

The discriminatory power of the model was evaluated by taking the 1527 greatest fitted values as predicted IPs. During the model building, AIC was used to compare between models in which the outcome variable was the same. The percentage of these predicted IPs that was actually IPs gave a measure of the ability of the model to discriminate between IPs and non-IPs. The discriminatory power of the model was further analyzed by plotting the receiver operator characteristic (ROC) curve for the model. The ROC curve plots the sensitivity and specificity of the model given a range of cut-off values, where the cut-off value is the threshold that determines whether a test result is considered positive or negative.[18] The sensitivity is the proportion of IPs that would be considered positive given its fitted value falling above a defined cut-off, and the specificity is the proportion of non-IPs that test
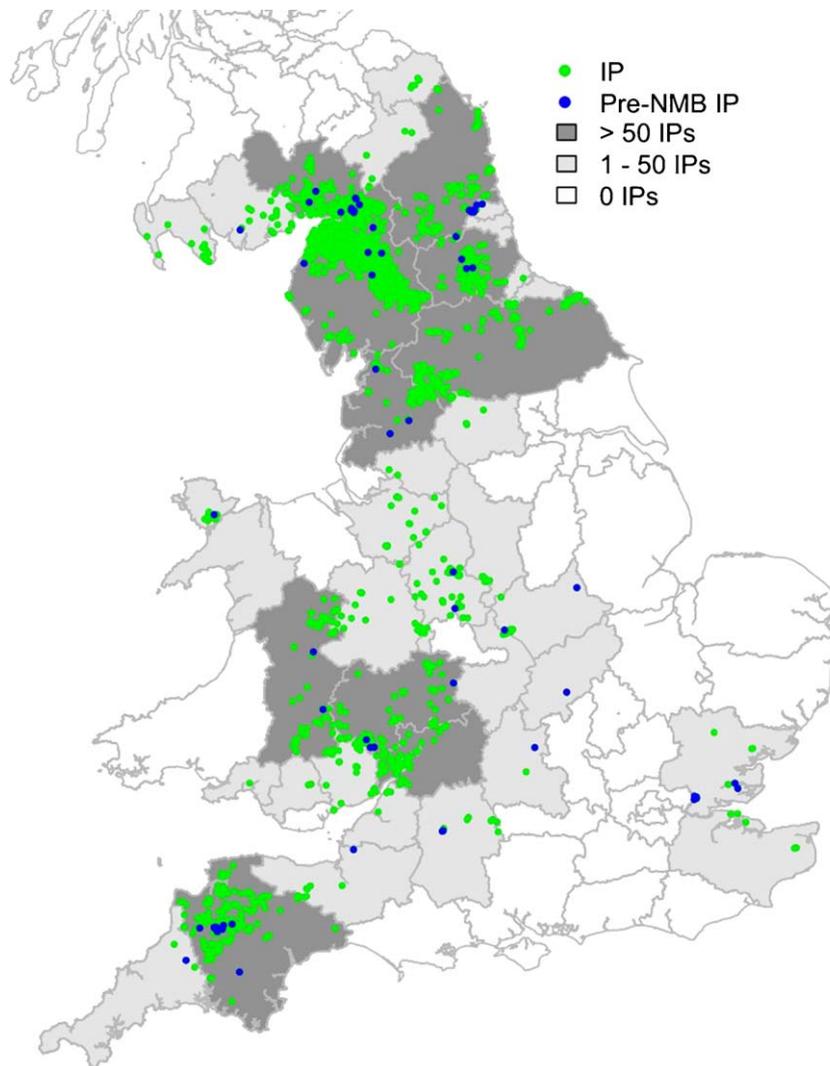
**Figure 1.** Map of GB counties showing infected premises (IPs) and the infection status of each county.

negative given their fitted value falling below that cut-off value. The cut-offs used to generate this ROC curve are 1000 evenly spaced points through the range of fitted values. The goodness of fit of the model was evaluated by calculating the area under the ROC curve (described above) and the Hosmer and Lemeshow Chi-square statistic, both of which use rankings of modeled values to give a measure of the overall model fit (or lack of fit in the case of the Hosmer and Lemeshow test).[18]

Spatial autocorrelation in the model was evaluated by carrying out a Moran's I test for spatial autocorrelation on the model residuals.[20] Spatial autocorrelation was evaluated across a range of numbers of nearest neighbors to test for significant ($p < 0.05$) autocorrelation at a range of scales. Significant autocorrelation was found at all scales (Moran's I statistic $p < 0.001$) and was adjusted for by overlaying a lattice of 5 km wide (edge–edge) hexagons generated using the repeating shapes[21] extension for ESRI ArcView 3.2 onto the farm distribution. Similar approaches have been used elsewhere to overcome this problem,[22,23] however these studies used political boundaries. For this study, a hexagonal grid was used in preference to political boundaries such as parishes, as the hexagon boundaries are arbitrary and all polygons are the same size. The grid of hexagons that covered the susceptible population in this study comprised 5033 hexagons, each one 21.6 km², with a median of 14 farms in each (25th percentile = 6, 75th = 24). The unique ID of the hexagon in question was incorporated into the

model as a random effect to form a generalized linear mixed model (GLMM) with binomial errors. The residuals of the GLMM were tested for spatial autocorrelation using the previously described method.

