General enquiries on this form should be made to:
Defra, Science Directorate, Management Support and Finance Team,
Telephone No. 020 7238 1612
E-mail: research.competitions@defra.gsi.gov.uk

SID 5 Research Project Final Report

- **Note**

  In line with the Freedom of Information Act 2000, Defra aims to place the results of its completed research projects in the public domain wherever possible. The SID 5 (Research Project Final Report) is designed to capture the information on the results and outputs of Defra-funded research in a format that is easily publishable through the Defra website. A SID 5 must be completed for all projects.

  A SID 5A form must be completed where a project is paid on a monthly basis or against quarterly invoices. No SID 5A is required where payments are made at milestone points. When a SID 5A is required, no SID 5 form will be accepted without the accompanying SID 5A.

- This form is in Word format and the boxes may be expanded or reduced, as appropriate.

- **ACCESS TO INFORMATION**

  The information collected on this form will be stored electronically and may be sent to any part of Defra, or to individual researchers or organisations outside Defra for the purposes of reviewing the project. Defra may also disclose the information to any outside organisation acting as an agent authorised by Defra to process final research reports on its behalf. Defra intends to publish this form on its website, unless there are strong reasons not to, which fully comply with exemptions under the Environmental Information Regulations or the Freedom of Information Act 2000.

  Defra may be required to release information, including personal data and commercial information, on request under the Environmental Information Regulations or the Freedom of Information Act 2000. However, Defra will not permit any unwarranted breach of confidentiality or act in contravention of its obligations under the Data Protection Act 1998. Defra or its appointed agents may use the name, address or other details on your form to contact you in connection with occasional customer research aimed at improving the processes through which Defra works with its contractors.

Project identification

1. Defra Project code: SE3265

2. Project title:

   Field approaches to identifying Mycobacterium bovis infection in badger populations.

3. Contractor organisation(s):

   AHVLA
   Block C
   Government Buildings
   Whittington Road
   Worcester
   WR5 2LQ

4. Total Defra project costs: £ 816,101

5. Project: start date ..... 01 April 2011

   end date ..... 30 April 2014

6. It is Defra's intention to publish this form.
Please confirm your agreement to do so.....**YES** ☒ **NO** ☐

(a) When preparing SID 5s contractors should bear in mind that Defra intends that they be made public. They should be written in a clear and concise manner and represent a full account of the research project which someone not closely associated with the project can follow.
Defra recognises that in a small minority of cases there may be information, such as intellectual property or commercially confidential data, used in or generated by the research project, which should not be disclosed. In these cases, such information should be detailed in a separate annex (not to be published) so that the SID 5 can be placed in the public domain. Where it is impossible to complete the Final Report without including references to any sensitive or confidential data, the information should be included and section (b) completed. NB: only in exceptional circumstances will Defra expect contractors to give a "No" answer.
In all cases, reasons for withholding information must be fully in line with exemptions under the Environmental Information Regulations or the Freedom of Information Act 2000.

(b) If you have answered NO, please explain why the Final report should not be released into public domain

Executive Summary

7. The executive summary must not exceed 2 sides in total of A4 and should be understandable to the intelligent non-scientist. It should cover the main objectives, methods and findings of the research, together with any other significant events and options for new work.

We used existing and contemporary data to investigate inter-relationships amongst diagnostic test results for detecting *M. bovis* infection in badgers and to explore the potential benefits of their application both in isolation and in combination for the diagnosis of infection in badgers at an individual and a group level.

The principal data source we used was the long-term dataset from the Woodchester Park badger study, supplemented by data from the Badger Vaccine Study (BVS) and the Randomised Badger Culling Trial (RBCT).

Objective 1: Investigation of sampling strategies for a sett-based badger management policy

(a) Investigation of relationships between different diagnostic tests
In individual live badgers there were positive correlations between StatPak and culture of clinical samples, and between the number of positive culture samples obtained from an individual badger and IFNg OD readings, suggesting that both StatPak and IFNg may be potential predictors of the likelihood of shedding *M. bovis* bacilli. PCR test results from faecal samples were positively correlated with IFNg OD values, but not with culture or StatPak test results.

Despite some evidence for positive correlation, there was no clear relationship between INFg test results and *post mortem* lesion severity scores in individual animals, consistent with immunology being a poor predictor of anticipated pathology.

(b). The value of diagnostic tests applied at the badger social group level
We modelled existing empirical diagnostic test data to explore the predictive value of currently available diagnostic tests, when used both singly and in combination, to classify the infection status of badgers at a sett or social group level. Results indicated that the sensitivity of StatPak, IFNg and culture of clinical samples was low when each test was used in isolation.

Parallel interpretation of the results of tests used in combination improved sensitivity but reduced specificity. Serial test interpretation improved the specificity of tests, but markedly reduced sensitivity.

The combined use of StatPak and INFg with a trigger of two animals testing positive to confirm positivity appeared to provide practical opportunities to identify infected setts or social groups at levels of trapping efficiency that have been observed in previous field studies. The proportion of the badger social group that needs to be sampled in order to correctly detect infection increases as prevalence decreases.

Any new diagnostic test applied to trapped badgers would need to be more than 80% sensitive in order to achieve better results than the combination of StatPak and IFNg. However, in its current format the IFNg cannot be performed as a trap-side test.

