Descriptive spatial analyses of *Mycobacterium bovis* infection in badgers *Meles meles* and cattle in four areas in Ireland.

Gabrielle E. Kelly [a*], Simon J. More[b], David H. Williams[a], Guy McGrath[b]

[a]UCD School of Mathematical Sciences, University College Dublin, Belfield, Dublin 4, Ireland.

[b]UCD Centre for Veterinary Epidemiology and Risk Analysis, University College Dublin, Belfield, Dublin 4, Ireland

Correspondence: Gabrielle Kelly, UCD School of Mathematical Sciences, University College Dublin, Belfield, Dublin 4, Ireland. (fax 00-353-1-7167182; email: Gabrielle.Kelly@ucd.ie)

Running title: Spatial association of TB in badgers and cattle in Ireland

Word count: 6,973

Summary

1. In Ireland, the herd prevalence of bovine tuberculosis has remained stable for several decades and efforts to eradicate the disease have not been successful. The disease has been linked to infected badgers *Meles meles*, a nationally protected species. To inform disease control, local geographical associations between *Mycobacterium* infections in cattle and badgers need to be explored further, as has been done in the UK.

2. Using data from a large-scale field trial in Ireland, the Four Area Project, K-functions and associated intensity functions were used to assess clustering of infection and the scale at which it occurs within and between cattle and badger populations. Kernel probability estimation was used to examine whether badgers with the same strain type of *M. bovis* were spatially segregated. Nearest neighbour analysis was also carried out to facilitate comparison with other studies.

3. *M. bovis* infections were locally clustered within the badger population at distances up to 8 km in 3 areas but not significantly in Donegal. Badgers infected with the same strain type were particularly closely associated in three of the four areas.

In the cattle population *M. bovis* infections were clustered at distances up to 6 km in Cork and 2 km in Donegal but not in the other two areas.

4. *M. bovis* infections in badgers and cattle were significantly spatially associated at distances up to 8 km in Cork, 4-8 km in Kilkenny and 5-8 km in Monaghan. Similar patterns of spatial associations between badgers and cattle were found to those in the Randomised Badger Culling Trial in England making the results in both countries very robust, despite differences in badger ecology and cattle management.

5. Synthesis and applications. The study provides evidence of local transmission of *M. bovis* infection between badgers and cattle in Ireland, at distances up to 8 km. and highlights the need to examine distances at which badger culling is carried out in cattle TB control policy.

Key-words: badger, bovine TB, Four Area Project Ireland, disease clustering and mapping, K-function, intensity function, RBCT.

Introduction

Bovine tuberculosis is a disease that affects cattle in Ireland where like the UK this disease is also present in the badger, wildlife species *Meles meles*. Large-scale field trials, the East Offaly Project and Four Area Project (FAP) in Ireland and the Randomized Badger Culling Trial (RBCT) in Britain found lower levels of bovine TB in areas subject to extensive badger culling than in matched reference areas where little or no experimental culls occurred (O' Mairtin *et al.* 1998; Griffin *et al*. 2005; Donnelly *et al*. 2006). In Ireland, the primary strategy for the control of cattle TB is to rapidly identify infected cattle, to restrict the movement and trading of cattle from infected herds and to test cattle on adjoining farms (Griffin *et al*. 2005). Where the epidemiological investigation has concluded that badgers are involved, culling in the immediate vicinity of TB-affected herds, known as localized or reactive culling, is also being conducted in the short to medium-term, pending the development of a TB vaccine for badgers (More and Good, 2006).

In an observational study over 16 years in the Irish midlands, Kelly *et al.* (2007) found a reduction in confirmed cattle herd TB incidence on land adjoining proactive culled areas compared to land further outside ($\leq$ 2 km , $\geq$ 2 km outside respectively). In contrast, the (randomized) RBCT indicated, that in lands $\leq$ 2 km outside culled areas and in lands subjected to localized culling only, cattle TB incidence was raised (Donnelly *et al*. 2003, 2006). However, a more recent study (Donnelly *et al*. 2007), indicated that this

detrimental effect diminishes with successive annual culls. Local and extensive badger culling differ primarily scale. The question at what scale culling is likely to be most effective does not have definitive answers at this time in Ireland.

We investigated geographical associations between *M. bovis* infection in badgers and cattle, using data from the FAP. Earlier work by Olea-Popelka *et al*. (2003; 2005) examined data over all years of the project. One of our purposes was to investigate if, by restricting the time frame and cattle and badger populations studied, definitive answers to questions of spatial association of *M. bovis* infection in badgers and cattle, as well as spatial clustering in both badger and cattle populations at a local level, could be established in the Irish context. The scale at which it occurs is also investigated, which relates to the question of culling scale. Thus we seek to provide insight into disease behaviour and control policy.

An analysis of these data, using the methods of Woodroffe *et al*. (2005) for the spatial analysis of the RBCT, is presented in Appendix S1 in Supplementary Material.

Differences in badger and cattle densities and infection rates between the FAP and RBCT are some of the factors that may affect results. By delineating similarities and differences between the two studies, it is hoped to inform policy on control strategies in both countries.

Materials and Methods

Study populations

The data for this study are drawn from the FAP, a formal badger removal project undertaken in Ireland from September 1997 to August 2002, to assess the effect of badger culling on the incidence of bovine TB. The study design and its results are published in detail in Griffin *et al*. (2005). Briefly, the FAP was conducted in matched removal and reference areas (average area of 245 km^2) in four counties in Ireland: Cork, Donegal, Kilkenny and Monaghan. In addition, where natural barriers were absent, 'buffer areas' were created, up to 6km in width, at the boundary of each selected removal area and shall be referred to as outer removal areas. Badger removal was intensive and proactive throughout the study period in the removal areas (inner and outer), but reactive (cull only those badgers spatially associated with farms that had experienced severe tuberculosis outbreaks in cattle) in the reference areas. We investigate spatial associations between *M. bovis* infections in badgers and cattle in the proactive removal areas of the FAP. Analyses here were restricted to the first year of the FAP, since the numbers of badgers captured in subsequent project years were too small to permit substantive analysis. Also years were not amalgamated, to avoid possible distorting effects of recent badger culling on the distribution of infection (Tuyttens *et al*. 2000, Woodroffe *et al*. 2006) and to permit comparison with other studies. The effect of badger culling on spatial associations between cattle herds over time, as discussed in Jenkins *et al.* (2007), will be considered in a later paper.

Badger population structure

Data were available on badgers culled in the removal areas throughout the study period. Records were complete for 2359 badgers regarding the date of capture, geographical area and specific sett from where the badgers were captured. The tuberculosis status was known for 2305 of these. The sett identifications used were based on surveys conducted as part of the FAP and the geographical position of the sett at which badgers were caught was recorded in a GIS database. Tuberculosis status of badgers was based on culture results. As noted above, we restricted our analyses to the first year during which two or three culls were carried out in each county. The dates of the last culls in that year 1998 were in late May or June varying with county. There were a total of 1,113 badgers culled in the proactive areas in the first year of the cull. Of these 15 were cubs i.e. badgers born in the previous 12 months, 1,069 were adults (157 of which were yearlings) and 29 badgers of undetermined age. Badgers without age data were excluded from the analysis, as were those without sex data or infection status and the 15 cubs (74 in total). We thus consider for study, the total of 1,039 adult badgers culled, 304 in the outer removal areas and 735 in the inner removal areas for which the infection status, sex and age were known.

Cattle population structure

Only those herds that had all their land contained in the removal or reference areas of the project (as in Griffin *et al*. 2005) were included in the analysis. The geographical

locations associated with cattle herds were based on the centroids of the land fragments owned by the farmer and recorded in a GIS database. During the period of the FAP all cattle herds in the study areas were tested at least annually using the single intradermal comparative tuberculin test (SICTT). A positive animal is then examined for lesions at slaughter. In addition animals sent for routine slaughter to a factory are also examined for TB-lesions. We included only those herds in the analysis, which had at least one full herd test in the first twelve months of the project. We defined a herd to be TB positive if it contained cattle that were: positive on the SICTT, TB-lesioned or both.

Diagnostic procedures for badgers and cattle

The TB status of culled badgers was established as described in Costello *et al*. (1997). Tuberculin reactors found in cattle herds were examined at slaughter and cultured (Costello *et al*. 1997). Strain typing was performed on badgers and cattle as described in Costello *et al*. (1999). However, *M*. *bovis* strain typing data from Irish reactor cattle during the FAP was very limited and as omissions could seriously bias results, only strain data for badgers were considered for analysis here. Appendix S1 has further details.

Statistical methods

Spatial associations of infection based on distances badger-badger, cattle-cattle and badger-cattle were investigated. As a herd may be located on multiple parcels of land, the calculation of distances is not straight-forward and the following methods were used throughout: the distance between two herds A, B, is found by taking the minimum distance between the centroids of all parcels in herds A to all parcels in herd B; the distance between a badger and a herd is found by taking the minimum distance between the centroids of all parcels in the herd to the badger capture location. An alternative distance method is discussed in Appendix S1.

K-Functions and Intensity functions

Spatial patterns of infection were visualized using K-functions and intensity functions, based on the theory of spatial point processes (Ripley, 1981); here the points are locations of herds or badger setts. We consider points of two different types, infected points labelled j = 1, and non-infected labelled j = 2. If we calculate all distances between points of type i and j of the pattern, then the K-function $K_{ij}(d)$ is proportional to the distribution function of these distances and the intensity function $I_{ij}(d)$ to the density function. By spatial clustering we mean a general tendency for cases, i.e. infected points, to occur more closely together than would be compatible with random sampling from the population at risk. No spatial clustering is equivalent to the random labeling hypothesis H, where the type j points constitute a random thinning of the unlabelled point process defined as the superposition of type 1 and type 2 events. Under H, $K_{11}(d) = K_{22}(d) = K_{12}(d)$ for all *d*. Departures from H are assessed by the significance of the estimated

differences $\hat{D}(d) = \hat{K}_{11}(d) - \hat{K}_{22}(d)$ and $I\hat{D}(d) = \hat{I}_{11}(d) - \hat{I}_{22}(d)$. The latter has the advantage of showing the exact distances at which clustering occurs but will have wider confidence bands. D(d) can also be interpreted as an expectation: $\lambda_1 D(d)$ represents the expected number of excess cases within a distance d of a typical case, by comparison with the number expected in the absence of clustering (where λ_1 is the intensity of type 1 points). The estimates $\hat{K}_{11}(d)$ and $\hat{K}_{22}(d)$ were computed as in Diggle & Chetwynd (1991) and then the estimates $\hat{I}_{ij}(d) = \hat{K}_{ij}(d+1) - \hat{K}_{ij}(d)$ were calculated i.e. by approximating the derivatives of the K-functions using a bandwidth of 1 km. All estimates were adjusted for edge effects as in Besag (1977). The null sampling distribution of $\hat{D}(d)$ and $I\hat{D}(d)$ was found by carrying out 99 Monte Carlo simulations in each of which disease labels were randomly assigned to points. Differences in K-functions D(d) (and thus also ID functions) tend to a positive constant as $d \to \infty$ and Diggle (2003, Chap. 5) suggests the statistical information is greatest at small values of d, quite apart from the limitations imposed by the physical dimensions of the region under study. Appendix S2 has further details of K-functions, intensity functions and edge-corrections.

Spatial variation in risk – kernel probability maps

In the case of strain data for infected badgers it is of interest to know if badgers with the same strain cluster. Spatial variation in risk was examined as described in Diggle (2007). With s strain types we denote the probability a case at location x will be of strain type j,

conditional on there being one of the s types at x, by $p_j(x)$. We say there is spatial segregation if the area can be partitioned approximately into sub-regions where one strain type predominates i.e complete segregation is if at each x in the sub-region, $p_j(x) = 1$ for one of the j. We used the kernel estimator of $p_j(x)$, $\hat{p}_j(x)$ given in Diggle (2007) where the smoothing parameter for $\hat{p}_j(x)$ is chosen by cross-validation. An ad-hoc statistic for a test of clustering is then given by,

$$T = \sum_j \sum_{x \varepsilon D} (\hat{p}_j(x) - \bar{p}(x))^2 \text{ where } \bar{p}(x) = \sum_j \hat{p}_j(x)$$

The significance level was obtained by Monte-Carlo simulation where labels were randomly assigned to strain types. The estimated type specific probability surfaces, $\hat{p}_j(x)$ versus x, were plotted for each county using ArcView 9.2 (ESRI Inc., Redlands, CA).

