



HAL
open science

Is a good predictor of final epidemic size: Foot-and-mouth disease in the UK

Michael J. Tildesley, Matt J. Keeling

► **To cite this version:**

Michael J. Tildesley, Matt J. Keeling. Is a good predictor of final epidemic size: Foot-and-mouth disease in the UK. *Journal of Theoretical Biology*, 2009, 258 (4), pp.623. 10.1016/j.jtbi.2009.02.019 . hal-00554581

HAL Id: hal-00554581

<https://hal.science/hal-00554581>

Submitted on 11 Jan 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Author's Accepted Manuscript

Is R_0 a good predictor of final epidemic size: Foot-and-mouth disease in the UK

Michael J. Tildesley, Matt J. Keeling

PII: S0022-5193(09)00082-4
DOI: doi:10.1016/j.jtbi.2009.02.019
Reference: YJTBI5476

To appear in: *Journal of Theoretical Biology*

Received date: 25 November 2008
Revised date: 28 January 2009
Accepted date: 14 February 2009

Cite this article as: Michael J. Tildesley and Matt J. Keeling, Is R_0 a good predictor of final epidemic size: Foot-and-mouth disease in the UK, *Journal of Theoretical Biology* (2009), doi:10.1016/j.jtbi.2009.02.019

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



www.elsevier.com/locate/jtbi

Is R_0 a good predictor of final epidemic size:
Foot-and-Mouth Disease in the UK.

Michael J. Tildesley¹ and Matt J. Keeling²

1 Centre for Infectious Diseases, Institute of Immunology and Infection Research, University of Edinburgh, Ashworth Labs, West Mains Road, Edinburgh, EH9 3JT, UK

2 Department of Biological Sciences, University of Warwick, Gibbet Hill Road, Coventry, CV4 7AL, UK

One of the main uses of an epidemic model is to predict the scale of an outbreak from the first few cases. In a homogeneous and non-spatial model there is a straightforward relationship between the basic reproductive ratio, R_0 , and the final epidemic size; however when there is a significant spatial component to disease spread and the population is heterogeneous predicting how the epidemic size varies with the initial source of infection is far more complex. Here we use a well-developed spatio-temporal model of the spread of foot-and-mouth disease, parameterised to match the 2001 UK outbreak, to address the relationship between the scale of the epidemic and the nature of the initially infected farm. We show that there is considerable heterogeneity in both the likelihood of a epidemic and the epidemic impact (total number of farms losing livestock to either infection or control) and that these two elements are best captured by measurements at different spatial scales. The likelihood of an epidemic can be predicted from a knowledge of the reproduction ratio of the initial farm (R_i), whereas the epidemic impact conditional on an epidemic occurring is best predicted by averaging the second-generation reproduction ratio ($R_i^{(2)}$) in a 58 km ring around the infected farm. Combining these two predictions provides a good assessment of both the local and larger-scale heterogeneities present in this complex system.

The role of R_0 in disease spread

Mathematical modelling of infectious diseases is a rapidly growing research area. Recent high profile epidemics in animals (e.g. Foot-and-Mouth Disease (FMD), Bluetongue Virus (BTV) and Avian Influenza (AI)), and in humans (e.g SARS) has increased the need for a greater understanding of the processes leading to the spatial spread of disease (Ferguson *et al* 2001; Keeling *et al* 2001; Savill *et al* 2006; Lipsitch *et al* 2003). Recent work on foot-and-

mouth disease has highlighted the role of local spatial transmission and spatial variation in host populations as two important concepts in understanding (and hence optimising the control of) infection. Since the 2001 epidemic, a range of publications (including Keeling *et al* 2001, Ferguson *et al* 2001, Le Menach *et al* 2005) have constructed maps of the local average farm-level basic reproductive ratio values (R_i) to act as a spatial prediction of at-risk regions and the likely spatial extent of an epidemic (see for example Fig 1*a*). Here we address two fundamental questions about such maps: firstly can they be used to inform epidemic properties such as the probability of an epidemic and the size of the ensuing epidemic; and secondly what is the appropriate scale to measure the local average R_i values.

From traditional mathematical models of infectious diseases, a single quantity (the population-level basic reproductive ratio, R_0) emerges which informs about the ability of a disease to invade and spread through a naive population. R_0 is defined as the number of secondary infections caused by a single infected source in a totally susceptible population (Diekmann *et al* 1990; Anderson and May 1991). As well as defining a threshold for invasion, R_0 can also inform about expected growth rates of cases, the likely success of an epidemic in a stochastic framework, the level of counter measures needed to control an epidemic and the total number of cases expected (Kermack & McKendrick 1927). However, when dealing with the spatial spread of infection, especially in a highly heterogeneous population, many other factors will contribute to the final size of the epidemic (May and Lloyd 2001).

From standard SIR models without births or deaths, it can be shown that the expected cases as a fraction of the entire population (also known as the final epidemic size, R_∞) is determined by the equation:

$$R_\infty = S(0) \left[1 - e^{(-R_0 R_\infty)} \right]$$

which assumes that the epidemic starts with a small seed of infection in a large homogeneous population, in which a proportion $S(0)$ are initially susceptible, and that the epidemic is not

driven extinct early by stochastic forces (Kermack & McKendrick 1927). From a probabilistic argument we find that this basic formulation is independent of the natural history of the pathogen (Keeling & Rohani 2007). However, the expected total number of cases is influenced by the risk of early stochastic extinction. Considering the early epidemic as a branching process, the probability of extinction (before a major epidemic) starting with a single infected individual is:

$$P_{\text{ext}}^{\text{exp}} = \frac{1}{R_0} \quad P_{\text{ext}}^{\text{fixed}} = 1 - R_{\infty}.$$

(Kermack & McKendrick 1927; Keeling 2004), where the two values give the probability of extinction assuming exponentially distributed infectious periods (a constant rate of recovery) and fixed duration infectious periods respectively (Keeling & Rohani 2007). The probability of an infection occurring is an important quantity when predicting final epidemic size – faced with an epidemic threat, policy makers are concerned with the risk of an epidemic occurring and, should an epidemic occur, the resultant epidemic size. We therefore see that for stochastic populations the expected total number of cases should be approximated by the standard final size expression combined with the probability that the initial infection manages to cause an epidemic:

$$\text{Expected number of cases} = \text{Population Size} \times R_{\infty} \times (1 - P_{\text{ext}}).$$

The above theory holds for randomly mixed and homogeneous populations. For more complex, realistic systems where transmission has a strong spatial aspect and there is considerable heterogeneity within the population, we generally rely on percolation theory to understand thresholds of spatial spread (Sander *et al.* 2002); however according to such models the threshold effectively separates parameter regions of small and large-scale (infinite) epidemics. Here we consider how such percolation ideas translate to populations in which there is considerable heterogeneity at large spatial scales, such that some regions can readily sustain an epidemic while in other regions epidemics are doomed to rapid extinction. In addition, when considering

the spread of foot-and-mouth disease we must not only count farms that lose livestock through infection, but also farms losing livestock owing to control measures.