All modeling was carried out in the R Statistical environment[24] using the spdep package to calculate Moran's I and the nmle4 package to generate the GLMMs.

**Results**

*Univariate analysis*

In univariate analysis, all predictors were statistically significant at the $p < 0.001$ level (Table 1). The principal predictors were the county of the farm, with farms in Cumbria at the greatest risk of infection. Mixed farms were at significantly greater risk than cattle farms, which in turn were at greater risk than sheep-only farms. The surface area of the holding and surrounding cattle and sheep densities were also positively associated with being an IP.

*Multivariate analysis*

All predictors were entered into the multivariate GLMM (Table 2) and all were retained with the exception of sheep density, which was non-significant ($p > 0.1$ at all scales). The fit of

**Table 1**
Univariate logistic regression analysis of IPs and non-IPs. Non-IPs are the reference level. The non-IP and IP columns relate to the number of each group falling into the different levels of each factor.

| Predictor | Unit | Non-IP | IP | z-Score | p-Value | OR (95% CI) |
|---|---|---|---|---|---|---|
| Distance to seed | $\sqrt{km}$ | 80 070 | 1527 | −32.67 | <0.001 | 0.57 (0.55, 0.59) |
| County | Rest of GB | 44 124 | 149 | – | – | 1 |
| | Cumbria | 4320 | 725 | 42.76 | <0.001 | 49.70 (41.55, 59.44) |
| | Devon | 7908 | 135 | 13.57 | <0.001 | 5.06 (4.00, 6.39) |
| | Dumfriesshire | 961 | 108 | 26.86 | <0.001 | 33.28 (25.77, 42.98) |
| | Durham | 1498 | 73 | 18.38 | <0.001 | 14.43 (10.86, 19.18) |
| | Gloucestershire | 2181 | 51 | 11.82 | <0.001 | 6.92 (5.02, 9.54) |
| | Hereford and Worcestershire | 4179 | 41 | 6.02 | <0.001 | 2.91 (2.05, 4.11) |
| | Lancashire | 3153 | 37 | 6.75 | <0.001 | 3.48 (2.42, 4.99) |
| | Northumberland | 1744 | 66 | 16.13 | <0.001 | 11.20 (8.35, 15.03) |
| | N. Yorkshire | 5599 | 99 | 12.75 | <0.001 | 5.27 (4.08, 6.81) |
| | Powys | 4443 | 43 | 6.06 | <0.001 | 2.87 (2.03, 4.02) |
| Species | Mixed | 30 153 | 996 | – | – | 1 |
| | Cattle | 29 584 | 408 | −14.72 | <0.001 | 0.42 (0.37, 0.47) |
| | Sheep | 20 333 | 123 | −17.69 | <0.001 | 0.18 (0.15, 0.22) |
| Holding area | $\log_{10}$ Ha + 1 | 80 070 | 1527 | 25.26 | <0.001 | 3.41 (3.10, 3.75) |
| Cattle density (kernel transformed) | $\sqrt{10^{-3}}$ head | 80,070 | 1,527 | 26.36 | <0.001 | 1.82 (1.74, 1.91) |
| Sheep density (kernel transformed) | $\sqrt{10^{-4}}$ head | 80 070 | 1527 | 18.34 | <0.001 | 3.41 (2.99, 3.89) |

IP, infected premises; OR, odds ratio; CI, confidence interval.