Statistical analyses

Statistical analyses were carried using SAS® software (SAS Institute, Cary, NC) and R (2007).

Logistic regression was used to examine associations between infection status of badgers and covariates such as age and sex.

K-functions and intensity functions from infected badgers to infected badgers and non-infected badgers to non-infected badgers were calculated and similarly for cattle. For badger-cattle, as there are two animals and two events, K-functions and intensity functions were calculated between (1) infected badgers and infected cattle; (2) non-

infected badgers and infected cattle; (3) infected badgers and non- infected cattle; and (4) non-infected badgers and non-infected cattle. Because these K-functions all involved two types of events they were calculated as in Diggle and Chetwynd (1991, equation (6); see also Appendix S2). Two D functions as well as two ID functions were then generated: (a) (1) – (2) and (b) (1) – (3). Confidence limits were generated (a) by considering only infected herds, permuting the badger disease labels and thus recalculating the K-functions and I functions and (b) only considering infected badgers, permuting the herd disease labels and thus recalculating the K- and I functions. D and ID values outside the upper confidence limit indicate clustering of infection.

Spatial variation was examined for the two most common strains in three counties. In Kilkenny, however, three strains were examined, as two of these were equally common (Table 2).

An alternative analysis of these data using the methods of Woodroffe *et al.* (2005) is presented in Appendix S1, for comparison purposes with that study.

Results

Figs S1&.S2 plot the spatial location of infected and non-infected badgers and cattle in each area.

BADGERS

Of the 1039 adults studied, 447 (43.0%) were male. The adult sex ratio was female biased in all counties (Table 1). Table 1 presents the prevalence of *M. bovis* infection in both male and female badgers. Using a logistic regression analysis we found prevalence varied substantially between areas ($p < 0.001$). There was no significant effect of sex or interaction between sex and area on the risk of *M. bovis* infection. Thus prevalence was the same for the sexes within each area and for the sexes overall. There were 572 capture locations, 61% of which were main setts. The estimated number of badgers per capture location is thus 1.82.

Clustering of infection within the badger population

Figs 1-4 display the difference in intensity functions for infected badgers and uninfected badgers for each of the four areas. A stabilization of the difference D(d) occurred at approximately d = 8 km in all counties. The figures show clear evidence of clustering of infection in badgers at all distances up to 8 km in counties Cork, Kilkenny and Monaghan, with differences in intensity functions above the confidence bands. For Donegal the difference in K-functions was positive up to about 4 km and significantly so up to 2.5 km, while there was no significant difference in intensity functions. The estimated number of excess infected badgers within a distance d of a typical case, by comparison with the number expected in the absence of clustering was about 35, 1 and 8 for all values d in Cork, Donegal and Kilkenny respectively, and in Monaghan it went from 14 at 3 km to 21 at 9 km.

Associations between strain types of *M. bovis* among badgers

Distance calculations related to strain types were done by two methods. In method 1, each badger contributed one observation for each strain type. In method 2, a capture location contributed one observation for each strain type as multiple strain types were occasionally found at the same location (Table 1). Of the 209 infected badgers under study, strain type information was available on 204. In all counties several strain types of *M. bovis* were found to infect badgers. Table 2 shows the distribution of each strain type for badgers by area. Fig.5 show the estimated type specific probability surfaces for the main strain types in Cork, Donegal and Kilkenny using method 1. The p-values associated with Diggle's test of spatial segregation were, Cork ($p < 0.001$), Donegal ($p = 0.128$), and Kilkenny ($p < 0.001$). In Monaghan the estimated smoothing parameter was so large that the estimated probability surface is constant and there is no spatial segregation ($p = 1.0$).

CATTLE

Summary statistics describing cattle populations in the 12 month period of the first year of culling are shown in Table 1.

Clustering of infection within the cattle population

Figs 1-4 display the difference in intensity functions for infected and uninfected cattle in the removal area of each county. The graphs show no evidence of clustering of infection among herds in Kilkenny and Monaghan, with all points lying within the confidence limits. The difference in intensity functions for Donegal is very variable due to the small number (9) of TB infected herds in that county but showed a difference up to 2 km. In Cork the difference in K-functions was positive, showing clustering of infection up to 8 km, significantly so up to 7 km. The differences in intensities were significant up to 6 km. The excess of infected herds at distances of 3 and 6 km from a reference case is 5 and 6 respectively, in Cork. This excess is negligible or non-existent in the other areas.

BADGERS-CATTLE

Associations between infections in badgers and cattle

In the analysis of distances from badgers to cattle, data from badgers captured at the same location (which shared the same distances to the nearest cattle herd) were condensed. A single location could contribute data both as TB infected (if one or more infected badgers were captured there) and as uninfected (if one or more uninfected badgers were captured there). The badger data thus consisted of 830 uninfected badgers at 491 locations and 207 infected badgers at 167 locations. Infected and uninfected badgers had comparable opportunities for contact with cattle. In the 12 months of the first year of badger culling, distances to the nearest herd were similar for infected and

uninfected badgers (median distance 0.55 km and 0.56 km respectively; p = 0.56, Wilcoxon rank sum test).

The differences in intensity functions and K-functions for badgers-cattle are displayed in Figs 1-4 and Fig.6, panels (c) and (d). There was clear evidence of clustering of infection from both panels. Figs 1-4 and Fig. 6 panel (c) are based on the difference (1) – (2), (i.e. differences in distances between infected and uninfected badgers to infected cattle; see Methods). The differences stabilized at about 5-6 km in all counties. Differences in K-functions showed clustering of infection in three counties (Donegal the exception) at distances up to 8 km. In Donegal there was a significant difference up to 1.7 km. Differences in corresponding intensity functions were significant up to 8 km for Cork, no differences in Donegal, at 4 and 6-8 km but not at other distances in Kilkenny and from 5 – 8 km in Monaghan. Figs 1-4 and Fig.6 panel (d) are based on the difference (1) – (3) (i.e. differences in distances between infected and uninfected cattle to infected badgers). The differences stabilized again at 5-6 km in all counties. Difference in the K-functions showed associations occurred at distances up to 8 km in all counties while differences in intensity functions were significant at 1 km in Cork and no differences in Donegal, Kilkenny and Monaghan. We note the difference based on a third comparison (1) - (4) (not shown) gave similar results to (1) – (2).

The excess of infected setts within d km of a reference cattle infection, than would be expected in the absence of clustering was about 36 for Cork at all distances, varied from 0.2 – 1.5 in Donegal, 5-10 in Kilkenny and 15-17 in Monaghan.

The excess of infected herds within d km of a reference badger sett infection, varied from 21-35 in Cork, 0.36-0.47 in Donegal, 6.7-7.7 in Kilkenny and 4-9 in Monaghan.

Discussion

BADGERS

The results show that tuberculosis clusters in Irish badger populations. Using two different methods (K-/intensity functions, nearest neighbour ratio approach in Appendix S1), we found significant evidence of clustering of infection in badgers in Cork, Kilkenny and Monaghan and weaker evidence from Donegal. The intensity functions also indicated that clustering occurred at all distances up to 8 km except in Donegal. We are uncertain as to the reason for reduced evidence of clustering in Donegal but note that it is geographically distinct with sea inlets being a key feature (Griffin *et al*. 2005). The excess of infected badgers due to clustering was highest in Cork which also had the highest rate of infected badgers per km^2. Reports listed below indicate the scale at which clustering was found may be affected by factors such as low badger density, increases in badger social group ranges due to culling, degree of contact between badgers and infection rates (Table 1). Several field studies have reported that badger removal operations lead to social disruption very soon after badger capture started and impact setts at some distance from the main capturing areas (O'Corry-Crowe *et al*. 1996; Tuyttens *et al*. 2000; Woodroffe *et al*. 2006). Two or three badger culls had taken place during the period considered here and some badger removal from Kilkenny and Monaghan occurred soon before the study began (Griffin *et al.* 2005). In addition, O'Corry-Crowe *et al*. (1993) report a less stable social structure in low density badger

populations. Further, mathematical models of infectious diseases (Becker 1989) suggest that prevalence of infection in animals is likely to be density dependent. The interaction of these factors – culling, density and infection rates is perhaps complex as suggested by two recent studies concerning the effects of perturbation on TB in badgers (Macdonald, Riordan & Mathews 2006; Woodroffe *et al*. 2007).

We believe that differences between this and earlier Irish work are as a result of methodological differences. Olea-Popelka *et al*. (2003) found minimal spatial clustering of tuberculosis in badgers using nearest neighbour methods, but their analysis did not adjust for the fact that negative badgers are more prevalent than positive ones and hence will be closer together.

Our results are in general agreement with reports from Britain of infection in badger populations (Cheeseman *et al*. 1981; Delahay *et al.* 2000; Woodroffe *et al.* 2005) although these did not investigate the extent of clustering.

In the RBCT analysis of Woodroffe *et al*. (2005), it was found that *M. bovis* infections were locally clustered within the badger populations; clustering was seen in nine of their ten trial areas (overall $p<0.001$). In terms of infection rates, we calculated an infection rate in this study of 0.17 badgers/km^2/year (1,039 adult badgers removed, area prevalence varying between 13 and 28%, overall prevalence 20%) compared with 0.29 badgers/km^2/year in the RBCT (2,699 adult badgers removed in the initial cull, area prevalence varying between 2 and 38%, overall prevalence 12%). Jenkins *et al.* (2007) in an alternative analysis of the RBCT data, noted that 40% of distances from an uninfected

badger to the nearest infected badger exceeded 1 km whereas the corresponding percentage for an infected badger to the nearest infected badger was about 20%. Similar percentages were found here for Cork, which had the highest infection rate in badgers (0.36 badgers/km^2/year, Table 1), and in this respect is the most similar to the RBCT. The spatial behaviour of infection in badgers in the FAP and RBCT appears to be similar, despite considerable differences in badger ecology (Smal 1995, O'Corry-Crowe *et al.* 1993), population density (Griffin *et al.* 2005, Donnelly *et al.* 2007) and infection rates (Table 1). There is undoubtedly a complex relationship between these factors, as noted above.

We found local associations between strains of *M. bovis* within the Irish badger population. We found that the main strains in three of the areas segregate, based on kernel probability estimates of strain specific probability surfaces. As reported previously (Olea-Popelka *et al.* 2005, Costello *et al.* 2006), certain strain types dominate in defined areas (A1A3A in Cork, A1A5A in Donegal, C1H1J in Kilkenny and B1C1C in Monaghan). There was a diversity of strain types from the same sett, explained perhaps by badger movement and densities, as described above. In addition, using a measure based on nearest neighbour distance ratios (see Appendix S1) clustering was seen but only when data from all counties were combined. Similar results using this type of measure were found by Olea-Popelka *et al.* 2005. However, they used a different reference group, used a subset of the badger setts and no formal statistical tests were carried out.

CATTLE

We found some clustering of infection in cattle populations using the K-function approach in Cork and Donegal, up to 7 km in Cork. We note again Cork had the highest rate of infected badgers (Table1). We note there are mechanisms of disease spread in cattle herds that are not spatially dependent (More and Good 2006).

Clustering of infection is also reported in cattle herds within the RBCT areas (Woodroffe *et al*. 2005, Jenkins *et al*. 2007). Woodroffe *et al.* found clustering of infection in nine of the ten trial areas and our results are consistent with Woodroffe using an analysis based on ratios (see Appendix S1 and Fig. S3), but although we found clustering overall using such an analysis we did not find it in all individual counties. In addition to badger density, herd density in both studies is quite different, 1.4 per km^2 here and 0.75 herds per km^2 in the RBCT (Woodroffe *et al.* 2005). Thus herd density is higher here and badger density and infection rates generally lower and these factors may contribute, in complex ways, to our results differing with the results of Woodroffe *et al*. (2005), as may the different method of analysis in that paper.