(c). Modelling intervention strategies
We used a current badger/cattle/bTB model to investigate the potential practical outcomes (numbers of infected badgers remaining) of interventions including live testing of captured badgers compared to no control, vaccination and unselective culling approaches.

As in previous simulations social perturbation countered the effect of culling to some extent in the core, and particularly in a surrounding 2km ring.

The use of the StatPak test at the group level with a threshold of two positive results required to trigger a group-wide cull was always had the worst outcomes and the untargeted cull had the best outcomes. The combined StatPak and gamma test at the group level with a trigger of one positive result leading to a group level cull had very similar outcomes to the untargeted cull in terms of the reduction in the number of infected badgers but resulted in less badgers being removed.

Most of the test and cull strategies investigated were improved to some extent when test-negative groups were vaccinated, although none led to an overall improvement on the untargeted culling strategy.

Objective 2: Disease progression and test outcomes

We used existing diagnostic test result data from live badgers to investigate whether IFNg test could be used to predict future positive results on other tests, and hence predict future disease progression.

Badgers with the highest IFNg OD values were most likely to subsequently test positive on both StatPak and culture, and this effect was detectable for up to 24 months after the IFNg test. Furthermore, the higher the original IFNg OD value, the greater the chance that a badger would subsequently test positive on StatPak. The relationship between IFNg and culture was less clear but may have been influenced by the low numbers of culture-positive badgers.

These results show the value in using continuous OD value data as opposed to binary positive/negative threshold data for diagnostic purposes.

Objective 3: Novel approaches to interpreting diagnostic test results

In these analyses we describe an approach for interpreting potentially disparate results from animals sampled multiple times over a protracted period and address the question of determining the probability that a badger is truly infected given its current and historical pattern of diagnostic test results.

We used a state-dependent capture-mark-recapture analysis of data from the Woodchester Park study which accounted for complexities in ecological epidemiology whilst estimating the probability that a badger was truly infected.

Our results indicate that this is a powerful approach for accounting for some of the underlying biological complexities which should not be ignored when attempting to evaluate epidemiological dynamics in wild animal populations.

Project Report to Defra

8. As a guide this report should be no longer than 20 sides of A4. This report is to provide Defra with details of the outputs of the research project for internal purposes; to meet the terms of the contract; and to allow Defra to publish details of the outputs to meet Environmental Information Regulation or Freedom of Information obligations. This short report to Defra does not preclude contractors from also seeking to publish a full, formal scientific report/paper in an appropriate scientific or other journal/publication. Indeed, Defra actively encourages such publications as part of the contract terms. The report to Defra should include:
   - the scientific objectives as set out in the contract;
   - the extent to which the objectives set out in the contract have been met;
   - details of methods used and the results obtained, including statistical analysis (if appropriate);
   - a discussion of the results and their reliability;

- the main implications of the findings;
- possible future work; and
- any action resulting from the research (e.g. IP, Knowledge Transfer).

Field approaches to identifying *Mycobacterium bovis* infection in badger populations

Background

Bovine tuberculosis (TB: infection with *Mycobacterium bovis*) is a zoonotic disease with a worldwide distribution, and which has a serious impact on livestock profitability, cattle health and welfare, and in some situations may present a risk to human health. In the UK, despite a variety of control measures (principally based on the test and slaughter of cattle), eradication has not been achieved. One impediment to this is the presence of infection in wildlife, most notably the European badger (*Meles meles*) which is the principal wild maintenance host of bovine TB in the UK.

Disease control measures in wildlife populations are challenging to apply owing to ecological complexities and practical difficulties, including for example, the absence of effective diagnostic tools for wild hosts. However, accurate recognition of the infectious state of a host is likely to significantly improve the effectiveness of disease control interventions. In the case of *M. bovis* infection in badgers, no gold standard diagnostic test is currently available. However, it is possible to combine available data on several existing but imperfect diagnostic tests and thereby increase diagnostic certainty (Drewe *et al*., 2010). The purpose of the work described below was to investigate inter-relationships amongst the existing diagnostic approaches and to explore the potential benefits of their application both in isolation and in combination for the diagnosis of *M. bovis* infection in badgers at an individual and group level.

Diagnostic approaches for use in live badgers that are considered in this report include the StatPak, the interferon gamma test (IFNg) and the culture of clinical samples. In addition we used data derived from *post mortem* examinations of carcases, and from a study employing a polymerase chain reaction (PCR) assay as a method of detecting *M. bovis* in badger faeces.

The StatPak test has been used in the long-term Woodchester Park study (WP) since 2006 and was also employed during the Badger Vaccination Study (BVS). The test identifies antibodies produced in response to specific antigens associated with *M. bovis* (Chambers *et al*., 2008). The results of the test are binary (positive or negative).

The IFNg test measures the secretion of interferon gamma cytokine by T-cells following stimulation with purified protein derivatives of bovine (PPD-B) and avian (PPD-A) tuberculin (Dalley *et al.*, 2008). INFg test results were available on a continuous scale as optical density (OD) readings of gamma interferon production. For each badger, an INFg OD value was calculated as the amount of INFg produced following stimulation with PPD-B minus the amount of INFg produced by stimulation with PPD-A. Binary values for the INFg test were produced by using an OD cut-off value of 0.044, as reported previously (Dalley *et al*., 2008). IFNg data came from the badger population at Woodchester Park.