For the purposes of the investigation presented in this paper, we aim to understand the role of local farm-level reproductive ratio, R_i , in disease spread. During the early stages of an outbreak, it is vital for policy makers to know the likelihood of a large scale epidemic occurring so that appropriate control measures can come into force. We therefore investigate whether the R_i values of the initially infected farm or region is an accurate guide of final epidemic impact (defined as number of farms infected or culled as part of any control policy), or whether an alternative parameter should be used from the early infection data to estimate the overall impact of the outbreak.

The Model

The first generation of the model used in the analysis presented in this paper was developed in 2001 by Keeling and collaborators to model the FMD epidemic in the UK (Keeling *et al.* 2001). It has since been adapted to investigate a variety of control strategies (Keeling *et al.* 2003, Tildesley *et al.* 2006), and has been refined by including power law scaling of farm-level susceptibility and infectiousness with the number of livestock.

The rate at which an infectious farm i infects a susceptible farm j is given by:

$$\text{Rate}_{ij} = ([N_{sheep,j}]^{p_s} S_{sheep} + [N_{cow,j}]^{p_c} S_{cow}) \times ([N_{sheep,i}]^{q_s} T_{sheep} + [N_{cow,i}]^{q_c} T_{cow}) \times K(d_{ij})$$

where $N_{s,i}$ is the number of livestock species s recorded as being on farm i , S_s and T_s measure the species-specific susceptibility and transmissibility, d_{ij} is the distance between farms i and j and K is the transmission kernel. The transmission kernel takes the same form as in previous work (Keeling *et al.* 2001) and is estimated from contact tracing, which captures how the rate of infection decreases with distance. p_s , p_c , q_s and q_c are power law parameters accounting for non-linear increases in susceptibility and transmissibility as animal numbers on a farm increase.

These parameters are estimated from the 2001 epidemic data; previous work has found that this power law model provides a closer fit to the 2001 data than one in which the powers are set to unity (Tildesley *et al* 2008). As in previous versions of the model, the parameters are estimated for five distinct regions: Cumbria, Devon, the rest of England (excluding Cumbria and Devon), Scotland and Wales. This enables us to account for differences in regional farming practices and culling levels (table 1).

In the 2001 epidemic, Dangerous Contact farms (DCs) were identified (and animals on these farms culled) for each Infected Premises (IP) based upon veterinarian judgement of risk factors and known movements from an IP. All farms sharing a common boundary with an IP were defined as Contiguous Premises (CPs) and were culled from 23rd March as part of an increased effort to control the spread of disease. Details of the modelling techniques used for CPs and DCs can be found in earlier papers (Keeling *et al.* 2001; Tildesley *et al.* 2008).

In line with previous versions of the model (Keeling *et al* 2001; Tildesley *et al* 2006), we make the assumption here that all farms are infected for 5 days before becoming infectious, and are infectious for 4 days before being reported with infection. Once a farm is reported, control measures are enforced. We assume a 24/48 hour policy, such that all IPs are culled within 24 hours of being reported and all DCs and CPs are culled within 48 hours.

For the above model we calculate the farm-level basic reproductive ratio R_i as:

$$R_i = \sum_{j \neq i} \text{Prob}_{ij} = \sum_{j \neq i} 1 - \exp(-\text{Rate}_{ij}P)$$

where P is the length of the infectious period and Prob_{ij} is the probability that farm i infects farm j over its entire infectious period. It is now well understood that for a wide variety of spatial models R_i is an over-estimate of the eventual growth of infection (Keeling & Eames 2005) as the strong negative spatial correlations between susceptible and infected individuals is yet to develop. As an acknowledgement of this factor, we calculate $C_i^{(2)}$ defined as the

Region	Parameters						
	$T_{cow} (\times 10^{-7})$	$T_{sheep} (\times 10^{-7})$	S_{cow}	q_c	q_s	p_c	p_s
Cumbria	15.9	9.2	10.0	0.42	0.49	0.41	0.20
Devon	8.1	8.4	8.9	0.37	0.42	0.37	0.40
Wales	8.0	1.9	3.6	0.25	0.22	0.31	0.43
Scotland	8.4	9.7	10.8	0.20	0.40	0.23	0.33
Rest of England	11.9	9.9	8.0	0.44	0.37	0.42	0.30

Table 1. Species-specific parameters for five regions of the UK that provide the best-fit to the cumulative number of cases and culls at a regional level. Levels of culling within each region have been scaled to match the temporal pattern in that given region. In all regions, S_{sheep} is set to equal 1.

average number of cases generated after 2 generations starting with infectious farm i in a totally susceptible population:

$$C_i^{(2)} = \sum_{k \neq i} 1 - \exp \left(-P \left[\text{Rate}_{ik} + \sum_{j \neq i, k} \text{Prob}_{ij} \text{Rate}_{jk} \right] \right)$$

From this definition we can determine the second-generation basic reproductive ratio $R_i^{(2)}$ defined as the (appropriately weighted) average number of cases generated by each secondary case when the primary infection is farm i :

$$R_i^{(2)} = \frac{C_i^{(2)}}{R_i} - 1$$

In a homogeneous mean-field (randomly mixed) model the primary and second-generation basic reproductive ratios are equal ($R_i^{(2)} = R_i = R_0$). When heterogeneities exist in the population the second-generation basic reproductive ratio is often larger than the primary, as the infection is most likely to have become focused within the high-risk individuals who are both more susceptible and more infectious (Eames and Keeling 2002). In contrast, when there is a strong

spatial element to transmission (but the population is otherwise homogeneous) then the second-generation basic reproductive ratio is generally smaller than the primary as the secondary cases find themselves in an environment with a reduced local density of susceptibles (negative spatial correlations have developed between susceptible and infectious individuals) and hence a reduced potential to cause further cases. In a heterogeneous spatial population either of these factors may dominate – this is explored later in Figure 2.

Finally for both of these basic reproductive ratios we define a local spatial average:

$$\bar{R}_i(D) = \frac{\sum_{d_{ij} < D} R_j}{\sum_{d_{ij} < D} 1} \quad \text{and} \quad \bar{R}_i^{(2)}(D) = \frac{\sum_{d_{ij} < D} R_j^{(2)}}{\sum_{d_{ij} < D} 1} \quad (1)$$

This local spatial average is the average over all farms within a distance D of the index farm.

Simulation Results

Before the UK 2001 epidemic, 188496 farms were identified as livestock farms, although only 142268 farms were actually part of the June 2000 census. For all 188496 livestock farms, the County/Parish/Holding number (CPH), the X-Y co-ordinates of the farmhouse, the area of the farm, and the number of cattle, pigs, sheep, goats and deer is recorded although the number of livestock on each farm is subject to variation throughout the year as animals are born and others are moved on or off the holding (Anderson 2002).