We note that considering infected herds as a random thinning of all herds, as is required for the K-function approach, may not be valid. For example, larger herds are more susceptible to infection (Griffin *et al*. 2005), and being located on large parcels of land, may be further from each other than smaller herds. This will have the effect of reducing the evidence for clustering of infection. However, the same criticism applies to the analyses conducted by Woodroffe *et al*. (2005).

In our analysis, cattle locations were represented by centroids of parcels of land, while in reality cattle can move to the boundaries of these parcels. We note, however, our findings are robust to the choice of distance method used when calculating distances involving cattle (Appendix S1).

BADGER-CATTLE

Using the K-function approach, there is a strong evidence of clustering of infection between badger and cattle populations. This is true at distances up to 8 km in all areas (except Donegal), a finding not previously reported or considered. The results from the intensity function differences are less strong. We note numbers of infected cattle are small, particularly in Donegal (Table 1). Thus confidence bands based on intensity functions are very wide and statistical significance difficult to obtain.

Using the distance ratio approach, we also found significant albeit weak, spatial association between infected badgers and infected cattle at distances as small as 1-2 km. Our results are consistent with those of Woodroffe *et al*. (2005) who found spatial associations of infections in badgers and cattle in 8 of 10 trial areas of the RBCT. The median distances from adult badgers to TB positive herds varied from 0.7 to 1.7 km over the four counties and are similar to those of Woodroffe *et al.* (2005) as shown in Fig.S4. In a study of spatial association in the Irish midlands (Martin *et al*. 1997) proximity to an infected badger sett was a risk factor for TB in cattle. In that study the median distance

from a herd to the nearest occupied sett was 0.9 km, while that to an infected sett was 1.3 km, distances comparable to here.

Conclusions

The data show evidence of spatial clustering of *M. bovis* infection within badger populations at a scale of up to 8 km. Moreover, badgers infected with the same strain of *M. bovis* were spatially segregated. There was some evidence of clustering within cattle populations using the K-function approach. The data provide clear evidence of a spatial association between *M. bovis* infection in cattle and badgers, at distances between 1 and 6 km. The extent of the associations found may in some analyses be limited by the dimensions of the study areas. While Griffin *et al*. (2005) demonstrated an association in a large scale trial, here we demonstrate local geographical associations exist. Based on the results here, the scale for culling to be both feasible and have optimum effect requires further study. The implication of clustering for control policy has been discussed in the British context (Macdonald, Riordan & Mathews (2006); Donnelly *et al.* 2006a, 2007; Woodroffe *et al*. 2005, 2006). In Ireland, we face ongoing challenges with TB control in an environment where badgers are a protected and valued wildlife species contributing to biodiversity and culture (Longley 2006) but are also an important reservoir of infection for cattle.

Acknowledgements

Thanks to the Department of Agriculture, Fisheries and Food, Ireland for providing the data for this study and in particular Eamon Costelloe of the Central Veterinary Research Laboratory for strain typing and pathology data.

References

Becker, N. G. (1989) *Analysis of infectious disease data.* Chapman and Hall, London.

Besag, J. E. (1977) Comments on Ripley's paper. *Journal of the Royal Statistical Society*, **B 39**, 193-195.

Cheeseman, C. L., Jones, G.W., Gallagher, J. & Mallinson, P.J. (1981) The population structure, density and prevalence of tuberculosis (*Mycobacterium bovis*) in badgers (Meles meles) from four areas in south-west England. *Journal of Applied Ecology*, **18**, 795-804

Costello, E., Quigley, F., O'Grady, D., Flynn, O., Gogarty, A., Mc Guirk, J., O'Rourke, J. & Griffin, J. (1997) Laboratory examination of tissues from badgers and cattle from the four area badger study for evidence of tuberculosis. *Selected Papers*, pp. 32-36, Tuberculosis Investigation Unit, University College Dublin,

Costello, E., O'Grady, D., Flynn, O., O'Brien, R., Rogers, M., Quigley, F., Egan, J. & Griffin, J. (1999) Study of restriction fragment length polymorphism analysis and spoligotyping for

epidemiological investigation of *Mycobacterium bovis* infection. *Journal of Clinical Microbiology*, **37**, 3217-3222

Delahay, R.J., Langton, S., Smith, G.C., Clifton-Hadley, R.S. & Cheeseman, C.L.. (2000) The spatio-temporal distribution of *Mycobacterium bovis* (bovine tuberculosis) infection in a high-density badger population. *Journal of Animal Ecology*, **69**, 428-441.

Diggle, P.J. & Chetwynd, A.G. (1991) Second order analysis of spatial clustering in inhomogenous populations. *Biometrics*, **47**, 1155-1163.

Diggle, P.J. (2003) *Statistical analysis of spatial point patterns*, 2nd edn. Arnold, London.

Diggle, P.J. (2007) Spatio-temporal point processes: methods and applications. *Statistical methods for spatio-temporal systems* (eds B. Finkenstadt, L.Held &V. Isham), pp. 1-47. Chapman & Hall/CRC, London.

Donnelly, C.A., Woodroffe, R., Cox, D.R., Bourne, J., Gettinby, G., Le Fevre, A.M., McInerney, J.P. & Morrison, W.I. (2003) Impact of localized badger culling on tuberculosis incidence in British cattle. *Nature*, **426**, 834-837.

Donnelly, C.A., Woodroffe, R., Cox, D.R., Bourne F.J., Cheeseman C.L., Clifton-Hadley, R.S., Wei G., Gettinby, G., Gilks, P., Jenkins, H., Johnston, W.T., Le Fevre, A. M., McInerney, J.P. & Morrison, W.I. (2006) Positive and negative effects of widespread badger culling on tuberculosis in cattle. *Nature,* **439**, 843-846.

Donnelly, C.A., Wei G., Johnston, W.T., Cox, D.R., Woodroffe, R., Bourne F.J., Cheeseman C.L., Clifton-Hadley, R.S., Gettinby, G., Gilks, P., Jenkins, H., Johnston, W.T., Le Fevre, A. M., McInerney, J.P. & Morrison, W.I. (2007) Impacts of widespread badger culling on cattle tuberculosis: concluding analyses from a large-scale field trial. *International Journal of Infectious Diseases*, **11**, 300-308.

Griffin, J.M., Williams, D.H., Kelly, G.E., Clegg, T.A., O'Boyle, I., Collins, J.D. & More, S.J. (2005) The impact of badger removal on the control of tuberculosis in cattle herds in Ireland. *Preventive Veterinary Medicine*, **67**, 237-266.

Jenkins, H.E., Woodroffe, R., Donnelly, C.A., Cox, D.R., Johnston, W.T., Bourne F.J., Cheeseman C.L., Clifton-Hadley, R.S., Gettinby, G., Gilks, P., Hewinson, R.G., McInerney, J.P. & Morrison, W.I. (2007) Effects of culling on spatial associations of *Mycobacterium bovis* infections in badgers and cattle. *Journal of Applied Ecology*, **44**, 897-908.

Kelly, G.E., Condon, J., More, S.J., Dolan, L., Higgins, I. & Eves, J. (2007) A long term observational study of the impact of badger removal on herd restrictions due to bovine TB in the Irish midlands during 1989-2004. *Epidemiology and Infection,* in press.

Longley, M. (2006) Badger; Between hovers; The Sett. *Collected poems*, p. 48, 167, 295. Jonathon Cape, London.

Macdonald, D. W., Riordan, P. & Mathews, F. (2006) Biological hurdles to the control of TB in cattle: A test of two hypotheses concerning wildlife to explain the failure of control. *Biological Conservation*, **131**, 268-286.

Martin, S. W., Eves, J.A., Dolan, L.A., Hammond, R.F., Griffin, J.M., Collins, J.D. & Shoukri, M.M. (1997) The association between the bovine tuberculosis status of herds in the East Offaly Project Area, and the distance to badger setts, 1988-1993. *Preventive Veterinary Medicine*, **31**, 113-125.

More, S.J. & Good, M. (2006) The tuberculosis eradication programme in Ireland: a review of scientific and policy advances since 1988. *Veterinary Microbiology*, **112**, 239-251.

Olea-Popelka F. J., Griffin, J.M., Collins, J.D., McGrath, G. &Martin, S.W. (2003). Bovine tuberculosis in badgers in four areas in Ireland: does tuberculosis cluster? *Preventive Veterinary Medicine* , **59**, 103-111.

Olea-Popelka, F.J., Flynn, O., Costello, E., McGrath, G., Collins, J.D., O'Keeffe, J.O., Kelton, D.F., Berke, O. & Martin, S.W. (2005) Spatial relationship between *Mycobacterium bovis* strains in cattle and badgers in four areas in Ireland. *Preventive Veterinary Medicine,* **71**, 57-70.

O'Corry-Crowe, G., Eves, J. & Hayden, R. (1993) Sett distribution, territory size and population density of badgers (Meles meles L.) in East Offaly. *The Badger* (ed T.J. Hayden), pp. 35-56. Royal Irish Academy, Dublin.

O'Corry-Crowe, G., Hammond, R., Eves, J. & Hayden, R. (1996) The effect of reduction in badger density on the spatial organization and activity of badgers Meles meles L. in relation

to farms in central Ireland. *Biology and Environment: Proceedings of the Royal Irish Academy*, **96B,** 147-158.

O'Mairtin, D., Williams, D.H., Griffin, J.M., Dolan, L.A. & Eves, J.A. (1998) The effect of a badger removal programme on the incidence of tuberculosis in an Irish cattle population. *Preventive Veterinary Medicine*, **34**, 47-56.

R Development Core Team (2007) *R: A Language and Environment for Statistics*. R Foundations for Statistical Computing, Vienna, Austria.

Ripley, B. (1981) *Spatial statistics*. Wiley, Chichester.

Smal, C. (1995) *The badger and habitat survey of Ireland*. Stationary Office, Government Publications, Dublin.

Tuyttens, F.A.M., Delahay, R.J., Macdonald, D.W., Cheeseman, C.l., Long, B. & Donnelly, C.A. (2000) Spatial perturbation caused by a badger *Meles meles* culling operation: implications for the function of territoriality and the control of bovine tuberculosis *Mycobacterium bovis*. *Journal of Animal Ecology,* **69**, 815-828.

Woodroffe, R., Donnelly C.A., Johnston, W.T., Bourne, F.J., Cheeseman, C.L., Clifton-Hadley, R.S., Cox D.R., Gettinby, G., Hewinson, R.G., Le Fevre, A.M., McInerney, J.P. & Morrison, W.I. (2005) Spatial association of *Mycobacterium bovis* infection in cattle and badgers *Meles meles*. *Journal of Applied Ecology*, **42**, 852-862.

Woodroffe, R., Donnelly C.A., Cox D.R., Bourne, F.J., Cheeseman, C.L., Delahay, R.J., Gettinby, G., McInerney, J.P. & Morrison, W.I. (2006) Effects of culling on badger *Meles meles* spatial organization: implications for the control of bovine tuberculosis. *Journal of Applied Ecology*, **43**, 1-10.

Woodroffe, R., Gilks, P., Johnston, W.T., Le Fevre, A.M., Cox D.R., Donnelly C.A., Bourne, F.J., Cheeseman, C.L., Gettinby, G., McInerney, J.P.& Morrison, W.I. (2007) Effects of culling on badger abundance: implications for tuberculosis control. *Journal of Zoology*, in press.

Table 1 Summary statistics describing (a) badgers captured in the initial 12-month period of proactive culling in the removal areas of the FAP and (b) cattle populations in the removal areas for the same period. .