Mycobacterial culture of clinical samples directly tests for the presence of *M. bovis* in faeces, urine, tracheal aspirate, oesophageal aspirate, swabs from bite wounds, burst sub-mandibular lymph nodes and non-bite-related wounds (Clifton-Hadley *et al*., 1993). Results were recorded as positive or negative for each of the culture samples. Culture data came from the badger population at Woodchester Park (see Delahay *et al*., 2000a & 2013).

Lesion severity scores derived from the *post mortem* examination of badgers, was used to determine if the StatPak, IFNg and culture tests described above could predict internal pathology and therefore provide an indication of likely infectiousness. Calculation of a pathology severity score on a scale from 0.0 to 4.0 has been used as a semi-quantitative measure of the severity of gross disease present *post-mortem* in published TB vaccine efficacy studies in possums and badgers (Corner *et al*., 2001 & 2007; Chambers *et al*., 2010). *Post mortem* data came from badgers culled during the RBCT.

A PCR test was conducted on faecal samples derived from the badger population at Woodchester Park as part of a separate defra-funded project led by Warwick University (SE3280). Faeces were collected from anaesthetised badgers with five technical replicates conducted on each badger. Individual quantitative (real-time) PCRs were considered positive if the cycle threshold (CT) was less than 40 (CT refers to the number of PCR cycles required for the fluorescent signal to exceed the background level). CT levels are inversely proportional to the amount of target nucleic acid in the sample (i.e. the lower the CT level the greater the amount of target nucleic acid in the sample). At least two replicates out of the five from each badger were required to test positive for that animal to be considered as TB positive using the PCR test.

Most of the data used in the following analyses were derived from historical and contemporary (i.e. within the lifetime of this project) diagnostic test results from the long-term Woodchester Park study. The methods employed in this study are well described (e.g. Delahay *et al*., 2000a & 2013) but supporting fieldwork is briefly described below.

Supporting Fieldwork

Routine fieldwork at Woodchester Park involves survey work to identify active badger setts and territories, the regular trapping of badgers and collection of samples for diagnostic testing, and provision of supporting contemporary ecological and epidemiological information from the badger population. In order to provide a sufficient number of social groups and variation in their demographic structure, an adequate number of individuals of all age and sex groups in a range of disease states, and to ensure representation of a range of ecological conditions, we monitored a cluster of 21 badger social groups in the Woodchester Park study area. Both bait-marking and capture-mark-recapture studies took place at this scale for the duration of the project. Below we briefly list field activities that were required to support the current project.

(1). Bait-marking
The spatial structure of badger social group territories was determined by bait-marking. This is both a practical requirement for trapping operations (by identifying active setts and their social group affiliations) and provided information necessary to investigate diagnostic approaches at the social group/sett level.

Bait marking was carried out each spring according to a well-established protocol (Delahay *et al*., 2000b). A palatable bait of peanuts and syrup laced with small coloured plastic pellets was fed for 10 days at the putative main sett of each social group. The subsequent distribution of the harmless indigestible pellets in faeces at badger latrines was used to map social group territories using a geographic information system (ArcView 3.2 or ArcGIS 8/9, ESRI, Redlands, California, USA).

(2). Live capture
Badgers were trapped using steel mesh box traps deployed at active setts, baited with peanuts and set after 4-8 days of pre-baiting. Traps were located on or near to badger 'runs' at the active setts in each territory. Trapped badgers were anaesthetised with a mixture of ketamine hydrochloride, medetomidine hydrochloride and butorphanol tartrate (de Leeuw et al., 2004), and on first capture each was given a unique identifying tattoo (Cheeseman & Harris, 1982). The location, sex, body weight and condition, reproductive status and age class was recorded. Trapping was suspended between 1st February and 30th April inclusive when most cubs are very young, confined to the sett, and totally dependent on their mother (see Woodroffe et al., 2005). During January (and weather dependent during December and May), when some females may be lactating, traps were checked during the night, and lactating females, or females deemed to be pregnant, were released immediately without sampling. After the collection of samples and a period of recovery, all trapped badgers were released at the point of capture. All badger carcasses found in the study area were submitted for *post mortem* examination for tuberculous lesions and tissue samples from lymph nodes and major organs taken for bacterial culture. All data arising from the capture and examination of badgers was added to a relational database (ACCESS 2000, Microsoft, USA).

It should be noted that the trapping of badgers and collection of diagnostic test data at Woodchester Park not only served the specific objectives of the present proposal, but also provided samples and

data for other defra-funded projects on diagnostic test development (e.g. SE3280: Development of an environmental PCR for *M. bovis* detection; SE3273: Systems for sample collection (blood and urine) from unanaesthetised badgers for diagnostic purposes; SE3281: Development of novel diagnostic tests for *M. bovis* detection in badgers).

Objective 1: Investigation of sampling strategies for a sett-based badger management policy

(a) Investigation of relationships between different diagnostic tests

Introduction

The focus of this task was the analysis of multiple existing datasets from the Woodchester Park study (WP), the Badger Vaccine Study (BVS) and the Randomised Badger Culling Trial (RBCT) to investigate relationships between the results of different diagnostic tests performed on the same animals. Individual badger test results using Stat-Pak, gamma interferon (IFNg), culture of clinical samples, and (where available) *post-mortem* examination, were analysed.