**Table 2**
Multivariate GLMM analysis of risk factors for FMD infection during the 2001 epidemic in GB.

| Predictor | Unit | z-Score | p-Value | OR (95% CI) |
|---|---|---|---|---|
| Intercept | | −20.73 | <0.001 | |
| Distance to seed | $\sqrt{km}$ | −10.31 | <0.001 | 0.613 (0.558, 0.672) |
| County | Rest of GB | – | – | 1 |
| | Cumbria | 15.63 | <0.001 | 25.38 (16.92, 38.07) |
| | Devon | 5.058 | <0.001 | 3.500 (2.154, 5.687) |
| | Dumfriesshire | 6.970 | <0.001 | 8.321 (4.586, 15.10) |
| | Durham | 6.729 | <0.001 | 9.587 (4.963, 18.52) |
| | Gloucestershire | 6.203 | <0.001 | 8.526 (4.332, 16.78) |
| | Hereford and Worcestershire | 2.673 | 0.008 | 2.471 (1.273, 4.798) |
| | Lancashire | 0.918 | 0.358 | 1.411 (0.677, 2.944) |
| | Northumberland | 7.976 | <0.001 | 10.86 (6.042, 19.51) |
| | N. Yorkshire | 8.478 | <0.001 | 10.86 (6.258, 18.85) |
| | Powys | 1.859 | 0.063 | 5.111 (3.577, 7.303) |
| Species | Mixed | – | – | 1 |
| | Cattle | −5.807 | <0.001 | 0.625 (0.533, 0.732) |
| | Sheep | −4.820 | <0.001 | 0.557 (0.439, 0.707) |
| Holding area | $\log_{10}$ Ha + 1 | 18.34 | <0.001 | 4.493 (3.826, 5.275) |
| Cattle density (kernel transformed) | $\sqrt{10^{-3}}$ head | 8.690 | <0.001 | 5.111 (3.577, 7.303) |

GLMM, generalized linear mixed model; FMD, foot and mouth disease; OR, odds ratio; CI, confidence interval.

the model was good; the area under the ROC is 0.962, Hosmer and Lemeshow Chi-square = 49.98 ($p > 0.1$), and it correctly identifies 688 (45.1%, 95% confidence interval (CI) = 42.6, 47.6) IPs. The ROC curve for this model (Figure 2) demonstrates that this model is very good at discriminating between IPs and non-IPs. Perfect discrimination would be demonstrated by a cut-off that maximizes both sensitivity and specificity so that the curve approaches the top left corner of the ROC curve. Furthermore, there was no spatial autocorrelation in the residuals of the GLMM (Moran's I statistic p-value >0.5) and there was little evidence for intercorrelation between predictors.

Farms in the county of Cumbria (odds ratio (OR) = 25.38, 95% CI = 16.92, 38.07) were at the greatest risk of infection relative to farms in the 'rest of GB' category. Decreased distance to the infectious seed (OR = 0.613 per square root km from an infectious seed, 95% CI = 0.558, 0.672) and higher local cattle densities (OR = 5.111 per square root $10^{-3}$ head of cattle, 95% CI = 3.577, 7.303) remained major risk factors. Furthermore, cattle only holdings (OR = 0.625, 95% CI = 0.533, 0.732) and sheep only holdings (OR = 0.557, 95% CI = 0.439, 0.707) were at significantly lower risk of infection than mixed holdings, and holdings that occupy a large surface area remained at greater risk of FMD than small holdings (OR = 4.493 per $\log_{10}$ Ha, 95% CI = 3.826, 5.275).

**Discussion**

A statistical model was developed to calculate the risk of a holding on the agricultural census becoming an IP. The model was a good fit and was able to identify 45% of the actual IPs; the corresponding figure from mathematical models that simulate the 2001 FMD outbreak is between 5% and 15%.[14] The NMB greatly reduced the range of disease transmission and FMD largely remained in the areas in which it was initially seeded, therefore one of the principal predictors of susceptibility to FMD is the distance to a seed infected before the NMB. The presence of FMD in the locality is further described by the county of the holding. However, additional variables relating both to a larger holding area and the surrounding cattle density are also risk factors.

The high quality of the data enabled the identification of individual seeds of infection. In a similar model constructed for susceptibility to highly pathogenic avian influenza (HPAI) in Vietnam,[23] the authors break the epidemic down into time periods and regions to describe the introduction of the virus into different regions in different years. For vector-borne diseases such as sleeping sickness, the point source can be modeled as the vector habitat and the risk of infection decreases with distance to the identified vector habitat.[25] In the 2001 FMD epidemic, the location
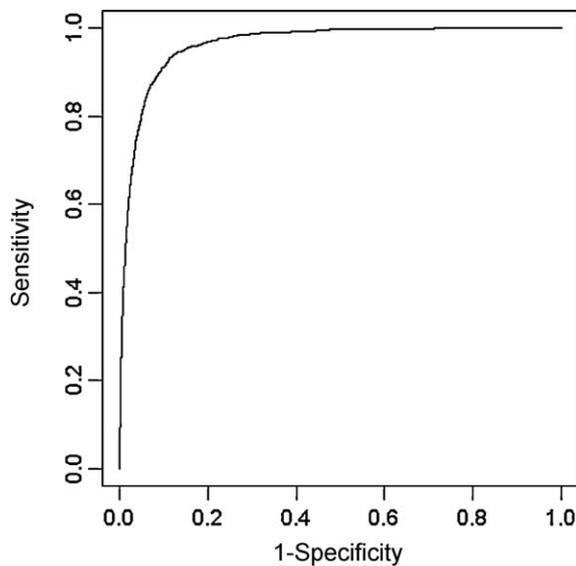
**Figure 2.** Receiver operator characteristic (ROC) curve for the multivariate model in Table 2. Sensitivity refers to the proportion of infected premises (IPs) that would be considered positive given a cut-off, specificity is the proportion of non-IPs that test negative at that cut-off.