(a) Badgers

	Trial area				
Number culled (%)	Cork	Donegal	Kilkenny	Monaghan	Total
Male	185	84	93	85	447
	(47)	(43)	(39)	(40)	(43)
Female	206	110	147	120	583
	(53)	(57)	(61)	(60)	(57)
Total	391	194	240	214	1039
Number infected (%)	109	27	30	43	209
	(28)	(14)	(13)	(20)	(20)
Number of capture locations	207	113	133	119	572
Number of capture locations with infected badgers (%)	81 (39)	23 (20)	25 (19)	38 (32)	167 (29)
Number of capture locations that were main setts	119	63	80	85	347
Number of *M. bovis* strains	13	5	8	11	
Area (km^2)	307	226	313	368	1214
Removal rate/km^2/year	1.27	0.86	0.77	0.58	0.86
Infection rate/km^2/year	0.36	0.12	0.10	0.12	0.17

(b) Cattle

Trial area	No. of TB tested herds	No.(%) of TB affected herds	Herd density/km^2/year	Infection rate/km^2/year
Cork	399	57 (14)	1.30	0.19
Donegal	369	9 (2)	1.63	0.04
Kilkenny	232	25 (11)	0.63	0.08
Monaghan	700	44 (6)	1.90	0.12
Total	1367	135 (10)	1.12	0.11

Table 2 Distribution of *M.bovis* strains in badgers captured in the initial 12-month period of proactive culling in the removal areas of the FAP.

	Trial area			
M. bovis strain	Cork	Donegal	Kilkenny	Monaghan
A1A1A	5	4	6	9
A1A1F	0	0	0	1
A1A3A	32	0	0	0
A1A5A	2	18	1	1
A1E2A	0	0	0	1
A2A1B	6	0	0	0
A4A1H	0	6	0	0
B1C1C	0	0	0	23
C1H1J	48	0	10	0
Other	15	5	6	5
Total	108	27	29	40

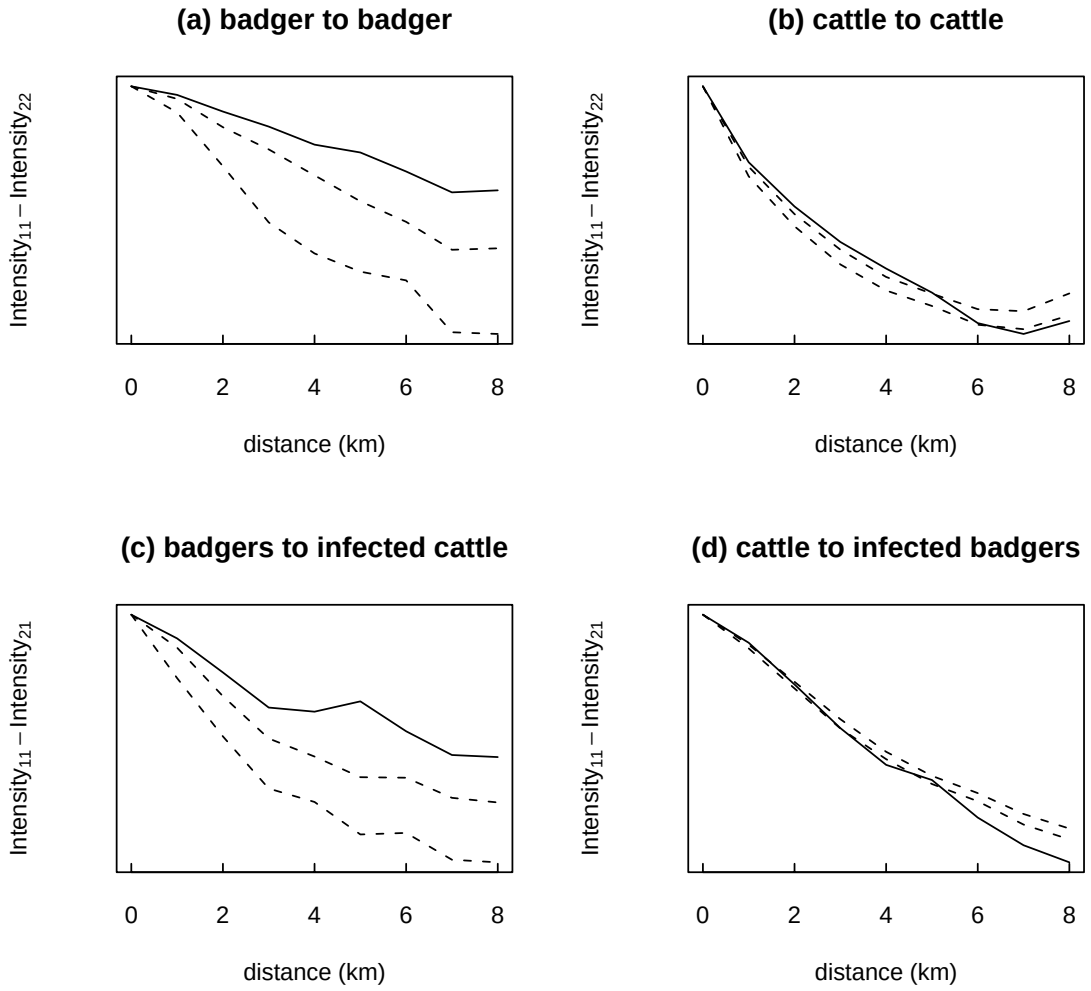

(a) badger to badger
(b) cattle to cattle
(c) badgers to infected cattle
(d) cattle to infected badgers
Intensity11 – Intensity22
Intensity11 – Intensity21
0
2
4
6
8
distance (km)

Fig.1.

Donegal

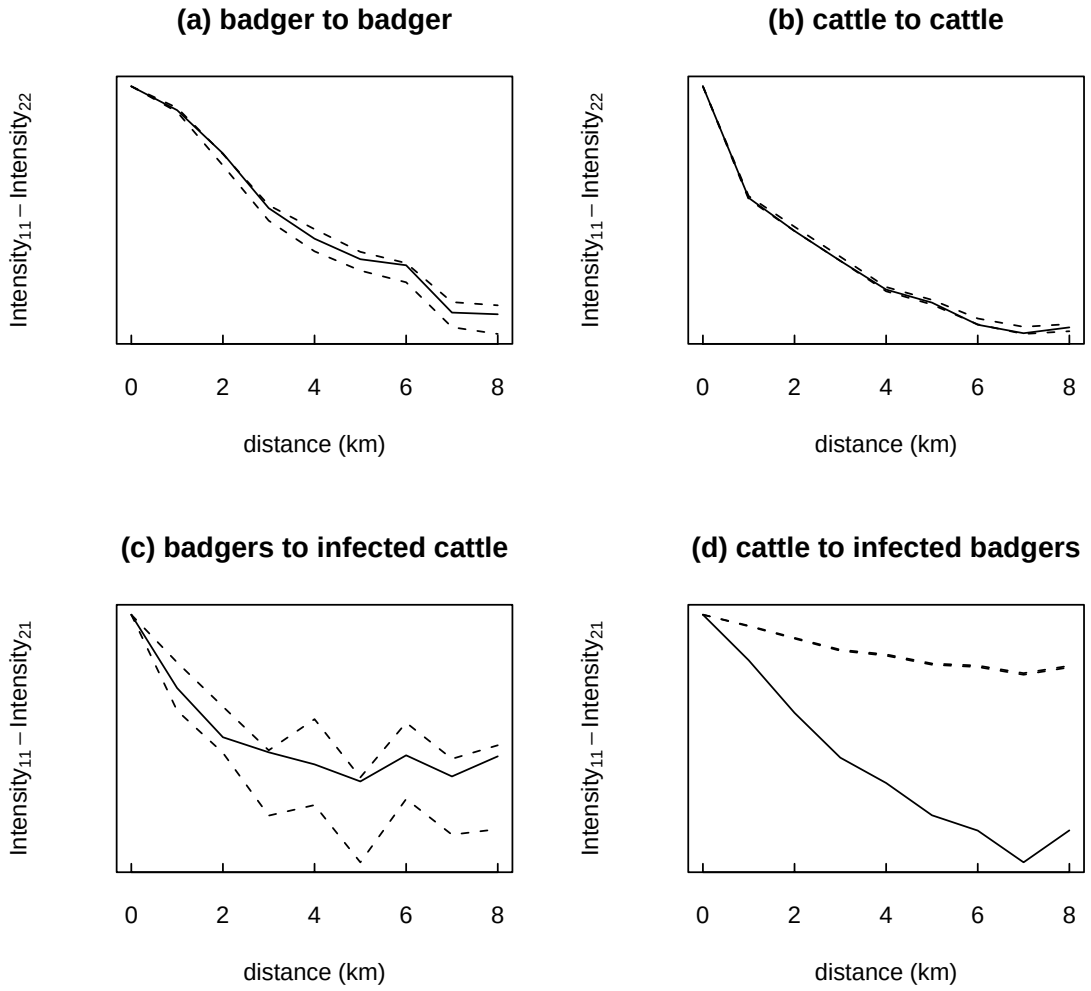


Fig.2.

Kilkenny

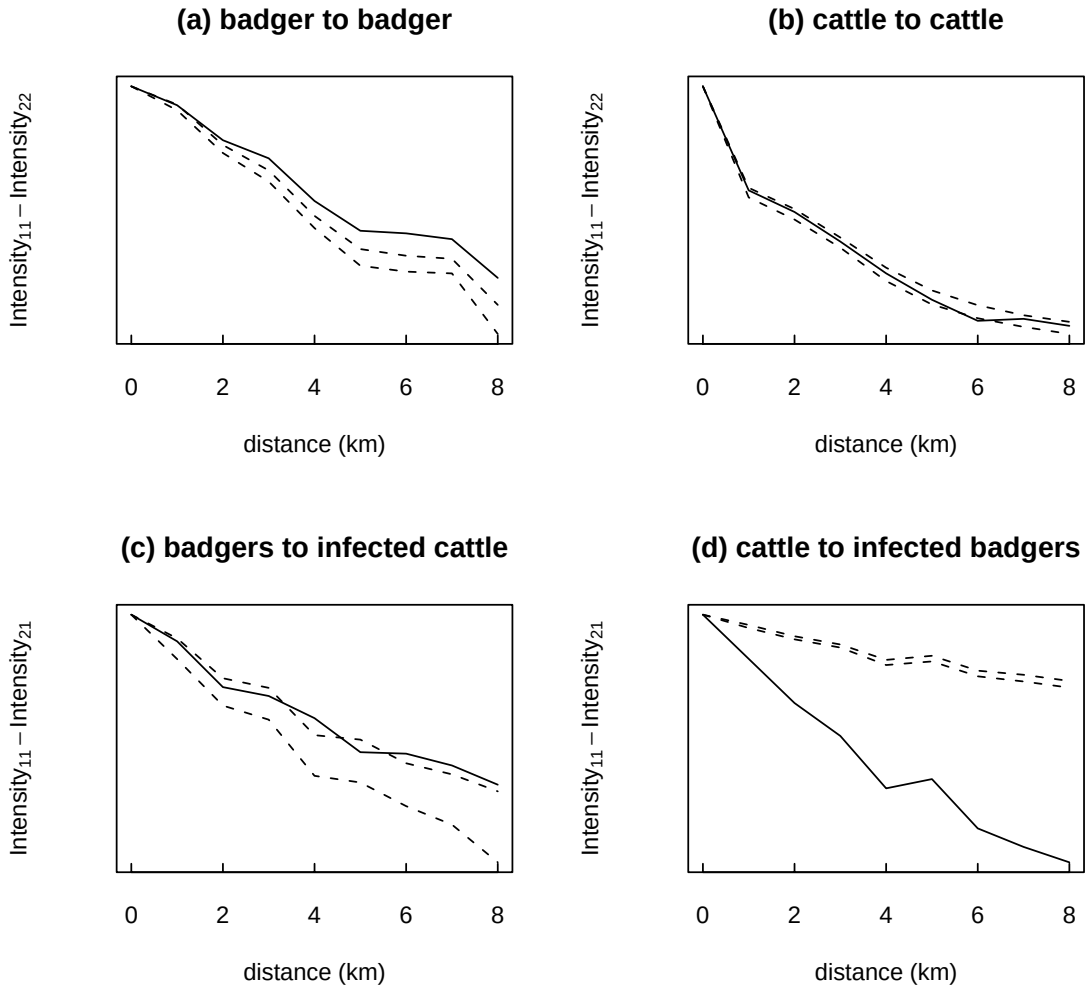


Fig.3.