Three questions were addressed:

1. Is there a correlation between *post mortem* lesion severity scores and IFNg test results which might make it possible to use IFNg as a predictor of pathology and therefore infectiousness? Such a relationship might allow the IFNg test to be used as a tool for the selective removal of individuals before they become infectious.
2. Can StatPak (binary result) or IFNg (continuous result) test results be used to predict multi-site culture-positive (i.e. exhibiting advanced disease) badgers?
3. How do results from the Warwick PCR study on detecting *M. bovis* in badger faeces (Defra project SE3280) compare with those from other diagnostic tests (StatPak, IFNg, faecal culture, other culture) at the individual animal level?

Methods

Diagnostic approaches

Correlations were explored between five diagnostic approaches: Stat-Pak, IFNg and mycobacterial culture of clinical sample results from the WP study, *post mortem* lesion severity scores and IFNg results from the RBCT, and PCR results from a separate study at WP. Details of these approaches are described above (background section).

Data analysis

Analysis was conducted using simple linear regression and mixed effects models (with individual animals as a random effect to avoid pseudo-replication). Non-parametric tests were used where appropriate to account for the skewed nature of some of the data.

Continuous INFg OD values were used to produce categories of INFg results (Table 1). As the distribution of variables was highly right-skewed it was not appropriate to simply divide the range of values by the number of categories in order to obtain cut-off values. Ideally, every category of the variable should have a similar number of observations to ensure equal weight in the analysis. For this part of the analysis, negative values (arising from cases where the OD of PPD-A was higher than that for PPD-B) were coded as category 0, while values higher than zero but less than 0.044 were coded as category 1. Categories 2 and 3 were equally spaced (starting from 0.044) using an interval step of 0.33. The category coded as 4 included values higher than 0.70 with the highest value of INFg OD in this category being 2.29 (Table 1).

For *post mortem* data, the degree of TB lesion severity had been established previously by an expert pathologist and recorded as a score on a continuous scale from 0 to 4 (see Corner *et al*., 2001 & 2007; Chambers *et al*., 2010). Data from 241 badgers with pathology scores and IFNg test values were available for analysis in the present study, of which the highest lesion severity score was 1.37. Lesion severity scores were grouped to produce categorical variables, by dividing the range of values by the number of categories (with the exception of the fifth category which included a broader range of score values due to the right skewed distribution). The number of observations falling into each lesion score category is presented in Table 2.

PCR test results from badgers captured at WP were matched by date to the results of other diagnostic tests from the same individuals. A total of 70 badgers were tested using PCR alongside the other three diagnostic tests (StatPak, IFNg and culture of clinical samples) on 87 occasions (14 badgers were tested twice and two badgers were tested three times).

Table 1. Sample sizes and OD values for 2,205 gamma interferon (IFNg) test results from 546 live badgers trapped at Woodchester Park from July 2006 to October 2013. These categories were used for the analysis of relationships between IFNg and other diagnostic test results.

Category	INF gamma OD values (PPD-B minus PPD-A)	Number of observations	Percentage of observations
0	<0.000	639	29.0
1	0.000 - 0.043	1,156	52.4
2	0.044 - 0.366	314	14.2
3	0.367 - 0.696	52	2.4
4	0.697 - 2.291	44	2.0
Total	**0.000 – 2.291**	**2,205**	**100.0**

Table 2. Sample sizes and *post mortem* lesion severity scores for 240 badgers culled in the Randomised Badger Culling Trial. These categories were used for the analysis of relationships between lesion scores and IFNg test results from the same animals just before death.

Category	*Post mortem* lesion severity scores	Number of observations	Percentage of observations
0	0	162	67.5
1	0.01 - 0.11	27	11.3
2	0.12 - 0.21	27	11.3
3	0.22 - 1.37	24	10.0
Total	**0.00 – 1.37**	**240**	**100.0**

Results

Correlations between IFNg and culture

On average, three different types of clinical sample (e.g. urine, faeces and an oesophageal aspirate) were obtained for the purposes of culture from a badger on any one particular trapping event at Woodchester Park. Therefore, it was possible to get more than one positive culture test result per badger per trapping event. A positive linear relationship (on a logarithmic scale) was found between the number of culture positive samples and the contemporary IFNg result, such that the greater the number of clinical samples that tested culture positive at a single time point, the higher the INFg log OD value for the same badger at that point in time (z-test = 6.02, $p<0.001$). The model output was log INFg OD = [-3.30 + (0.84 multiplied by the number of culture-positive samples)]. This equation predicts that for each additional positive culture result, an increase in the log OD value of 0.84 would be expected. The mixed effects linear regression model produced a very good fit to the WP data for both INFg and culture tests (likelihood ratio test $p<0.001$). Figure 1 illustrates the level of correlation between culture and IFNg test results. This result should be interpreted with caution however, owing to the limited number of results available for badgers testing positive on more than one culture sample (see Table 3). Nonetheless, the model proved to be robust to removal or regrouping of the two badgers in the highest category of culture test results (i.e. those that tested culture positive on three or

more samples at one point in time) indicating that these small number of observations did not carry sufficiently significant statistical weight to modify the slope of the regression line.