of disease introductions was known precisely, and these were included as point sources for infection. However, in the absence of a movement ban, the range of spread would be over much greater distances and the nature of the effect of distance to an infectious seed would be significantly less, particularly as the epidemic progressed. Previous studies[19] have demonstrated the effect of the NMB on controlling longer range spreads of FMD.

In addition to data regarding the point sources of virus introduction, the nature of the susceptible holding is also important in determining its susceptibility. Mixed cattle and sheep holdings are at statistically significantly greater risk than single species holdings; farms covering a greater surface area are also at greater risk because they will generally have more animals, although this varies with the type of farming. Furthermore, larger numbers of cattle in a 10-km radius surrounding a holding is an important risk factor. In spite of the importance of higher cattle densities in this epidemic, sheep density was non-significant in multivariate analysis. This may reflect the infectious challenge to the holding–sheep have been estimated to be an order of magnitude less infectious than cattle.[26] Whilst this analysis has identified demographic risk factors over the entire population of susceptible farms, there remain risk factors at the much smaller scale, such as river and rail barriers, which continue to influence the pattern of FMD cases.[9,27]

The model also shows that the epidemic was operating differently in different parts of the UK, over and above the contribution of other factors in the model; farms in Cumbria were at greatest risk of infection, whilst those in Devon were at relatively low risk compared to some counties with fewer IPs such as Dumfriesshire and Gloucestershire (Tables 1 and 2). These differences may reflect regional heterogeneities in the nature of farming or farming practices, such as the distribution of animals in fields or biosecurity, which result in holdings in certain areas being more susceptible to infection with FMD. These differences may also reflect differences in the management of the epidemic by the different regional disease control centers (DCCs). In spite of the county predictor being highly significant, its removal from the model in sensitivity analysis resulted in slightly greater discriminatory power as measured by area under the ROC curve (0.976 opposed to 0.962). This is because upon the removal of county, some of that variation is taken up by other continuous predictors (in particular distance to seed); this

introduces greater heterogeneity into the fitted values upon which the area under the ROC curve is based.

One of the limitations of using the GLMM framework was that the numbers of animals on the holding could not be included as a predictor due to their zero-inflated distribution. Previous analyses[10,28] found a positive relationship between the numbers of animals by species on the holding and susceptibility, and have found cattle to be an order of magnitude more susceptible than sheep. This has been overcome to an extent in this study by using the holding area and the numbers of animals by species on the holding. There are regional differences in stocking densities, however broadly the holding area will be correlated with the number of animals on the premises.

In conclusion, this analysis demonstrates that post-NMB risk of FMD can be modeled relatively easily using a simple statistical framework and this approach is better at identifying the actual IPs than similar analysis using mathematical model simulations of the outbreak.

## Conflict of interest

No conflict of interest to declare.