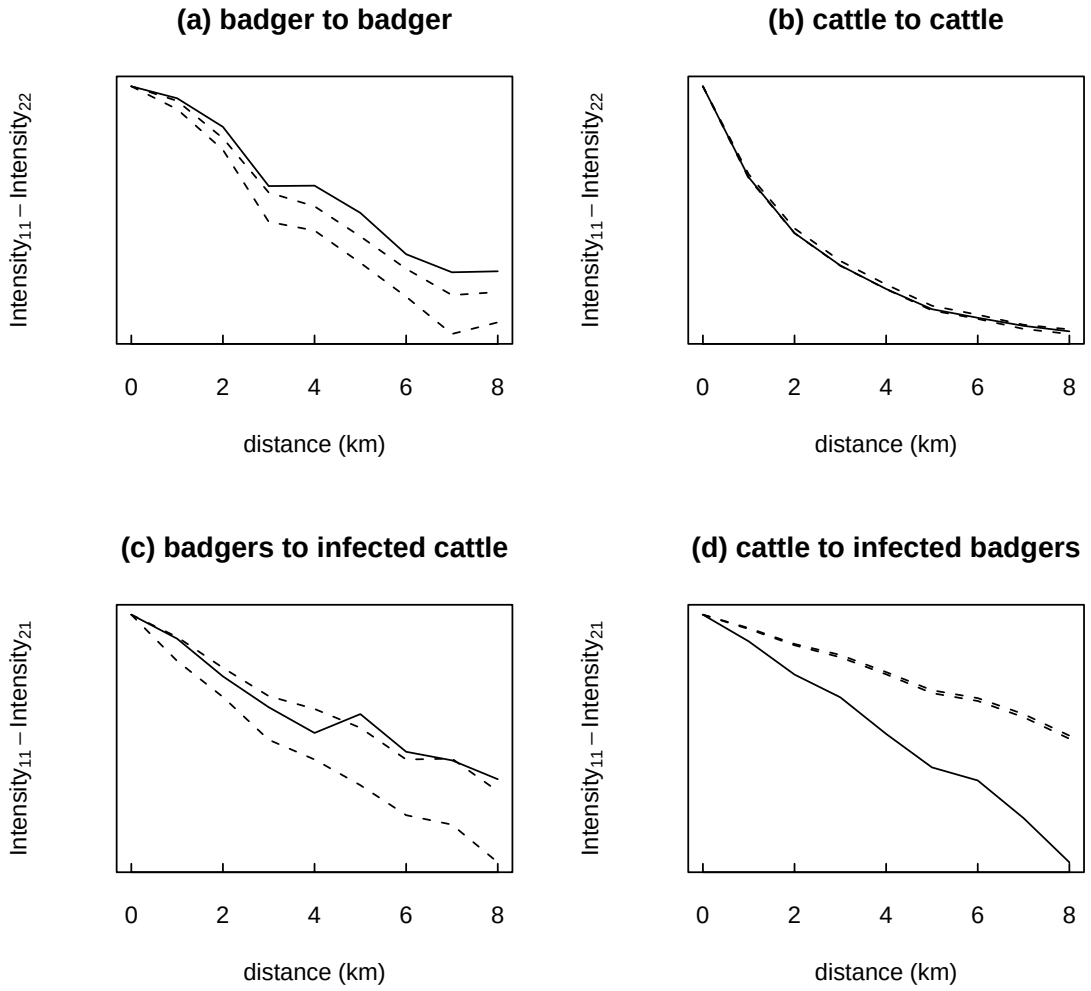
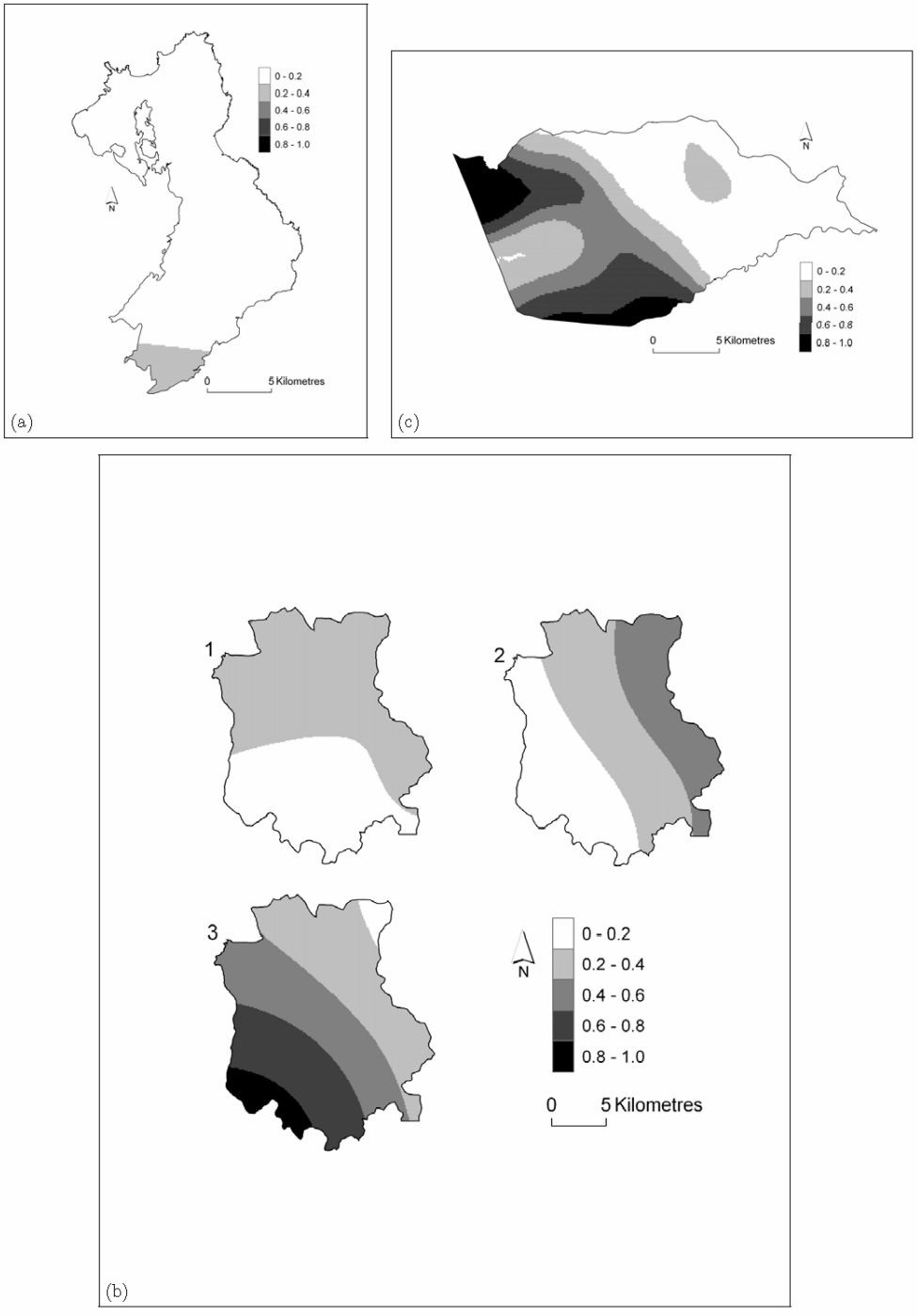


Fig.4.

Figs 1-4. Differences (with confidence bands dashed) in intensity functions of infected and uninfected animals. Animals were located in the removal areas of the FAP. Panels compare differences in intensity functions between infected and uninfected (a) badgers and (b) cattle. Panel (c) shows the difference between intensity functions of infected badgers to infected cattle and uninfected badgers to infected cattle while that in panel (d) shows the difference between infected cattle to infected badgers and uninfected cattle to

1 infected badgers. Distances involving herds were calculated using minimum distance
2 method. Infection status of herds was taken in the first year of badger culling. Badger
3 locations were condensed in the intensity functions (c) and (d). A single location could
4 contribute data both as TB infected (if one or more infected badgers were captured there)
5 and as uninfected (if one or more uninfected badgers were captured there).



1

Fig.5. Kernel estimates of badger strain specific probability surfaces in three areas of the FAP.

(a) Donegal: Kernel estimate of the A1A5A strain specific probability surface. The estimate for A1A1A is one minus this. $p = .128$ indicating some spatial segregation.

(b) Kilkenny: Kernel estimates of the A1A1A, A4A1H, C1H1J strain specific probability surfaces. $p < .001$ indicating spatial segregation.

(c) Cork: Kernel estimate of the A1A1A strain specific probability surface. The estimate for C1H1J is one minus this. $p < .001$ indicates spatial segregation.

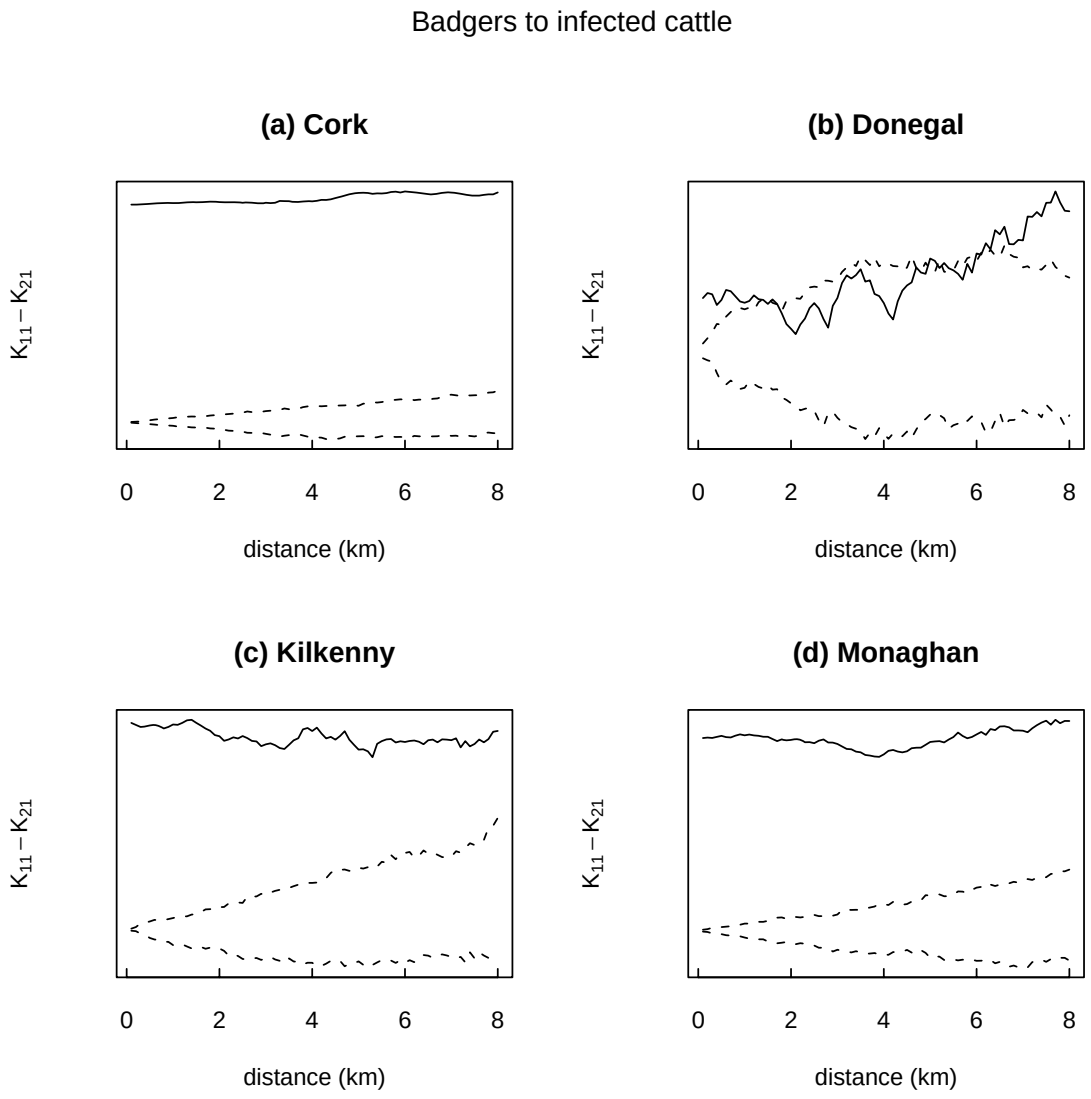


Fig.6. Differences (with confidence bands dashed) in K-functions of infected and uninfected animals. The difference is between K-functions of infected badgers to infected cattle and uninfected badgers to infected cattle. Animals were located in the removal areas of the FAP. Distances were calculated using minimum distance method. Infection status of herds was taken in the first year of badger culling. Badger locations were condensed. A single location could contribute data both as TB infected (if one or more

infected badgers were captured there) and as uninfected (if one or more uninfected badgers were captured there).