Figure 1. Relationship between the number of positive culture samples obtained from a badger on a single sampling occasion and the concurrent INFg OD value. Results from a mixed effect linear regression model which predicted an increase in INFg log OD values of 0.84 for each additional culture sample that tests positive. Data was derived from 2,205 trapping events at Woodchester Park (July 2006 to October 2013) involving 546 badgers.

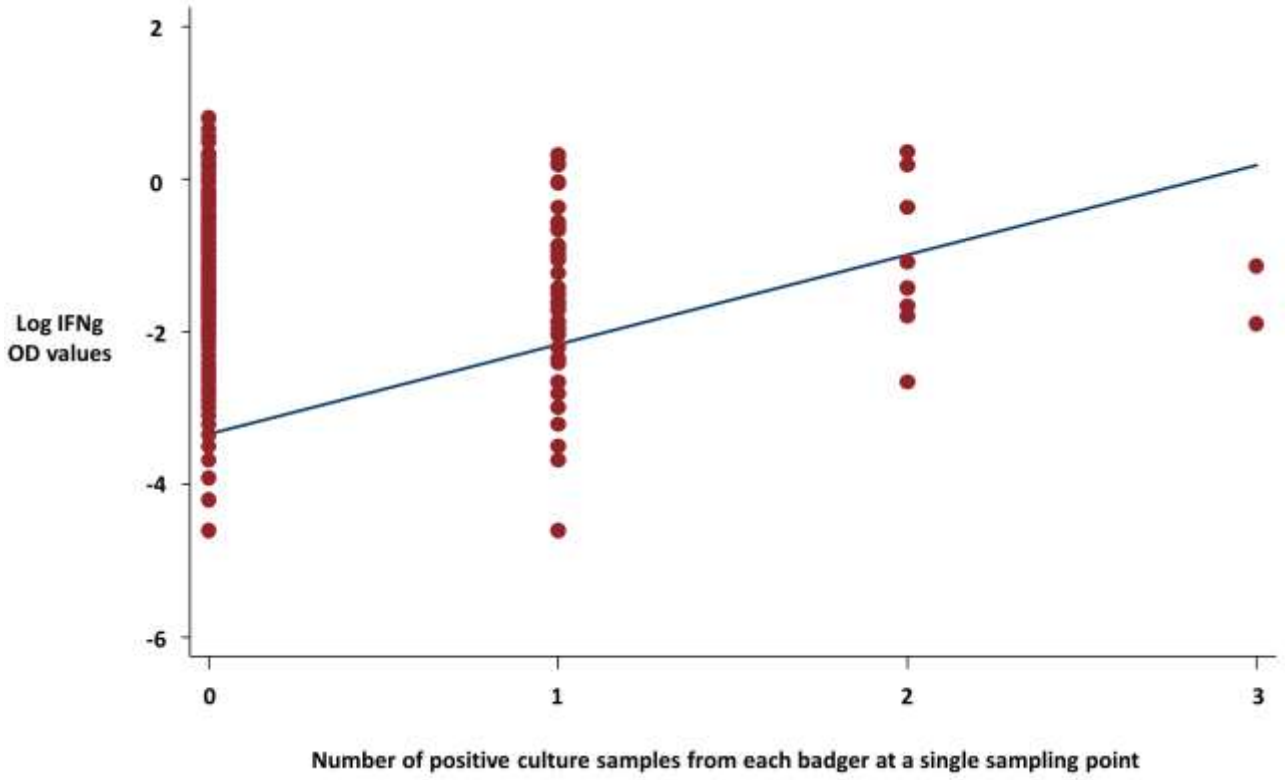


Table 3. The number of observations of live-captured badgers from Woodchester Park (July 2006 to October 2013) in which none, one or more clinical samples from a single capture event were *M. bovis* positive on culture. These data were used for the analysis of relationships between culture and IFNg tests results in 546 live badgers.

Number of positive culture samples per badger	**Number of observations**	**Percentage of observations**
0	2,149	97.5
1	44	2.0
2	10	0.5
3	2	0.1
Total	**2,205**	**100.0**

The likelihood of obtaining a positive culture result from urine, faeces or bite wound swab samples was positively correlated with the INFg OD value. Hence, predicted increases in OD values were 0.035 (z-test = 3.68, $p<0.001$; 95% CI: 0.018 – 0.069) for a positive urine culture result, 0.033 (z-test = 2.77,

p=0.006; 95% CI: 0.014-0.077) for a positive bite wound culture result, and 0.035 (z-test =3.19, p=0.001; 95% CI: 0.016-0.074) for a positive faeces culture result. However, INFg OD values were not significantly correlated with the probability of obtaining any other positive culture results for a tracheal or oesophageal aspirate, a swab from a burst lymph node or a non-bite-related wound.

Correlations between Stat-Pak and culture
StatPak and culture test results were significantly positively correlated (Pearson's chi-squared test = 147.63, p<0.001). However, for binary results of StatPak and culture (i.e. classifying the results of each as either positive or negative), a Kappa coefficient of 0.15 (p<0.001) indicated poor agreement between tests. A badger was much more likely to test positive on culture if its Stat-Pak test results were also positive (odds ratio = 78, 95% CI: 23.3- 266.4, z-test = 7.02, p<0.001: result from a mixed effect logistic regression).