We now perform multiple epidemic simulations – starting with a single infected farm (labelled farm i) we allow the epidemic to propagate forwards while performing the appropriate culling measures,. For each of the possible 188496 index farms we perform 10 replicate simulations; for each index farm we record the average epidemic impact, E_i , (the average number of farms either infected or culled) and the proportion of replicates that fail to generate more than 10 cases, P_i^{fail} . We additionally define the conditional epidemic impact, E_i^{success} , calculated as the average epidemic impact of those replicates that give rise to more than 10 cases; when the

epidemic fails in all replicates ($P_i^{\text{fail}} = 1$) then the conditional epidemic impact is undefined. In addition, 100 replicate simulations are used to determine P_i^{fail} more accurately, as these can be halted once the number of cases exceeds 10.

Results of these simulations are shown in figure 1. Graph *b* shows the average epidemic impact E_i for all 188496 livestock farms in the UK. The two most striking features are the sheer range of epidemic impacts predicted and the extreme heterogeneity at a range of scales. There is clearly broad qualitative agreement between regions of the country with high local average R_i (as shown in Figure 1a) and regions that are predicted to give rise to large-scale epidemics, but even within these regions there are considerable differences between individual farms. Notably infected premises in Cumbria and Dumfries & Galloway often give rise to epidemic impacts in excess of 4000 farms, with values exceeding 8000 farms not uncommon, but even in these regions are farms which fail to generate any sizable epidemics. For the 2001 outbreak it has been estimated that 73 premises were infected before stringent control measures were implemented (Mansley *et al* 2003) and this gave rise to an epidemic impact of 10242 farms – a results that is in agreement with our model simulations for the same conditions. In comparison the large epidemic impacts predicted here from a single infected farm highlight the strong non-linearities in the epidemic process.

Graph 1c shows the probability of each infected farm generating a significant outbreak (defined as being more than 10 cases). Again there is good qualitative agreement between regions likely to give rise to an epidemic and regions of high average R_i (comparing maps *a* and *c*). It should be noted that for many areas of the country – including the location of the 2007 Surrey outbreak – a substantial epidemic is highly unlikely. We now wish to relate E_i , P_i^{fail} and E_i^{success} to the farm-level and local reproductive ratios to see if we can capture some of the underlying heterogeneity.

Figure 2 considers how R_i , $R_i^{(2)}$ and E_i co-vary. In general we find that there is a strong

positive correlation between first and second generation reproductive ratios; however, the relationship is non-linear and far from the homogeneous prediction of $R_i^{(2)} = R_i$. We find that the correlation coefficient for R_i and $R_i^{(2)}$ is 0.345 ($p < 0.001$). For large values of R_i we observe that $R_i^{(2)}$ is generally lower than R_i , with the average $R_i^{(2)}$ values being between 2 and 3; this reduction reflects both the spatial structure that develops during the first generation but also highlights that farms with high R_i may be imbedded in a population that lies closer to the average. For small R_i values it is still possible for $R_i^{(2)}$ to be relatively large, reflecting both the local heterogeneity in the population and the greater susceptibility associated with high-risk farms. In general we find that the average R_i is 0.61 while the average $R_i^{(2)}$ is 0.728; given that the average second generation reproductive ratio exceeds the first generation average, we see that on average the natural focusing of the epidemic towards high-risk farms outweighs the impact of spatial structure. In addition, it is clear that the smallest epidemic impacts (coloured blue) are associated with small values of R_i and $R_i^{(2)}$, while larger reproductive ratios generally lead to larger epidemic impacts. We can also investigate the correlation between the predicted Epidemic Impact E_i and R_i or $R_i^{(2)}$, the correlation coefficients are found to be 0.376 and 0.444 respectively ($p < 0.001$). However we find that simple linear regressions are poor predictors of the dynamics; instead we use a non-parametric model to capture the degree to which individual or local averages of the reproductive ratios can be used to predict the epidemic impact and other large-scale aggregate features of the epidemics.

For a given observed quantity x_i associated with farm i (for example its first generation reproductive ratio), we use this quantity to predict an aggregate epidemic feature A_i (for example the epidemic impact) as follows:

$$\text{Predicted}(A_i) = \frac{1}{n} \sum_{x_j \in \text{nearest}(n, x_i)} A_j$$

where $\text{nearest}(n, x_i)$ allows us to identify the n farms that have the nearest x value to x_i . We therefore wish to see if the aggregate epidemic feature associated with a given farm can be

predicted by examining the aggregate epidemic feature associated with farms that have similar local quantities – that is, do observed local quantities allow us to predict more global epidemic patterns? For this analysis, we seek to minimise the error, $\text{Error}(A)$, such that:

$$\text{Error}(A) = \sqrt{\sum_i (\text{Predicted}(A_i) - \text{True}(A_i))^2}$$

Throughout we take n to be 0.1% of the population size. Although more sophisticated methods of prediction (rather than taking a near average) could be employed, given the sheer number of farms these are unnecessary and do not provide any noticeable benefits.

The most obvious question to address is whether the individual first-generation reproductive ratio, R_i , is a good predictor of epidemic impact E_i . Using the method outlined above we find that the value of R_i explains 29.3% of the standard deviation in E_i – although we note that 17.8% of the standard deviation may be attributable to the large variation between epidemics with the same seed. However, much greater precision can be achieved by decomposing the average epidemic impact (E_i) into two elements, the probability an epidemic is successfully spawned ($1 - P_{\text{fail}}$) and the epidemic size conditional on it successfully taking off (E_i^{success}).

In agreement with standard results, the best predictor of epidemic failure is the individual reproductive ratio of the farm (R_i). The results of this analysis are shown in figure 3a. As expected the probability of an epidemic taking off (generating more than 10 cases) increases with R_i ; however $R_i = 1$ no longer acts as a threshold. In fact at the threshold value of $R_i = 1$ we find that 20% of initial seedings generate epidemics; this probability increases to around 50% for the largest R_i values. These results clearly highlight the stochastic nature of the disease in its early stages, and the dependence of the ensuing epidemic on favourable local conditions in the neighbourhood of the initial infection. The reason why low R_i farms are still capable of generating a large epidemic is exemplified in figure 2: farms with low R_i may still have values of $R_i^{(2)}$ above the threshold of one and therefore if the infection spreads beyond the initial farm its likely success is dramatically increased. Using additional information on

$R_i^{(2)}$ provides a slightly more accurate prediction of epidemic success, but in general the same qualitative features hold.

Predicting the average epidemic impact, E_i , is far more involved as the size of an epidemic (and the control measures used) depend upon the population structure in general. In particular, once an epidemic has taken off, the structure of all farms in that region will affect spread of disease. Therefore, a more appropriate indicator of epidemic size may come from averaging individual farm-level reproductive ratios in a local region around each farm. We therefore investigate how the accuracy with which we can predict the epidemic impact varies as we adjust the scale of the region in which we take the average (parameter D in equation 1). We consistently find that averaging within a radius of $D \approx 50 - 60\text{km}$ provides the best predictors, and more specifically that $\overline{R_i^{(2)}}(58\text{km})$ predicting the conditional epidemic impact E_i^{Success} offers the best match between local properties and large-scale epidemic potential (figures 3b and c). This scale of 58km is clearly a compromise between a range of epidemiological factors, in particular the ability to find regions of sufficient livestock density to support a continuing epidemic in both the short and long term. Table 2 provides a more detailed examination of errors associated with predicting the epidemic impact; we consistently find that $R_i^{(2)}$ provides a better prediction (lower error) than R_i , and that averages over 58km provide a better prediction than the individual-level measures. Finally, only considering the average epidemic impact from outbreaks that exceed 10 cases further improves our ability to predict the outcome.