Supplementary Material

The following supplementary material is available below:

Appendix S1. Alternative analysis of the data using the distance ratio methods of Woodroffe *et al.* (2005) for the RBCT.

Appendix S2. Further details and results on K-functions, intensity functions, edge-corrections and kernel probability maps.

Fig. S1. Plots showing the locations of infected and non-infected badgers and cattle in Cork and Donegal removal areas.

Fig. S2. Plots showing the locations of infected and non-infected badgers and cattle in Kilkenny and Monaghan removal areas.

Fig.S3. Clustering of infection within badger populations in removal areas and within cattle populations in removal and reference areas.

Fig.S4. Spatial association of M.bovis strains in badgers in removal areas.

Fig.S5. Spatial association of M.bovis infection in cattle and badgers in removal areas.

Table S1. Summary statistics describing cattle populations in both the removal and reference areas.

Appendix S1. Spatial analysis of the data using the distance ratio methods of Woodroffe *et al.* (2005).

Introduction

This document presents the analysis of spatial clustering of *M. bovis* infections and strain types within badger populations, of infections within cattle populations and spatial associations of infections between badgers and cattle using the distance ratio method of Woodroffe *et al*. (2005), omitted from the main text for reasons of space. The data are based on the Four Area Project (FAP) a large-scale badger removal trial carried out in four counties in Ireland from 1997-2002, and described in detail in Griffin *et al.* (2005). In the removal area of each county badger removal was intensive and proactive but reactive (in response to severe tuberculosis outbreaks in cattle) in the reference areas. Analyses here, as in the main paper, are restricted to the first twelve months of the FAP, since the numbers of badgers captured in any year after the first year of the project were too small to permit substantive analysis and to avoid possible distorting effects of recent badger culling on the distribution of infection

Methods

Statistical analyses

Analyses were conducted to examine spatial associations of infection using the methods. of Woodroffe *et al*. (2005) on the RBCT data in the UK. These are based on nearest neighbour distances badger-badger, cattle-cattle and badger-cattle. Firstly distances badger-badger, cattle-cattle and badger-cattle were computed. The calculation of

distances between herds or between badgers and herds is not straight-forward, as a herd may be located on multiple parcels of land. Two distance methods were used.

Distance method 1: For each parcel the centroid was located using the GIS system. Then the distance between two herds A, B, is found, by taking the minimum distance between the centroids of all parcels in herds A to all parcels in herd B.

Distance method 2: For each herd all those parcels adjacent to each other were amalgamated to make one larger parcel. The parcel with the largest area was chosen first and the area of the second largest parcel was accumulated to the area of the first and then the area of the third largest parcel was accumulated to this and so on. This was continued until we got at least 80 % of the total area of the farm. Those parcels that did not contribute to this 80% were dropped from the analysis. The centroids for the remaining parcels were found using GIS. The distance between two herds A, B, was then calculated as the minimum distance between the parcels constituting the 80% in herd A and those in herd B. The reason for using method 2 was that if a herd had a small outlying parcel of land where perhaps cattle were infrequently grazed, then distance to another herd should not be based on distance to this small parcel.

Similarly the distance between a badger and a herd is found by taking the minimum distance between the centroids of all parcels (or parcels constituting the 80%) in the herd to the badger capture location.

The Wilcoxon rank sum test was then used to compare summary distance measures from infected and uninfected animals and the sign test and Wilcoxon signed rank test were used to compare distance measures from the same animal.

Results

Figures S1 & S2 plot the spatial location of infected and non-infected badgers and cattle for each area.

BADGERS

The main text of the paper presents analyses based on K-functions and intensity functions of the extent to which clustering of infection was found in badgers in the removal areas of the FAP. Here we considered for study, the total of 1,039 adult badgers culled, 304 in the outer removal areas and 735 in the inner removal areas for which the infection status, sex and age were known, as in the main text.

Clustering of infection within the badger population

As in Woodroffe *et al*. (2005), we calculated, for each badger, the distance to the nearest infected badger and the distance to the nearest uninfected badger and used the ratio between these two distances as our outcome measure. In order not to generate infinite ratios for badgers captured at the same location, 1m was added to all distances between pairs of badgers. The ratio between the distance to the nearest infected badger, and the distance to the nearest uninfected badger, was much lower for infected animals (1.52) than for uninfected animals (61.31), Wilcoxon rank sum test ($p < 0.0001$), indicating that infected badgers were closer to other infected badgers than would be expected if the distribution of infection within the population was random. This difference was observed in all counties. However in Donegal the difference was not statistically significant. The results are displayed in Figure S3. These results agree with those found in the analyses

used in the main text. In the main analysis however, clustering of infection was statistically significant in all counties.

In the RBCT the analysis of Woodroffe *et al*. (2005), it was found that *M. bovis* infections were locally clustered within the badger populations. In that study, clustering was seen in nine of their ten trial areas (overall $p<0.001$). Ratios varied between 1 and 1,000 comparable to those in this study. This was true despite considerable differences in badger density and infection rates between the FAP and RBCT (see main text).

Associations between strain types of *M. bovis* among badgers

The main text of the paper presents analyses based on kernel probability maps of the extent to which different strain types of *M. bovis* infection were spatially segregated in badgers in the removal areas of the FAP. Here, as described in the main text, distance calculations related to strain types were done by two methods. In method 1, each badger contributed one observation for each strain type. In method 2, a capture location contributed one observation for each strain type as multiple strain types were occasionally found at the same location. Here we used an analysis similar to Woodroffe *et al*. (2005), and calculated for each infected badger, the distance to the nearest badger with the same strain type and to the nearest badger with a different strain type, and compared these distances using both the Wilcoxon signed rank test and a paired sign test. In cases where there was only one badger with a certain strain type, the distance to the nearest badger with the same strain type was set to a very large number so that the difference entering into the sign test and Wilcoxon test was negative and in the Wilcoxon it received the largest absolute rank. Using method 1, a statistically significant difference

was found in Monaghan (p=.0243 sign test, p=.0408 Wilcoxon), but significance values < 0.05 were not reached in the other counties. An overall difference was found when all counties were combined (p=.0276 sign test) with badgers closer to other badgers infected with the same strain type as themselves. The results are displayed in Figure S4. Using method 2, a statistically significant difference was found in Cork (p=.0675 sign test, p=.0016 Wilcoxon). An overall difference was found when all counties were combined (p=.0251 sign test) with locations closer to other locations infected with the same strain type as themselves. In the main analysis significant spatial segregation of strain types was found in Cork and Kilkenny with borderline statistical significance in Donegal. No segregation of strain types was found in Monaghan. The results for Monaghan here are in overall agreement with the main analysis since that used a different measure of spatial association i.e. spatial segregation.

CATTLE

Clustering of infection within the cattle population

The main text of the paper presents analyses based on K-functions and intensity functions of the extent to which clustering of infection was found in cattle in the removal areas of the FAP. Here both removal and reference areas are examined and summary statistics for the cattle populations are given in Table S1. Then the analysis of Woodroffe *et al*. (2005) was carried out, as it was above for the badgers. In summary, for all counties combined, using distance method 1, the median ratio for infected herds was 2.07 (1.87 reference area, 2.48 removal area) and for non-infected herds was 4.70 (4.40 reference area, 4.95

removal area). Using the Wilcoxon test the difference in ratio was significant both in removal ($p < 0.0001$) and reference areas ($p < 0.0001$) and overall ($p < 0.0001$). Using distance method 2, the median ratio for infected herds was 2.12 (1.76 reference area, 3.17 removal area) and for uninfected herds was 3.97 (3.64 reference area, 4.31 removal area). The results are displayed in Figure S3. Using the Wilcoxon test the difference in ratio was significant both in removal ($p < 0.0005$) and reference areas ($p < 0.0001$) and overall ($p < 0.0001$). We note that these differences were seen but were not statistically significant for most of the individual removal and reference areas within counties. Median distances were smaller using method 1 than method 2, as was expected. However, median ratios of distances are not necessarily smaller under either method and were close in all areas considered here. Only in the Donegal reference area did the method of calculating distance alter the results in terms of comparing infected to uninfected (Figure S3). We note in Donegal the very small number of infected herds. Thus, distances from infected to infected herds involve only nine herds. A change in distances related to one of these herds altered the results substantially but not significantly so. Thus, conclusions remain the same with both distance methods. The conclusions are also in agreement with those of the analyses in the main text.

Woodroffe *et al*. (2005) considered 821 herds in ten trial areas of which 185 were TB-affected (22.5%) and found clustering of infection in nine of the ten trial areas. In calculating distances a single location for each herd was used - the centroid. Ratios in that study, ranged from approximately 1 to 6 and are comparable to those in Figure S3 here other than Donegal. Thus our results are consistent with Woodroffe *et al*. but although we did find clustering overall using such an analysis we did not find it in all

individual counties. We note, as discussed in the main text, that in addition to badger density, the herd density in both studies is quite different, 1700/1214=1.4 per km^2 here and 0.75 herds per km^2 in the RBCT (Woodroffe *et al.* 2005).

BADGERS-CATTLE

Associations between infections in badgers and cattle

In the analysis of distances from badgers to cattle, as described also in the main text, data from badgers trapped at the same location (which shared the same distances to the nearest cattle herd) were condensed. A single location could contribute data both as TB infected (if one or more infected badgers were trapped there) and as uninfected (if one or more uninfected badgers were trapped there). The badger data thus consisted of 830 uninfected badgers at 491 locations and 207 infected badgers at 167 locations.

In the 12 months of the first year of badger culling, distances to the nearest herd were similar for infected and uninfected badgers ($p=0.56$, Wilcoxon rank sum test), both when distance was calculated by method 1 or 2 as described previously. Using distance method 1, the median distance was 0.55 km for uninfected badgers and 0.56 km for infected badgers. For distance method 2 the corresponding medians were 0.61 km and 0.67 km.

The main text of the paper presents analyses based on K-functions of the extent to which *M. bovis* infections were spatially associated in badgers and cattle in the removal areas of the FAP. Here we examine this question using the Woodroffe *et al*. (2005) type analysis. Using distance method 1, for the same time period as above, there was an overall significant difference in the ratio between the distance to the nearest TB-affected cattle

herd and the distance to the nearest unaffected herd for infected and uninfected badgers (p=0.02, Wilcoxon rank sum test). This indicated that infected badgers were spatially associated with infected cattle herds. The pattern was seen in all counties (but statistical significance was not reached in any county) except Donegal (Figure S5 (a)). Results were similar using distance method 2. Figure S5 (b) displays the median of the distances from badgers to the nearest infected herd, separately for infected and uninfected badgers. This shows associations occurred at a scale of 1-2 km. The results are in agreement with the main text where however statistically significant associations were found in all counties. In addition, spatial associations were shown there up to 8 km in most counties. Woodroffe *et al*. (2005) also found significant spatial association between infected badgers and infected cattle in 8 of 10 trial areas using the ratio method. We note that median distances from adult badgers to TB positive herds varied from 0.7 to 1.7 km over the four counties and are similar to those of Woodroffe *et al.* (2005) as shown in Figure S5. While our results are significant overall, significance was often not attained in individual counties. There are a number of possible explanations. Woodroffe *et al*. (2005) considered the 12 months prior to badger culling. Our data did not permit us to do that; instead our data spanned the first year of badger culling. As two or three removal operations occurred yearly, badger culling may already have affected our data, affecting associations both within the badger and cattle populations, as well as between them (O'Corry-Crowe *et al.* 1996; Tuyttens *et al*. 2000; Woodroffe *et al.* 2006). Secondly, the badger density was much higher in the areas studied in the RBCT - increasing the possibility for possible associations (O'Corry-Crowe *et al.* 1993). The interaction of these two factors is complex as two recent studies concerning the effects of perturbation

on TB in badgers (Macdonald, Riordan & Mathews 2006; Woodroffe *et al*. 2007) suggest. Related to this perhaps, thirdly, badger numbers were insufficient to see statistical significance in individual counties.