Correlations between IFNg and post mortem lesion severity scores
There was weak positive correlation between *post mortem* lesion severity scores and INFg OD values on a continuous scale (Spearman's rank correlation coefficient = 0.21, p=0.001, indicating strong evidence for a weak correlation). However, there appeared to be strong evidence for an association between categories of INFg OD values and categories of *post mortem* lesion severity score (Pearson's chi-squared test = 27.37, p=0.007), suggesting that badgers in the lower INFg OD categories were likely to also have lower categories of lesion severity scores. But in contrast, simple linear regression did not provide a good fit to the data and the output was not statistically conclusive despite normal distribution of residuals. This suggests that the relationship between IFNg OD values and lesion severity scores may not be linear. To verify this, multinomial logistic regression and logistic regression models were fitted to the data. The multinomial regression model did not provide a good fit, while logistic regression fitted to binary INFg and lesion severity scores produced only a marginally better fit. In conclusion, there is no clear relationship between INFg test results and *post mortem* lesion severity score results.

Correlations between PCR and other live diagnostic tests
There was a positive relationship between IFNg OD values and PCR test results (linear regression with IFNg OD values on a continuous scale and PCR as the binary predictor variable). The likelihood of obtaining a positive PCR result was correlated with an increase in the IFNg OD value of 0.36 (z-test = -13.04, p<0.001) compared to an OD value of 0.04 (z-test = 3.10, p=0.002) when PCR was negative. For badgers with a positive IFNg result, the odds of also obtaining a positive PCR test result were 49 (95% CI: 3 – 739) times higher than for a negative PCR test result (z-test = 2.84, p=0.004: results from a mixed effect logistic regression model). A kappa coefficient of 0.5 (z-test = 5.09, p<0.001) suggested moderate agreement between PCR and INFg test results. No evidence of correlations between PCR and Stat-Pak or PCR and culture was detected using either the Kappa test or a mixed effects logistic regression.

Discussion
The positive correlations between StatPak and culture of clinical samples, and between the number of positive culture samples obtained from an individual badger and IFNg OD readings, suggests that both StatPak and IFNg may be potential predictors of the likelihood of shedding *M. bovis* bacilli. StatPak and culture test results are likely to be similar in any individual (either both test positive or both test negative) but this does not of course mean that they are accurate, particularly for negative test results owing to the limited sensitivity of both tests. The practical application of this correlation are that a badger testing StatPak positive is at least more likely to also test culture positive, and that the use of culture adds little to the diagnostic accuracy of *M. bovis* detection in live badgers. In the analysis of the relationship between IFNg results and lesion severity scores, the limited number of multi-site culture-positive badgers with concurrent IFNg results meant that the sample size was insufficient to produce a robust prediction model. The predictive ability of these tests in relation to disease progression is explored further in section (c), Objective 2 below. Also, a recent study by (Tomlinson *et al*., submitted) found that the magnitude of the IFNg response at the disclosing test event was positively correlated with subsequent progression of disease to a seropositive or excreting state.

Further evidence for the potential role of IFNg in predicting infectiousness arises from the observed positive correlation with PCR test results. The observation of no correlations between PCR and other

diagnostic tests could be real or may be a product of the small sample size, as only 87 badgers were included in this analysis which was limited by the number of PCR results available.

The results presented here support the contention that immunology is a poor predictor of anticipated pathology, as there was no clear correlation between IFNg test results and the severity of lesions identified at an immediately subsequent *post mortem* examination. This may be partly related to the pathology scoring system being relatively crude, but is also affected by the skewed distribution of the data (as most badgers had no visible lesions). More valuable information might be obtainable from experimental studies where there may be opportunities to use a more sophisticated pathology score in combination with immunological tests. However, the major drawback of such an approach is that it would be performed in the context of artificial infection which may not replicate the pattern of lesions seen in naturally-acquired infections.

In conclusion, evidence was found for positive correlations between IFNg and all other tests. The strongest correlations appeared to be between IFNg and both StatPak and culture of clinical samples. The correlation between IFNg and *post mortem* lesion severity scores appeared weaker. These results, in combination with those presented below (section (c), Objective 2), indicate that IFNg may be the most useful test for predicting current and future infection with *M. bovis* in live badgers. However, specialist facilities and the incubation of stimulated T-cells for 16-24 hours are required to currently perform the IFNg test, which severely limits its value as a test for use in field interventions to control disease in badger populations.

(b). The value of diagnostic tests applied at the badger social group level

Introduction

This series of analyses sought to investigate the predictive value of currently available diagnostic tests, when used both singly and in combination, to classify the infection status of badgers at a sett or social group level. This is a critical question for determining the potential value of existing tests to be used to identify infected badger social groups as part of any targeted intervention.