The above results can be related back to correlation coefficients discussed earlier, we find that the correlation coefficient between $\overline{R_i^{(2)}}(58\text{km})$ and E_i is 0.526 ($p < 0.001$) which is much stronger than for local measures. We can also measure the residual errors from simple linear regression models which are shown in the right-hand column of Table 2; clearly our non-parametric fit performs better than a linear regression, especially as our predictive ability increases. We can also extend this linear analysis to far higher dimensions, taking into ac-

Parameter	Error(E_i)	Residual Error
R_i	4.30×10^5	4.31×10^5
$R_i^{(2)}$	4.16×10^5	4.17×10^5
$\overline{R}_i(58km)$	3.27×10^5	4.08×10^5
$\overline{R}_i^{(2)}(58km)$	3.08×10^5	3.95×10^5
Parameter	Error(E_i^{success})	Residual Error
$\overline{R}_i^{(2)}(58km)$	2.84×10^5	3.91×10^5

Table 2. Table showing the errors in predicting epidemic impact when errors are calculated based upon R_i and $R_i^{(2)}$ values of each farm, and averaged over the optimal radius of 58km. Values are shown for both the non-parameteric fit and linear regression.

count R_i and $R_i^{(2)}$ values averaged over a range of distances ($E_i \approx \sum_{0 \leq D \leq 60km} A_D \overline{R}_i(D) + B_D \overline{R}_i^{(2)}(D) + C$). However, even this more elaborate linear fit produces an error of 3.81×10^5 , which is only a marginal improvement on the simple regression using $\overline{R}_i^{(2)}(58km)$ and is still worse than our non-parameteric fit.

Finally, in figure 4 we bring together the two predictive measures of epidemic impact: the predicted probability of success based upon the individual R_i and the predicted conditional epidemic impact based upon $\overline{R}_i^{(2)}(58km)$. By combining predictions that utilize individual-level farm information at two distinct spatial scales we are able to generate far more reliable predictions of the total average epidemic impact starting in each farm – and by using a combination of measures we are able to capture some of the local heterogeneity in the results. Figure 4 shows the predicted value, which should be compared to Figure 1b; we note that these predictions explain 31.7% of the standard deviation. We therefore have a relatively simple measure which will allow us to translate from individual-level properties of farms (which are relatively quick

to calculate) to emergent population-level values. It is therefore maps such as figures 1*b* and 4, that should be used to provide an understanding of the risk posed by an outbreak in any given location.

Discussion

One of the most valuable lessons learnt from the 2001 FMD epidemic is that information regarding the risk of spread of infection is required as soon as possible during the early stages of an outbreak of disease to enable policy makers to put control measures in place. An analysis of the role of R_i on final epidemic impact when infection occurs on farm i sheds some light on the ability of models to predict the risk of large scale outbreaks of disease based upon early epidemic behaviour. A large range of epidemic impacts is observed, with largest epidemics generally found to occur when primary cases are in Wales, Devon, the North of England and Southern Scotland. This is unsurprising – these regions were epidemic hotspots during the 2001 epidemic. We note that the work presented in this paper focuses on Foot-and-Mouth disease and the 2001 strain of FMD specifically and, whilst the general results could be applied to other diseases, more work would need to be done to establish the conclusions for a range of livestock and human diseases, as well as for other strains of FMD.

The conclusion reached in this paper is that individual farm-level R_i is not the main predictor of epidemic impact for an outbreak of FMD in the UK. We find that $R_i^{(2)}$ averaged within a 58km radius around infected cases provides the best predictive measure of final epidemic impact. This indicates that it is not only the nature of the primary case but the potential of other farms in the locality that influence the extent of an outbreak of disease. The number of cases infected by the source farm in the first generation provides a measure of the local spatial distribution of infected farms. However, the number of subsequent cases produced by an average secondary case ($R_i^{(2)}$) is far more informative about the spread of infection, as it includes the establishment of spatial correlations and the competition for local susceptible farms. In theory, this process could be

extended to examine $R_i^{(3)}$ terms, but these higher-order terms would need to be calculated by multiple simulations rather than numerically reducing their immediate use.

The spatial scale over which $R_i^{(2)}$ is averaged is clearly a compromise between capturing the early ability of the disease to spread and its ability to generate a large-scale epidemic. The optimal scale of 58km is likely to depend on several factors including the national farm landscape, the degree of spatial spread of the pathogen and the use of control measures. We are still a long way from solving the general problem of percolation of infection through a highly heterogeneous environment, but by analysing the sensitivity of our result to the culling effort we gain some understanding of its robustness. We consider two other epidemic scenarios: doubling the control culling levels and only culling infected farms. These two extreme cases lead to similar optimal averaging scales of 56.5km and 62km respectively, although the non-parametric function used to predict the epidemic impact is very different in both cases. This provides some evidence that for foot-and-mouth disease in the UK livestock industry, taking a spatial average within a 50-60km radius consistently generates a useful measure of the likely extent of an epidemic.

Related research has been conducted to search for signatures of epidemic outbreaks using simple epidemic models in a number of spatial network structures with the goal of determining epidemic features that can provide significant "separability" to epidemics developed in well-mixed systems using a multi-dimensional scaling approach (Burr and Chowell 2008). The effects of network clustering on the final epidemic size and the probability of outbreaks occurring has been previously investigated (Vazquez *et al* 2003; Boguna *et al* 2003), and similar tools could be applied to investigate the role of clustering, R_i and $R_i^{(2)}$ in the context of an FMD outbreak. Alternatively, we could use the transmission kernel to generate a static network of connected farms; the local and global properties of such a network may also provide a fruitful avenue of future research and has much in common with the recent work by Burr and Chow-

ell (2008) who used multi-dimensional scaling to elucidate which features most influenced the characteristics of an epidemic spreading through a spatial network.

This work provides a useful tool for epidemic predictions and, in turn ascertaining the necessity for the introduction of control measures - risk maps such as those shown in figure 1b are computationally intensive, whilst those in figure 4 are not. The close similarity between the two figures means that risk can be determined more easily by such a method particularly when investigating the effects of variation in parameter values. Of course, in an outbreak with multiple initially infected cases in the same region, the infection processes will interact with one another (i.e. the predicted epidemic impact in such a scenario will not simply be the sum of the epidemic impacts assuming only one primary case).

It is somewhat surprising that a simple average within a given radius provides the greatest predictive ability, but this measure consistently out-performs a range of more involved spatially weighted averages, including measures using the 2001 transmission kernel. It remains to be seen what epidemiological factors contribute to the optimal scale of 58km; we feel that this scale coincides with the spatial scale of heterogeneity within the UK livestock industry – the spatial aggregation of cattle in Cumbria occurs at a comparable scale.