Olea-Popelka *et al*. (2005) report a negative association between strain types in badgers and cattle. However, *M*. *bovis* strain typing data from Irish reactor cattle during the four area project was very limited. For example, in that first year of the FAP, only 43% of TB-lesioned herds were typed and at the very most 58% of lesioned cattle were typed. Since frequencies for a particular strain type are necessarily smaller than frequencies for the 'rest of the strain types', distances between same strain types will be much more affected by incomplete data, and will be larger than reality. These omissions could seriously bias the results, including those in Olea-Popelka *et al.* (2005) and therefore strain data for cattle are not considered here. In the RBCT, Woodroofe *et al*. (2005) report that all lesioned herds had at least one cattle strain typed. In that study, strain type data were available for 5,469 cattle with *M. bovis* infections confirmed by culture that were tested within 10 km of badger capture locations, in the 12 months before or after badger culling. It was found that infected badgers were closer to cattle infected with the same strain type as themselves ($p < 0.001$, sign test).

Summary

All our results are consistent with the finding of Woodroffe *et al*. (2005), demonstrating clustering of infections within badger and cattle populations and associations between infections in badgers and cattle. The associations found here were somewhat weaker, particularly for cattle populations. The K-/intensity function approach in the main text

shows better agreement. It is possible the K-/intensity function approach accounted in more detail for the different infection rates and spatial distribution in and between badgers and cattle than the summary ratio measure used here. This summary measure is subject to long tails as is evidenced by the different results between the Wilcoxon and sign tests.

References

Griffin, J.M., Williams, D.H., Kelly, G.E., Clegg, T.A., O'Boyle, I., Collins, J.D. & More, S.J. (2005) The impact of badger removal on the control of tuberculosis in cattle herds in Ireland. *Preventive Veterinary Medicine*, **67**, 237-266.

Macdonald, D. W., Riordan, P. & Mathews, F. (2006). Biological hurdles to the control of TB in cattle: A test of two hypotheses concerning wildlife to explain the failure of control. Biological *Conservation*, **131**, 268-286.

O'Corry-Crowe, G., Eves, J. & Hayden, R. (1993). Sett distribution, territory size and population density of badgers (Meles meles L.) in East Offaly. *The Badger* (ed T.J. Hayden), pp. 35-56. Royal Irish Academy, Dublin.

O'Corry-Crowe, G., Hammond, R., Eves, J. & Hayden, R. (1996). The effect of reduction in badger density on the spatial organization and activity of badgers Meles meles L. in relation to farms in central Ireland. *Biology and Environment: Proceedings of the Royal Irish Academy*, **96B,** 147-158.

Olea-Popelka, F.J., Flynn, O., Costello, E., McGrath, G., Collins, J.D., O'Keeffe, J.O., Kelton, D.F., Berke, O. & Martin, S.W. (2005) Spatial relationship between *Mycobacterium bovis* strains in cattle and badgers in four areas in Ireland. *Preventive Veterinary Medicine,* **71**, 57-70.

Tuyttens, F.A.M., Delahay, R.J., Macdonald, D.W., Cheeseman, C.l., Long, B. & Donnelly, C.A. (2000). Spatial perturbation caused by a badger *Meles meles* culling operation: implications for the function of territoriality and the control of bovine tuberculosis. *Mycobacterium bovis*. *Journal of Animal Ecology,* **69**, 815-828.

Woodroffe, R., Donnelly C.A., Johnston, W.T., Bourne, F.J., Cheeseman, C.L., Clifton-Hadley, R.S., Cox D.R., Gettinby, G., Hewinson, R.G., Le Fevre, A.M., McInerney, J.P. & Morrison, W.I. (2005) Spatial association of *Mycobacterium bovis* infection in cattle and badgers *Meles meles*. *Journal of Applied Ecology*, **42**, 852-862.

Woodroffe, R., Donnelly C.A., Cox D.R., Bourne, F.J., Cheeseman, C.L., Delahay, R.J., Gettinby, G., McInerney, J.P. & Morrison, W.I. (2006) Effects of culling on badger *Meles meles* spatial organization: implications for the control of bovine tuberculosis. *Journal of Applied Ecology*, **43**, 1-10.

Woodroffe, R., Gilks, P., Johnston, W.T., Le Fevre, A.M., Cox D.R., Donnelly C.A., Bourne, F.J., Cheeseman, C.L., Gettinby, G., McInerney, J.P. & Morrison, W.I. (2007). Effects of culling on badger abundance: implications for tuberculosis control. *Journal of Zoology*, in press.

Table S1. Summary statistics describing cattle populations in both the removal and reference areas. Data refer to the initial 12-month period of proactive culling of the FAP

Trial area	No. of TB tested herds	No. (%) of TB affected herds	Infection rate per km^2
Cork removal	399	57 (14)	0.19
Cork reference	262	48 (18)	0.24
Donegal removal	369	9 (2)	0.04
Donegal reference	303	9 (3)	0.03
Kilkenny removal	232	25 (11)	0.08
Kilkenny reference	219	37 (17)	0.15
Monaghan removal	700	44 (6)	0.12
Monaghan reference	518	63 (12)	0.17

Figure S1. Plots shows the locations of infected and non-infected badgers and cattle in Cork and Donegal removal areas.

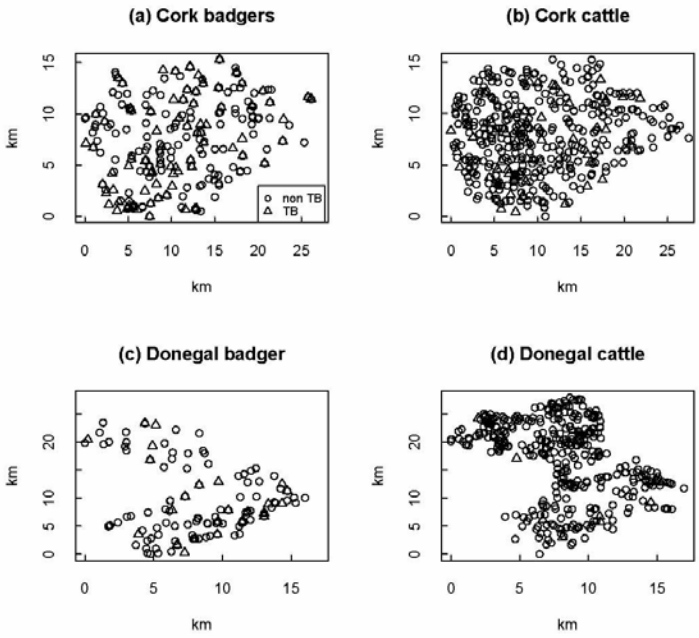
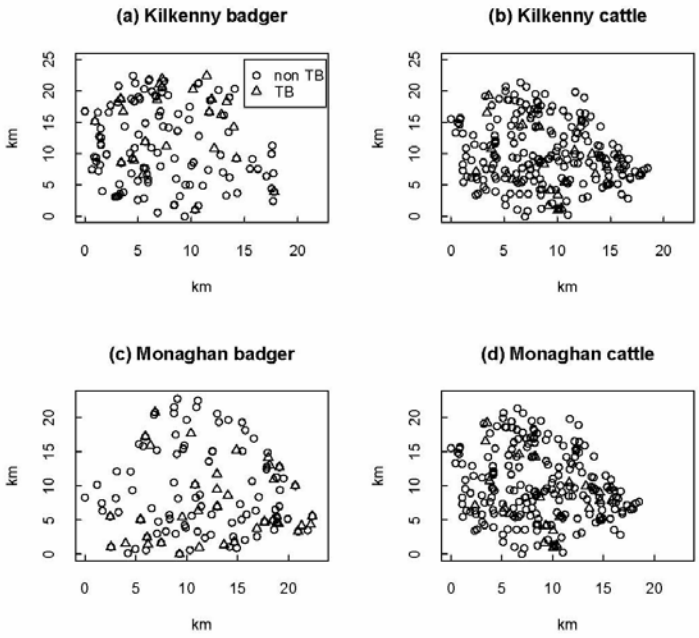
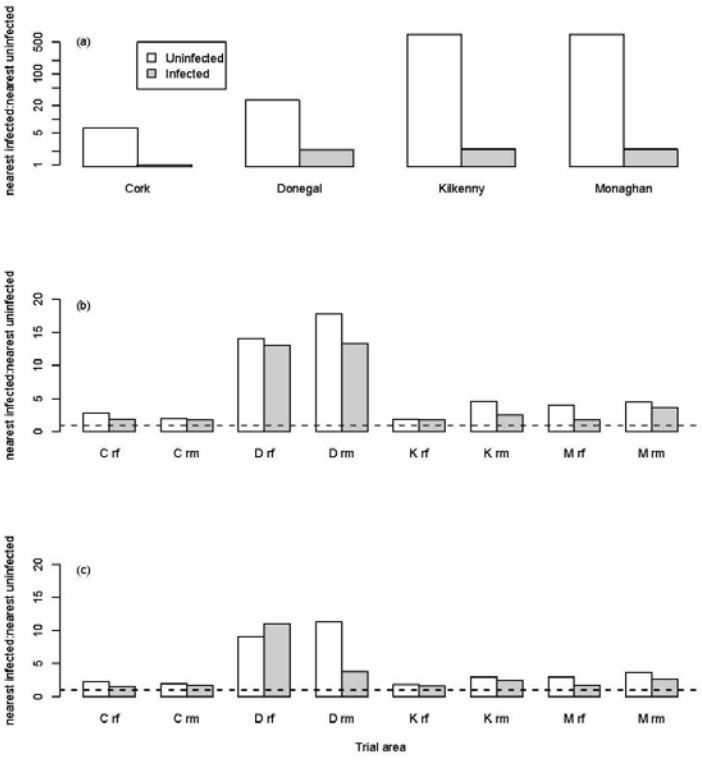


Figure S2. Plots shows the locations of infected and non-infected badgers and cattle in Kilkenny and Monaghan removal areas.



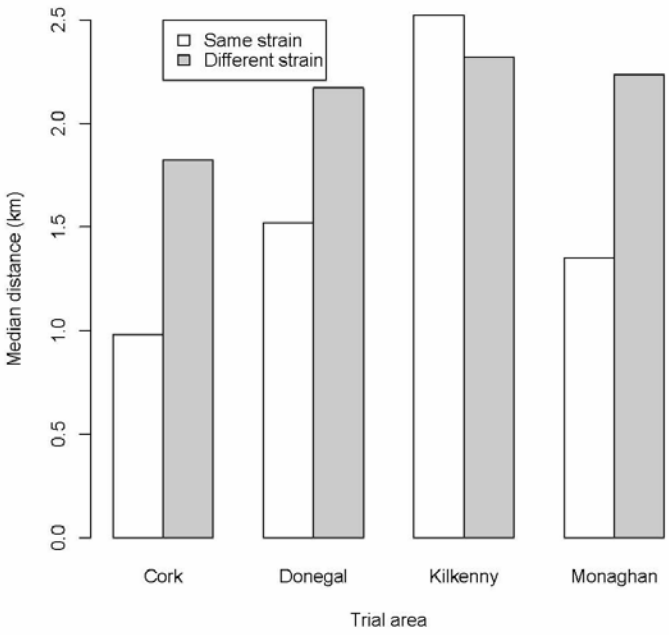
1

Figure S3. Clustering of infection within badger populations in removal areas and within cattle populations in removal and reference areas. Plots show the ratio between the distance from an animal to its nearest infected neighbour, and the distance to its nearest uninfected neighbour, expressed as medians across all uninfected (open bars) and infected (grey bars) animals, in each of the four counties: Cork(C), Donegal (D), Kilkenny (K), Monaghan (M).
(a) badgers: Ratios (log scale) in each of the four counties (removal area only)are smaller for infected badgers, indicating clustering of infection. (b) and (c) cattle herds : (b) Distances were calculated using minimum distance method. (c) Distances were calculated using the minimum distance based on 80 % of farmland only. Infection status was taken in year prior to last badger cull. In general, ratios in each of the four counties (removal(rm) and reference (rf))are smaller for infected herds, indicating clustering of infection.