## Acknowledgements

## References

1. Domingo E, Baranowski E, Escarmis C, Sobrino F. Foot-and-mouth disease virus. Comp Immunol Microbiol Infect Dis 2002;**25**:297–308.
2. Sorensen JH, Mackay DKJ, Jensen CO, Donaldson AI. An integrated model to predict the atmospheric spread of foot-and-mouth disease virus. Epidemiol Infect 2000;**124**:577–90.
3. Anderson I. Foot and Mouth disease 2001: lessons to be learned inquiry. London: The Stationary Office; 2002.
4. National Audit Office. The 2001 outbreak of foot and mouth disease. London: The Stationary Office; 2002.
5. Gibbens JC, Wilesmith JW. Temporal and geographical distribution of cases of foot-and-mouth disease during the early weeks of the 2001 epidemic in Great Britain. Vet Rec 2002;**151**:407–12.
6. Alexandersen S, Kitching RP, Mansley LM, Donaldson AI. Clinical and laboratory investigations of five outbreaks of foot-and-mouth disease during the 2001 epidemic in the United Kingdom. Vet Rec 2003;**152**:489–96.
7. Mansley LM, Dunlop PJ, Whiteside SM, Smith RGH. Early dissemination of foot-and-mouth disease virus through sheep marketing in February 2001. Vet Rec 2003;**153**:43–50.
8. Thrusfield M, Mansley LM, Dunlop PJ, Taylor J, Pawson A, Stringer L. The foot-and-mouth disease epidemic in Dumfries and Galloway, 2001. 1: Characteristics and control. Vet Rec 2005;**156**:229–52.
9. Savill NJ, Shaw DJ, Deardon R, Tildesley MJ, Keeling MJ, Woolhouse ME, et al. Topographic determinants of foot and mouth disease transmission in the UK 2001 epidemic. BMC Vet Res 2006;**2**:3.
10. Keeling MJ, Woolhouse ME, Shaw DJ, Matthews L, Chase-Topping M, Haydon DT, et al. Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. Science 2001;**294**:813–7.
11. Tildesley MJ, Savill NJ, Shaw DJ, Deardon R, Brooks SP, Woolhouse ME, et al. Optimal reactive vaccination strategies for a foot-and-mouth outbreak in the UK. Nature 2006;**440**:83–6.
12. Ferguson NM, Donnelly CA, Anderson RM. Transmission intensity and impact of control policies on the foot and mouth epidemic in Great Britain. Nature 2001;**413**:542–8.
13. Ferguson NM, Donnelly CA, Anderson RM. The foot-and-mouth epidemic in Great Britain: pattern of spread and impact of interventions. Science 2001;**292**:1155–60.
14. Tildesley MJ, Deardon R, Savill NJ, Bessell PR, Brooks SP, Woolhouse ME, et al. Accuracy of models for the 2001 foot-and-mouth epidemic. Proc R Soc Lond Ser B-Biol Sci 2008;**275**:1459–68.
15. Wilesmith JW, Stevenson MA, King CB, Morris RS. Spatio-temporal epidemiology of foot-and-mouth disease in two counties of Great Britain in 2001. Prev Vet Med 2003;**61**:157–70.

16. Morris RS, Wilesmith JW, Stern MW, Sanson RL, Stevenson MA. Predictive spatial modelling of alternative control strategies for the foot-and-mouth disease epidemic in Great Britain, 2001. *Vet Rec* 2001;**149**:137–44.
17. Honhold, N., Taylor, N. M. Data quality assessment: comparison of recorded and contemporary data for farm premises and stock numbers in Cumbria, 2001 in Proceedings for the Society of Veterinary Epidemiology and Preventive Medicine 152-163 (Exeter, 2006). .
18. Hosmer DW, Lemeshow S. Applied logistic regression. Chichester: Wiley; 2000.
19. Haydon DT, Chase-Topping M, Shaw DJ, Matthews L, Friar JK, Wilesmith JW, et al. The construction and analysis of epidemic trees with reference to the 2001 UK foot-and-mouth outbreak. *Proc R Soc Lond Ser B-Biol Sci* 2003;**270**:121–7.
20. Crawley M. The R book. Chichester: Wiley; 2007.
21. Jenness, J. Repeating shapes (repeat_shapes.avx) extension for ArcView 3.x. Jenness Enterprises; 2005.
22. Haine D, Boelaert P, Pfeiffer DU, Saegerman C, Lonneux JF, Losson B, et al. Herd-level seroprevalence and risk-mapping of bovine hypodermosis in Belgian cattle herds. *Prev Vet Med* 2004;**65**:93–104.

23. Pfeiffer DU, Minh PQ, Martin V, Epprecht M, Otte MJ. An analysis of the spatial and temporal patterns of highly pathogenic avian influenza occurrence in Vietnam using national surveillance data. *Vet J* 2007;**174**:302–9.
24. R Development Core Team. R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2006.
25. Odiit M, Bessell PR, Fèvre EM, Robinson T, Kinoti J, Coleman PG, et al. 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* 2006;**100**:354–62.
26. Donaldson AI, Alexandersen S, Sorensen JH, Mikkelsen T. Relative risks of the uncontrollable (airborne) spread of FMD by different species. *Vet Rec* 2001;**148**:602–4.
27. Bessell PR, Shaw DJ, Savill NJ, Woolhouse ME. Geographic and topographic determinants of local FMD transmission applied to the 2001 UK FMD epidemic. *BMC Vet Res* 2008;**4**:40.
28. Diggle PJ. Spatio-temporal point processes, partial likelihood, foot and mouth disease. *Stat Methods Med Res* 2006;**15**:325–36.