These analyses were theoretical simulations using diagnostic test performance values derived from the long-term study at WP where StatPak, IFNg and culture of clinical samples have been routinely applied to captured badgers with known affiliations to social groups. We sought to compare a range of sampling strategies in order to identify the optimum approach for detection of *M. bovis* infection at the badger sett or social group level using existing diagnostic tests. A critical component in these analyses is the performance of the diagnostic tests when used in isolation or in combination with one another. Test performance is therefore considered primarily in terms of the trade-off between sensitivity (the ability of a test to correctly identify true positive cases) and specificity (the ability of a test to correctly identify true negative cases). In order to replicate the likely scenario when trapping badgers to conduct diagnostic testing, we focused on determining the ability to correctly detect infection in social groups where not all badgers could be sampled. Analysis was conducted in two complementary stages, first by examining the performance of tests at the individual level, and then by examining the test characteristics when interpreted at the social group level.

Methods

This analysis took the form of a theoretical model where we simulated a range of approaches to examine how much each test result influenced the diagnosis of infection in groups of live badgers. This allowed us to therefore quantify the usefulness of each test in contributing to detection of infection at the sett or social group level. The sensitivity and specificity of each diagnostic test was estimated in the absence of knowledge of true infection status using Bayesian methods (Branscum *et al.*, 2005). Data were analysed using WinBUGS freeware (Spiegelhalter *et al.*, 1996) to run a Markov chain Monte Carlo model containing five over-dispersed chains. Priors for the sensitivity and specificity estimations of the three diagnostic tests were taken from Drewe *et al.* (2010). Prevalence was expected to vary over the study period and so was estimated on an annual basis using uniform (0, 1) priors. Estimates of sensitivity, specificity and prevalence were generated from 50,000 posterior samples collected after a burn-in of 5,000 iterations. Convergence was assessed by visual checking of trace plots of all chains

for each parameter. Code from the UC Davis website[1] was used, modified to fit the assumption of independence between tests which was considered appropriate because each test detects a different biological marker (i.e. antibody, cytokine, or bacteria; Cousins and Florisson, 2005). For this analysis we used the WP dataset from July 2006 to October 2013, which represented 2,022 capture (sampling) events involving 541 individual badgers.

Diagnostic tests were interpreted in parallel, meaning that they were considered together and an animal was categorised as infected if one or more of the tests yielded a positive result. We also investigated serial interpretation of test results, where test results from the same animal at any given capture event needed to all be positive in order for the animal to be considered infected.

A sample size of 15 animals per 'sett' was chosen as the unit for analysis, as this was a sufficiently large number for use in the necessary calculations. In reality, this number is more likely to represent the total social group size (at the higher end of the expected range in high density populations) rather than the number of occupants of a single sett. The average number of badgers per social group in Woodchester Park has been estimated at 9.4 (range 4.9-12.4; Delahay *et al*., 2013) and so in reality two main setts in close proximity may be considered together as the unit for this analysis. Results of tests were interpreted at an aggregated rather than an individual animal level, meaning that two or more badgers in a sett (or cluster of setts) would need to test positive in order for it to be considered infected. This threshold was chosen due to the limited specificity of tests, and hence it reduced the chances of incorrectly identifying a sett as positive when, in fact, there were no truly infected animals present (see also Woodroffe *et al*., 1999).

The performance of combinations of diagnostic tests was examined across a range of values for TB prevalence from 10% to 50%. Thus the 'true' number of infected individuals used for comparison in each case was calculated by multiplying each prevalence level, at intervals of 10%, by the number of badgers in the social group. This 'true' number of infected animals represents the situation that would be seen if the diagnostic tests were perfectly accurate (i.e. 100% sensitive and 100% specific).

The influence of the proportion of badgers trapped on diagnostic accuracy was another important consideration, so we tested the effects of a range of trapping efficiency values (10% to 100%). The results from various combinations of tests were assessed by comparing the numbers of infected animals identified by each combination of tests to the 'true' number of infected animals in the social group (estimated at varying prevalence intervals, and each time assuming 15 animals per social group as the unit of study).

Finally, we used an alternative complementary approach to examine the accuracy of the testing regime at the social group level, by calculating the herd sensitivity and herd specificity; these are epidemiological terms which refer to the ability of test(s) to correctly identify infected groups as positive and uninfected groups as negative (Dohoo *et al.* 2009). In this application, 'herd' refers to a badger social group, 'herd sensitivity' refers to the ability of diagnostic test(s) to correctly identify badger groups infected with *M. bovis*, and 'herd specificity' refers to their ability to correctly identify uninfected badger groups. Herd-level sensitivity is calculated when individual animal test results are interpreted at an aggregated (group) level. A certain (stated) number of animals need to test positive in order for the herd to be considered positive.

Herd-level sensitivities and specificities were calculated as follows (from Dohoo *et al.,* 2009):

$$AP = P * Se + (1 - P)(1 - Sp) \qquad \text{(Equation 1)}$$

$$HSe = 1 - \sum_0^{k-1} * C_{k-1}^n * AP^{k-1} * (1 - AP)^{n-(k-1)} \qquad \text{(Equation 2)}$$

$$HSp = Sp^n, \ when\ k = 1 \qquad \text{(Equation 3)}$$

$$HSp = \sum_0^{k-1} * C_{k-1}^n * (Sp)^{n-(k-1)} * (1 - Sp)^{(k-1)}, \ when\ k > 1 \qquad \text{(Equation 4)}$$

[1] http://www.epi.ucdavis.edu/diagnostictests/2dep1ind3t2p.html (accessed on 10 March 2014)

Where:

AP = the apparent prevalence (refers to the proportion of animals testing positive which is usually not the same as the proportion of animals actually infected, due to false negative and false positive results).
P =the true prevalence.
Se = the sensitivity of a diagnostic test (or combination of tests).
Sp = the specificity of a diagnostic test (or combination of tests).
HSe = the herd-level sensitivity (ability to detect infected groups).
k = the threshold number of animals required to test positive in order to consider the badger social group to be infected.
n = the number of animals tested.
C_k^n = the number of combinations of k positives when n animals are tested.
HSp = the herd-level specificity (ability to correctly identify uninfected groups). HSp is calculated assuming infection is absent (equations 3 and 4).