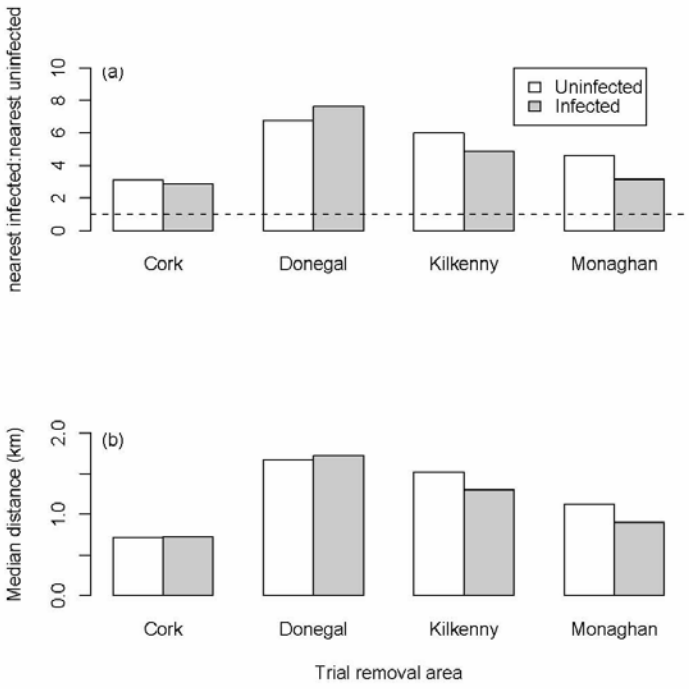
1

Figure S4. Spatial associations of *M. bovis* strains in badgers in removal areas. Plot shows the median distance from each infected badger to the nearest badger infected with the same strain (open bars) and the nearest infected with a different strain (grey bars).



1

Figure S5. Spatial association of *M. bovis* infection in cattle and badgers in removal areas. (a) Plot shows the ratio between the distance from a badger capture location to the nearest infected cattle herd, and the distance to the nearest uninfected cattle herd, expressed as medians across all uninfected (open bars) and infected (grey bars) badgers. Overall, ratios are smaller for infected badgers. (b) Plot shows the median distances from badger capture locations of uninfected (open bars) and infected (grey bars) to TB-positive herds. In (a) - (b) distances were calculated using the minimum distance method and infection status was taken in year prior to last badger cull. Badgers and herds in the proactive treatment area (i.e. removal and buffer areas) were used.



1
2

3 **Appendix S2**. Further details and results on K-functions, intensity functions, edge-

4 corrections and kernel probability maps.

Introduction

This document presents additional details and results of the K-functions and intensity functions together with the edge-corrections used in the analysis of spatial clustering of *M. bovis* infections within and between badger and cattle populations omitted from the main text for reasons of space. The data are based on the Four Area Project (FAP) a large-scale badger removal trial carried out in four counties in Ireland from 1997-2002, and described in detail in Griffin *et al.* (2005). In the removal area of each county badger removal was intensive and proactive but reactive (in response to severe tuberculosis outbreaks in cattle) in the reference areas.

K-Functions

Spatial patterns of infection were visualized using K-functions, based on the theory of spatial point processes (Ripley, 1981). Let A denote the region under study. The K-function is defined by

$K(d) = \lambda^{-1} E[\text{number of further points within a distance d of an arbitrary point}]$. (1)

where E() denotes expectation and λ is the intensity or the mean number of points per unit area. Here the points are locations of herds or badger setts. An estimator for $\lambda K(d)$ is the average, over all points of the pattern, of the number seen within a distance d of that point. Ripley suggested an edge-correction for points outside A. Let p(x,y) be the

proportion of the circumference of a circle with centre x passing through y that lies within A. Thus we use

$$\hat{K}(d) = \frac{|A|}{n^2} \sum_{x,y \varepsilon A, d(x,y) \leq d} \frac{1}{p(x,y)}$$

to estimate K(d). Essentially, the K-function describes the extent to which there is spatial dependence in the arrangements of the points. As noted by Gatrell *et al*. (1996) and Diggle (2003) it makes little sense to examine individual K-functions as we may expect to observe a certain amount of clustering due to environmental heterogeneity. We consider points of two different types, infected points labelled j = 1, and non-infected labelled j = 2. For a labelled stationary, isotropic point process, $x_i \varepsilon R^2 : i = 1,2,\ldots,$ in which points are of qualitatively different types j=1, 2 (here infected and uninfected), Diggle & Chetwynd (1991) similarly define a set of K-functions

$K_{ij}(d) = \lambda_j^{-1}$ E[number of (further) type j points within distance d of an arbitrary type i point], (2)

where λ_j is the intensity of type j points. When i=j, eqn (2) agrees with eqn (1). Also $K_{12}(d) = K_{21}(d)$. By spatial clustering we mean a general tendency for cases, i.e. infected animals, to occur more closely together than would be compatible with random sampling from the population at risk. This is a description of the underlying disease process rather than the study region itself. The implication of clustering is that the conditional intensity of cases at an arbitrary location y, given a case at a nearby location x, is greater than the unconditional intensity of cases at y, i.e. clustering involves a form of dependence between cases.

Under the null hypothesis of no clustering, cases form a spatially random sample from the underlying population. By design, controls necessarily form a spatially random sample from this same population. Hence, no spatial clustering is equivalent to the random labelling hypothesis H, where the type 1 points constitute a random thinning of the unlabelled point process defined as the superposition of type 1 and type 2 points. Under H,

$K_{11}(d) = K_{22}(d) = K_{12}(d)$ for all d. (3)

Note that eqn (3) does not require any parametric assumptions about the underlying unlabelled process. We consider departures from H by assessing the significance of the difference $D(d) = K_{11}(d)-K_{22}(d)$, estimated by $\widehat{D}(d) = \widehat{K}_{11}(d) - \widehat{K}_{22}(d)$

For data $x_i \varepsilon A : i = 1,\ldots,n$ where $n=n_1 + n_2$ with the first n_1 points of type 1 and the remainder of type 2 unbiased estimators for the $K_{ij}(d)$ can be found as follows, as in Diggle & Chetwynd (1991). Let w(x,d) be the reciprocal of the proportion of the circumference of the circle with centre x and radius d which lies within A. Let d_{ij} be the distance between x_i and x_j. Let $\delta_{ij}(d)$ be the indicator of the event $d_{ij} \le d$. Write $w_{ij} = w(x_i, d_{ij})$ for $j \neq i$, and $w_{ii} = 0$. Then

$$\hat{K}_{11}(d) = |A|\{n_1(n_1-1)\}^{-1}\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} w_{ij}\delta_{ij}(d), \quad (4)$$

$$\hat{K}_{22}(d) = |A|\{n_2(n_2-1)\}^{-1}\sum_{i=n_1+1}^{n}\sum_{j=n_1+1}^{n} w_{ij}\delta_{ij}(d), \quad (5)$$

and

$$\hat{K}_{12}(d) = |A|\{n(n_1-1)(n_2-1\}^{-1}\sum_{i=1}^{n_1}\sum_{j=n_1+1}^{n}(n_2 w_{ij} + n_1 w_{ji})\delta_{ij}(d) \quad (6).$$

Significantly positive values of $\hat{D}(d) = \hat{K}_{11}(d) - \hat{K}_{22}(d)$ would constitute evidence of spatial clustering of the disease in question.

We evaluate the null sampling distribution of $\hat{D}$(d) by carrying out 99 Monte Carlo simulations in each of which disease labels were randomly assigned to locations. Upper and lower confidence bands were thus obtained. Also D(d) can be interpreted as an expectation: λ_1D(d) represents the expected number of excess cases within a distance d of a typical case, by comparison with the number expected in the absence of clustering (where λ_1 is the intensity of type 1 points).

Intensity functions

If K(d) is known for a particular process, the second-order intensity is given by

$$I(h) = \frac{\lambda^2}{2\pi h}\frac{dK(h)}{dh}$$

By taking simple differences $K_{ij}(d+s)-K_{ij}(s)$ and dividing by s to estimate the derivative of $K_{ij}(d)$, estimators for the intensities corresponding to K_{11}, K_{22} and K_{12} were found.. Differences between these intensities were also examined and confidence intervals for these differences generated by re-estimating I_{ij} using Monte Carlo simulation, as above. These have the advantage of showing the exact distances d at which clustering occurs. Thus plots of $I\hat{D}(d) = \hat{I}_{11}(d) - \hat{I}_{22}(d)$ versus d were also generated. Here $I_{ij}(d)=K_{ij}(d+1)-K_{ij}(d)$ was used , i.e. a bandwidth of h=1 km. We note however that confidence bands for D(d) are narrower than ID(d). Further details regarding intensity functions can be found in Schabenberger & Gotway (2005).

Edge corrections for K-functions and intensity functions.

Note the weight w_{ij} in eqn (4)-(6) for K-functions involving cattle, is based on the centroid of that land parcel of farm i corresponding to the minimum distance to farm j (or sett j). The weights w(x,d) in equations (1) – (6) above, were found by invoking the geographic information system software ArcView version 9.2 (ESRI Inc., Redlands, CA) and writing a macro to generate circles and clip them to the shape of the study areas. The proportions of the areas of the circles within the study area boundaries were then calculated. Ripley (1988, Chapter 3) argues that it is more correct to calculate proportions of the circumferences but Besag (1977) argues this correction gave an excessive weight to the furthest neighbours and proposed using area as done here. The robustness of the results to the choice of weight was checked by performing some analyses using the weights $w(x,d)^{0.5}$, a choice based on calculations with rectangles and other mathematically regular regions. In all counties results were remarkably similar for the badger K-functions using $w(x,d)^{0.5}$.

Spatial variation in risk – kernel probability maps.

Spatial variation in risk between strains was examined as described in Diggle (2007). By spatial variation in risk we mean (first–order) intensity functions for each strain are not proportional. With s strain types, the pattern of strains is assumed to be generated by a

multivariate Poisson process with intensities $\lambda_k(x)$, $k=1, \ldots, s$ at each location x. Then the probability a case at location x will be of type j, conditional on there being one of the s types at x, is $p_j(x) = \lambda_j(x)/\Sigma \lambda_k(x)$. We say there is spatial segregation if the area can be partitioned approximately into sub-regions where one strain type predominates i.e complete segregation is if at each x in the sub-region, $p_j(x) = 1$ for one of the j. We used the kernel estimator of p_j given in Diggle (2007) where the smoothing parameter for p_j is chosen by cross-validation.

References

Besag, J. E. (1977) Comments on Ripley's paper. *Journal of the Royal Statistical Society*, **B 39**, 193-195.

Diggle, P.J. & Chetwynd, A.G. (1991) Second order analysis of spatial clustering in inhomogenous populations. *Biometrics*, **47**, 1155-1163.

Diggle, P.J. (2003) *Statistical analysis of spatial point patterns*, 2nd edn. Arnold, London.

Diggle, P.J. (2007) Spatio-temporal point processes: methods and applications. *Statistical methods for spatio-temporal systems* (eds B. Finkenstadt, L.Held &V. Isham), pp. 1-47. Chapman & Hall/CRC, London.

Gatrell, A.C., Bailey , T.C., Diggle, P.J. & Rowlingson, B.S. (1996) Spatial point pattern analysis and its application in geographical epidemiology. *Transactions of the Institute of British Geographers*, **21**, 256-274.

Griffin, J.M., Williams, D.H., Kelly, G.E., Clegg, T.A., O'Boyle, I., Collins, J.D. & More, S.J. (2005) The impact of badger removal on the control of tuberculosis in cattle herds in Ireland. *Preventive Veterinary Medicine*, **67**, 237-266.

Ripley, B. (1981) *Spatial statistics*. Wiley, Chichester.

Ripley, B. (1988) *Statistical inference for spatial processes*. Cambridge University Press, Cambridge.

Schabenberger, O. & Gotway, C.A. (2005) *Statistical methods for spatial data analysis*, Chapter 3. Chapman & Hall/*CRC*, London.