In summary, we find that although spatial maps of locally-averaged farm level R_i (eg Figure 1a, which uses 5×5 km squares) provide a rapid means of assessing the risk associated with an initial seed of infection of FMD, this only provides a limited qualitative understanding. We have shown that it is dangerous to simply average R_i without paying particular attention to issues of spatial scale, and that no single measure can capture the two elements of risk and consequences. In this example, both the individual level values of R_i and a more wide-scale spatial average $\overline{R_i^{(2)}}(58km)$ control the potential for an epidemic and its subsequent impact. Of course, we are yet to reach the stage of obtaining true understanding of percolation in a heterogeneous spatial environment. However, such results as presented in this paper may help us to identify high-risk

regions in other countries or when different strains lead to alternative epidemiological parameters.

References

1. Anderson, I. Foot-and-Mouth Disease 2001: Lessons to be Learnt Enquiry. (2002).
2. Anderson, R.M. and May, R.M. Infectious Diseases of Humans: Dynamics and Control *Oxford University Press* (1991).
3. Boguna, M., Pastor-Satorras, R. & Vespignani, A. Epidemic Spreading in Complex Networks with Degree Correlations. *Lecture Notes in Physics*, Springer Berlin / Heidelberg. **625** (2003).
4. Burr, T.L. & Chowell, G. Signatures of non-homogeneous mixing in disease outbreaks. *Math. Comp. Model.* **48** Issues 1-2, 122-140 (2008).
5. Diekmann *et al.* On the definition and the computation of the basic reproduction ratio R_0 in models for infectious diseases in heterogeneous populations. *Journal of Mathematical Biology* (1989) vol. 28 (4) pp. 365-82)
6. Eames K.T.D. and Keeling M.J. Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. *PNAS* **99**, 13330-13335 (2002).
7. Ferguson *et al.* Transmission intensity and impact of control policies on the foot and mouth epidemic in Great Britain. *Nature* **413** 542-548 (2001).
8. Keeling, M.J. and Rohani P. Modelling Infectious Diseases in Humans and Animals. *Princeton University Press* (2007).
9. Keeling, M.J. *et al.* Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science* **294**, 813-817 (2001).
10. Keeling, M.J. *et al.* Modelling vaccination strategies against foot-and-mouth disease. *Nature* **421**, 136-142 (2003).
11. Keeling, M.J. The implications of network structure for epidemic dynamics. *Theoretical*

Population Biology **67**, 1-8 (2005).

12. Keeling, M.J. and Eames, K.T.D. Networks and epidemic models *J. Roy. Soc. Interface* **4** 295-307 (2005).

13. Kermack, W.O and McKendrick, A.G. A contribution to the mathematical theory of epidemics, *Proc. Roy. Soc. A* **115**, 700-721 (1927).

14. Le Menach et al. Modeling spatial and temporal transmission of foot-and-mouth disease in France: identification of high-risk areas. *Vet Res* **36** 699-712 (2005).

15. Lipsitch M. *et al.* Transmission dynamics and control of severe acute respiratory syndrome. *Science* **300**, 1966-1970 (2003).

16. Mansley L.M. *et al.* Early dissemination of foot-and-mouth disease virus through sheep marketing in February 2001. *Veterinary Record* **153**, 43-50, (2003).

17. May R.M and Lloyd A.L. Infection dynamics on scale-free networks. *Phys. Rev. E.* **64**, 066112 (2001).

18. Sander, L.M. *et al.* Percolation on heterogeneous networks as a model for epidemics. *Mathematical Biosciences* **180**, 293-305 (2002).

19. Savill, N.J. *et al.* Silent spread of H5N1 in vaccinated poultry. *Nature* **442**, 757 (2006).

20. Tildesley, M.J. *et al.* Optimal reactive vaccination strategies for a foot-and-mouth outbreak in Great Britain *Nature* **440**, 83-86 (2006).

21. Tildesley, M.J. *et al.* Accuracy of models for the 2001 foot-and-mouth epidemic. *Proc. Roy. Soc. B* **275** (1641), 1459-1468 (2008).

22. Vazquez, A., Boguna, M., Moreno, Y., Pastor-Satorras, R., Vespignani, A. Topology and correlations in structured scale-free networks. *Phys Rev E Stat Nonlin Soft Matter Phys.* 2003 **67**(4 Pt 2), 046111 Epub (2003).

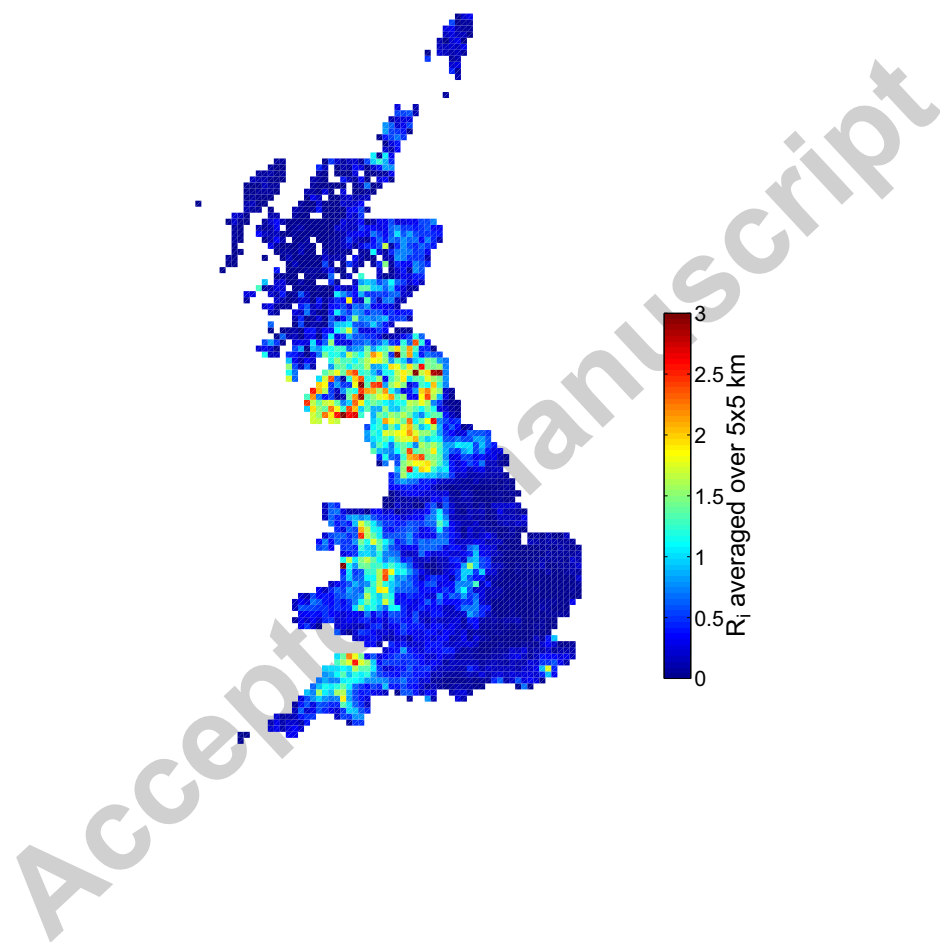
Figure Legends

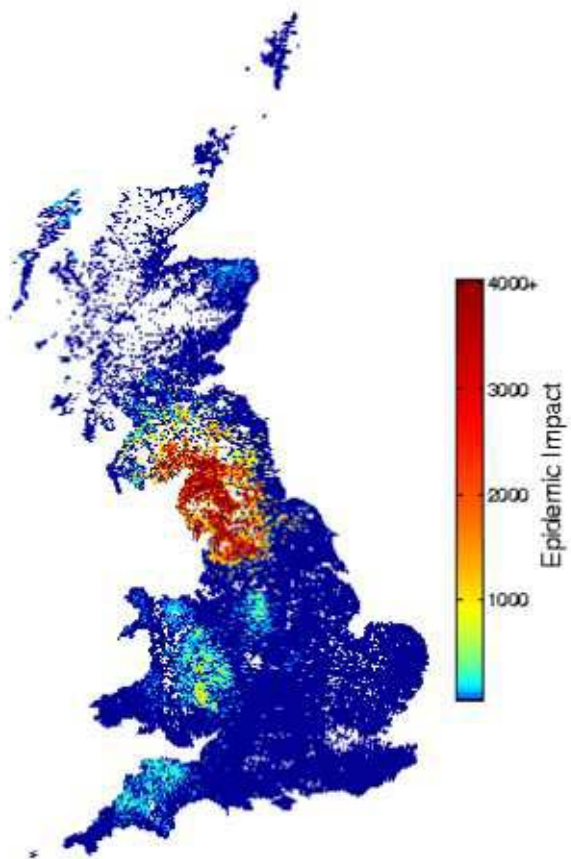
Figure 1 Spatial plots of (a) R_i , (b) Epidemic Impact given that an epidemic is seeded on an individual farm and (c) the probability of an epidemic occurring from infection on an individual farm. In (a) the colour scale shows the average value in 5×5 km grid squares. In (b) and (c) the colour scale shows the value for each individual farm. In (c) infection results in an epidemic if 10 or more farms are infected.

Figure 2 Graph of $R_i^{(2)}$ against R_i . The colour scale shows the epidemic impact for each value of R_i and $R_i^{(2)}$. Each point corresponds to the R_i and $R_i^{(2)}$ values for a single farm in the UK. The black line is shown for $R_i = R_i^{(2)}$.

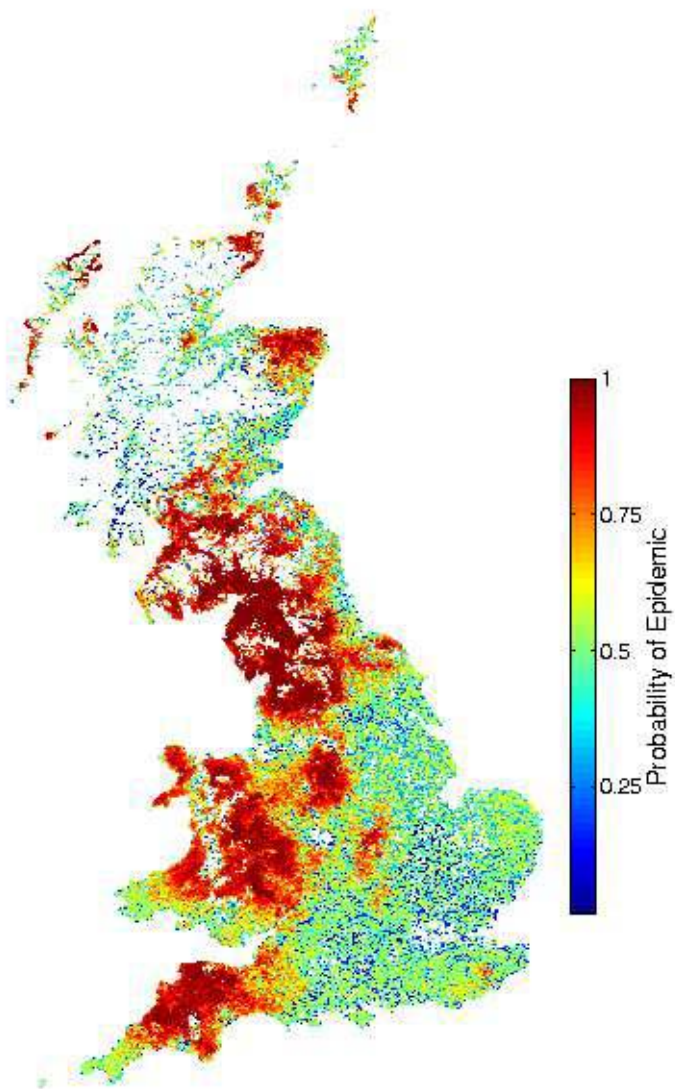
Figure 3 (a) The probability of an epidemic against R_i and (b) Epidemic Impact against $R_i^{(2)}$. In both graphs 95% confidence intervals on epidemic impact are shown. The shaded region shows the 95% confidence intervals of the mean of the y-axis quantity. Graph (c) shows the errors in predicting epidemic impact when errors are calculated based upon R_i and $R_i^{(2)}$ values of each farm, where these values are averaged over rings of varying radii.

Figure 4 Spatial plot of the predicted epidemic impact based upon the value of $\overline{R_i^{(2)}}(58\text{km})$ multiplied by the predicted probability that an epidemic occurs. Each point corresponds to the predicted epidemic impact for an epidemic seeded on each individual farm in the UK. An epidemic is defined as occurring if 10 or more farms are infected.





ed manuscript



Manuscript

Figure 2

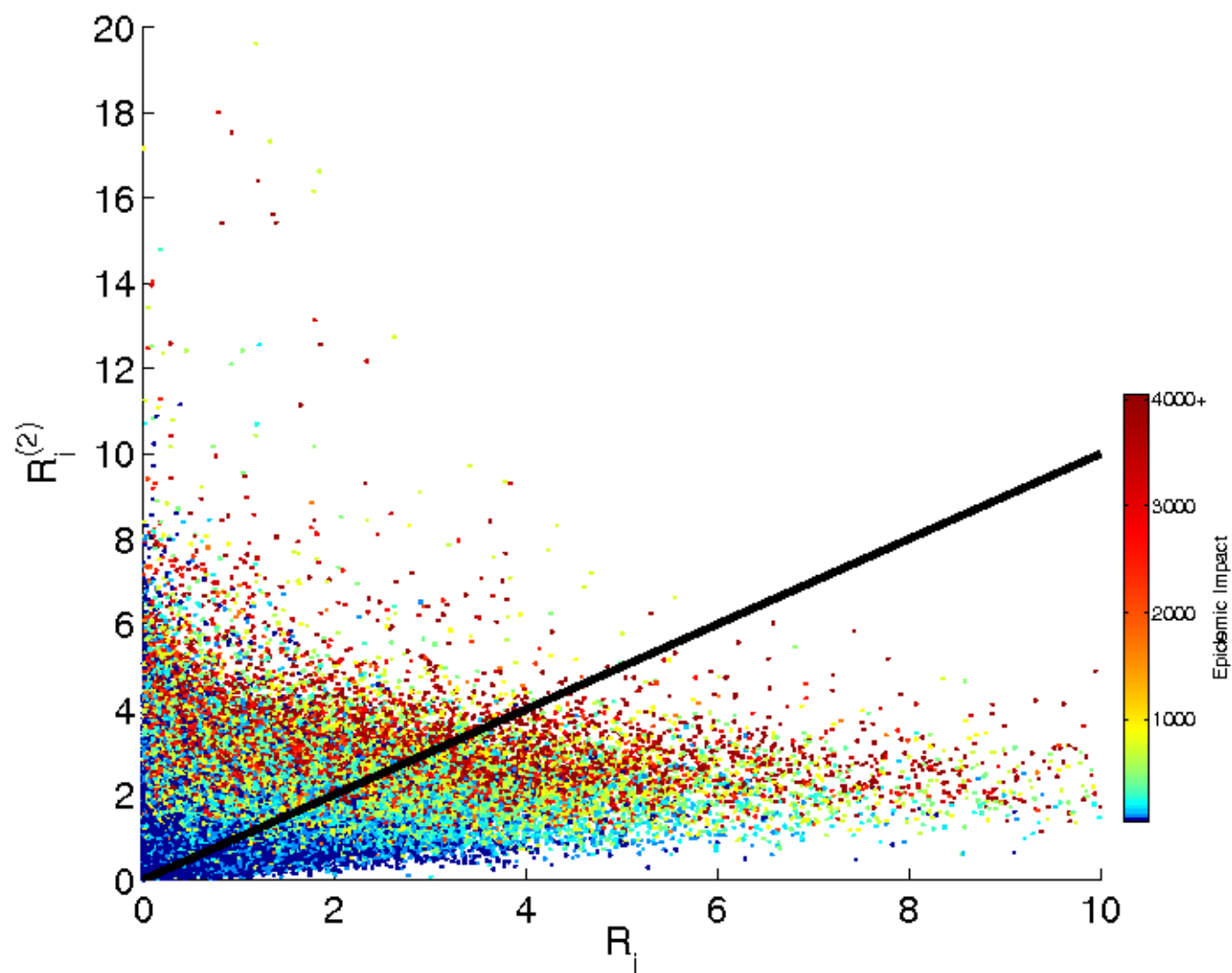


Figure 3a

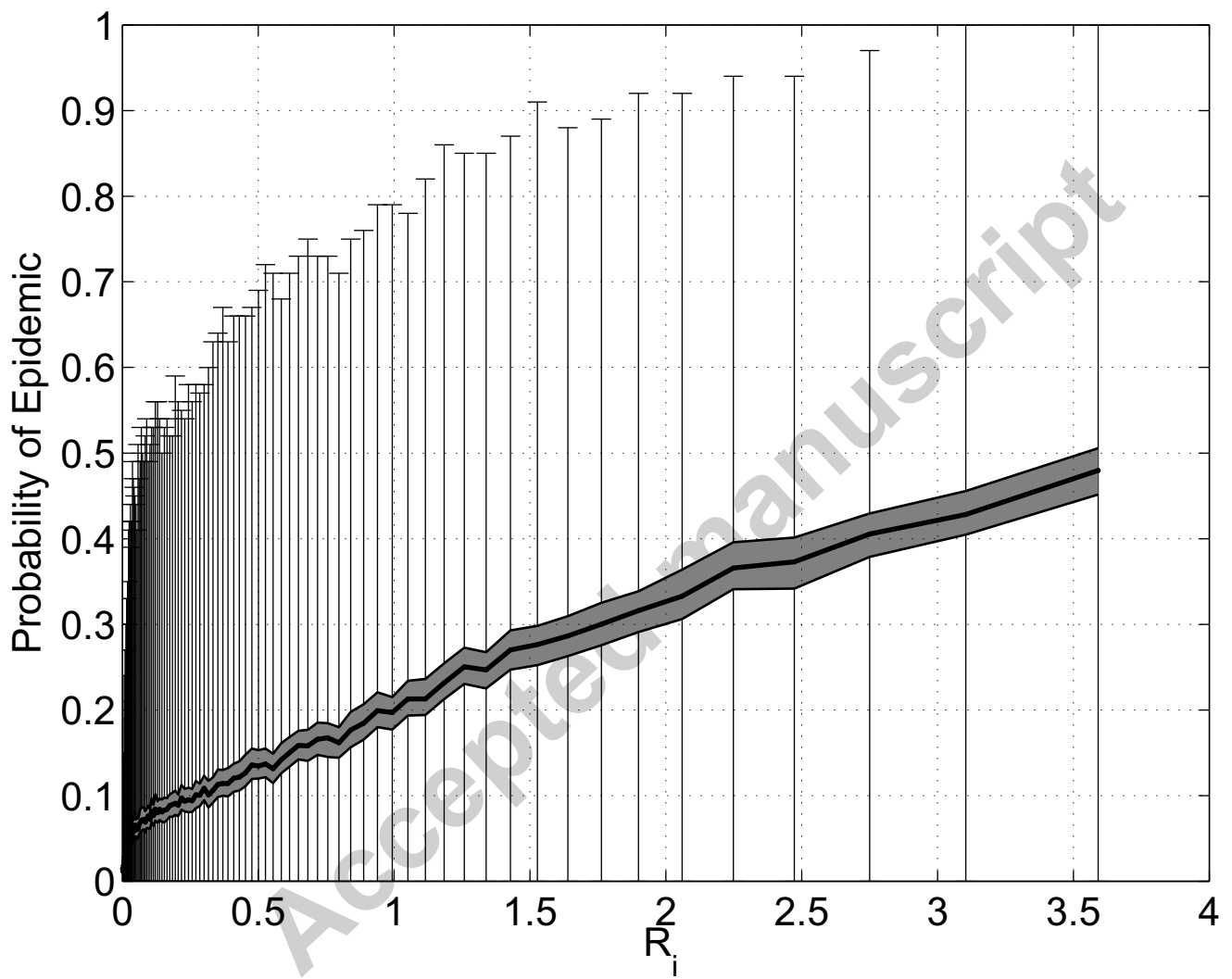


Figure 3b

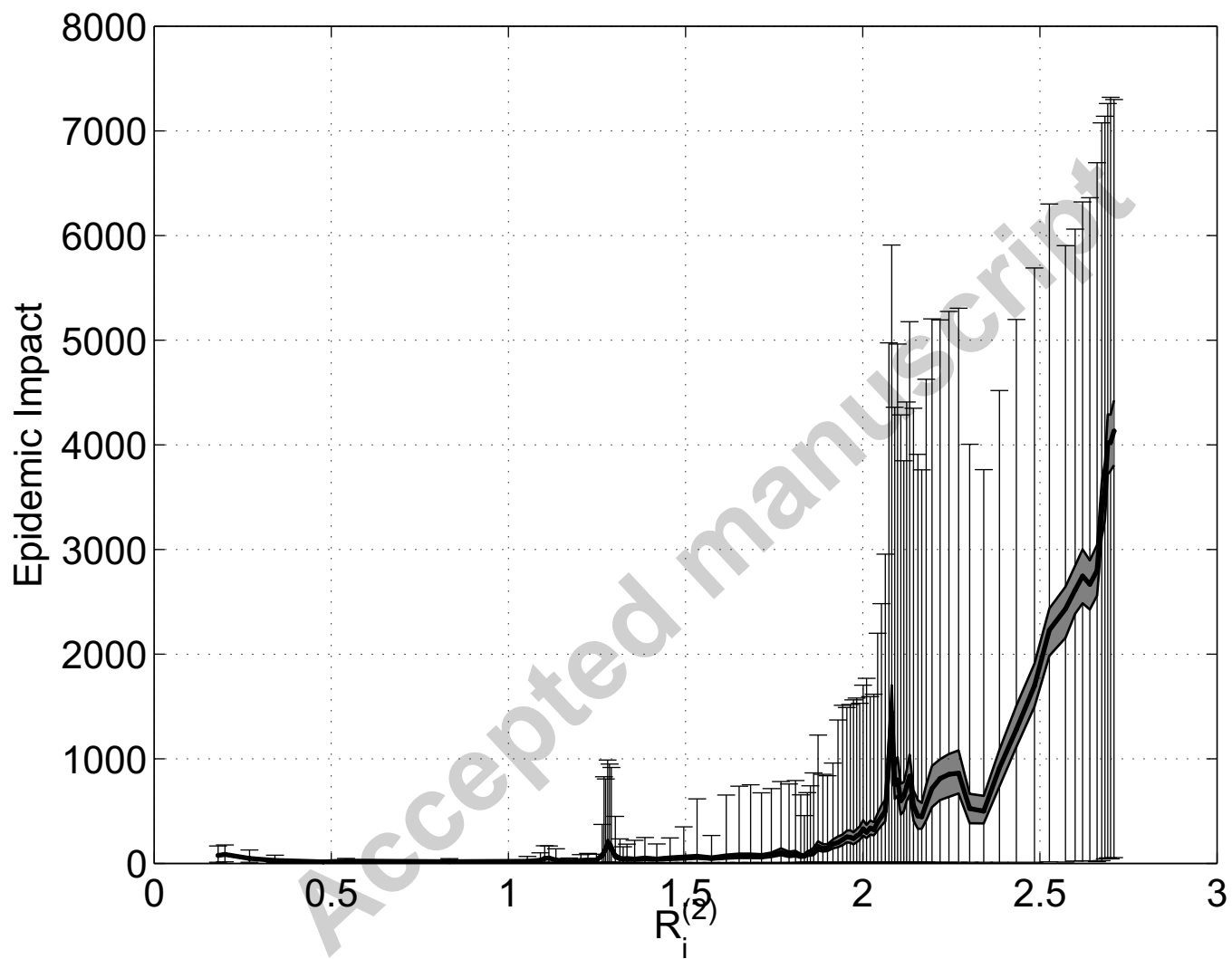
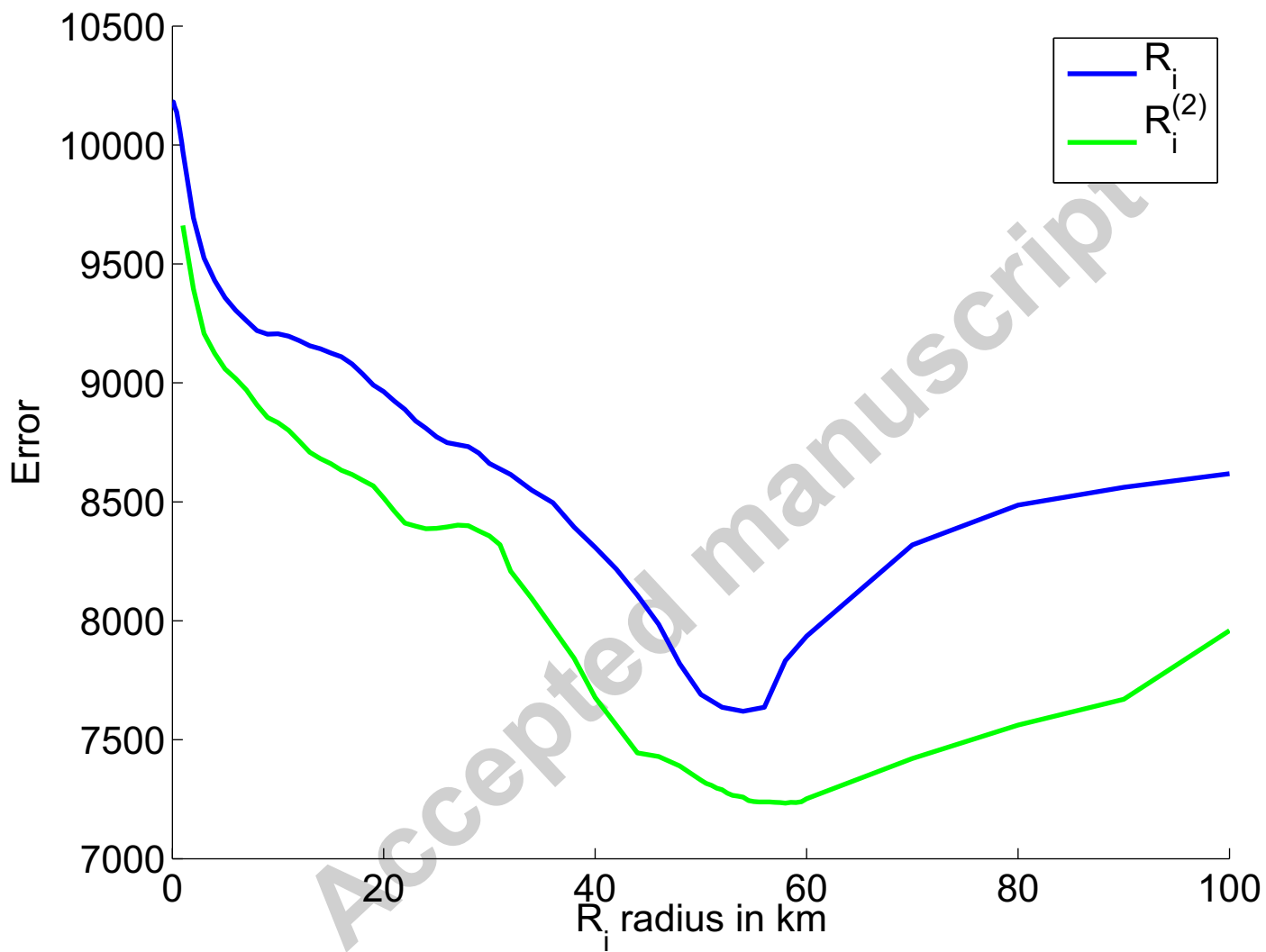


Figure 3c



Accepted manuscript