As can be seen from the formulae, the value of HSe is directly dependent on both the apparent prevalence and the number of animals tested. Conversely, HSp does not depend on infection prevalence, but is only sensitive to the number of animals tested and the chosen threshold number of animals required to test positive in order for a social group to be considered infected.

The following range of parameters was used for calculating the accuracy of diagnosis of *M. bovis* infection at the badger social group level:

1. Values of apparent prevalence ranged from 11% to 52%. These figures equate to a true prevalence range of 10% to 50%.

2. Social group size was set at 15 badgers, with the number of animals sampled per group ranging from 2 to 15.

3. The number of animals required to test positive in order for the herd (social group) to be considered positive ranged from 1 to 3.

Results

The sensitivity and specificity of each diagnostic test for diagnosing *M. bovis* infection in live badgers, estimated using Bayesian methods in the absence of knowledge of individuals' true infection status, are presented in Table 4.

Ability of each diagnostic test to detect infection at the social group level

Initially, tests were evaluated using a theoretical TB prevalence of 20% and a social group size of 15 animals. Under these assumptions, none of the tests when used singly was able to correctly identify all infected animals in the social group (Figure 2). However, in a scenario where the threshold for a sett to be categorised as positive was for two individuals to be detected as test positive, then StatPak appears able to detect infection at the group level if 90% of badgers are tested, and IFNg appears able to detect infection at the group level if 100% of badgers are tested. Within the parameters of this analysis, culture was not able to detect any infected animal (Figure 2).

In contrast, when all three diagnostic tests were interpreted in parallel at the sett level, a social group could be correctly identified as infected if only 50% of the animals were tested (0.5 on the x-axis in Figure 3). Two combinations of multiple tests [(Stat-Pak + INFg) and (Stat-Pak + INFg + Culture)] produced virtually identical results (top two lines – orange and brown – in Figure 1). This suggests that the addition of culture adds little to the diagnostic accuracy of the remaining tests for TB in live badgers.

Table 4. Estimated values for the sensitivity (Se) and specificity (Sp) of three diagnostic tests for the detection of *M. bovis* infection in live badgers, when used in isolation and in combination. Values estimated from empirical diagnostic test results from 2,022 sampling events involving 541 individual badgers trapped at Woodchester Park from July 2006 to October 2013.

Diagnostic approach	Combination of tests	Sensitivity (95% CI)	Specificity (95% CI)	Note
(a) Use of each test on its own	Stat-Pak	0.58 (0.53-0.63)	0.97 (0.93-0.99)	1
	INFg	0.52 (0.46-0.63)	0.97 (0.94-0.99)	1
	Culture	0.08 (0.06-0.11)	1.00 (0.99-1.00)	1
(b) Use of two or three tests together (parallel interpretation)	INFg + Culture	0.55	0.97	2
	Stat-Pak + Culture	0.61	0.97	2
	Stat-Pak + INFg	0.79	0.94	2
	Stat-Pak + INFg + Culture	0.81	0.94	3
(c) Use of two or three tests together (serial interpretation)	INFg + Culture	0.04	1.00	4
	Stat-Pak + Culture	0.04	1.00	4
	Stat-Pak + INFg	0.30	1.00	4
	Stat-Pak + INFg + Culture	0.02	1.00	5

Notes

1. Values estimated from the Woodchester Park dataset (2022 capture and sampling events involving 541 individual badgers from July 2006 to October 2013) using a Bayesian approach (Drewe *et al.,* 2010).
2. Values obtained by using formulae for two tests from Dohoo *et al.* (2009):
   $Se_{parallel} = 1 - (1-Se_x) * (1-Se_y)$
   $Sp_{parallel} = Sp_x * Sp_y$
3. Values obtained by using formulae for three tests from Dohoo *et al.* (2009):
   $Se_{parallel} = 1 - (1 - Se_x) * (1 - Se_y) * (1 - Se_z)$
   $Sp_{parallel} = Sp_x * Sp_y * Se_z$
4. Values obtained by using formulae for two tests from Dohoo *et al.* (2009):
   $Se_{series} = Se_x * Se_y$
   $Sp_{series} = 1 - (1 - Sp_x) * (1-Sp_y)$
5. Values obtained by using formulae for three tests from Dohoo *et al.* (2009):
   $Se_{series} = Se_x * Se_y * Se_z$
   $Sp_{series} = 1 - (1 - Sp_x) * (1 - Sp_y) * (1 - Se_z)$

Impact of prevalence on test performance

Given the requirement that two or more badgers must test positive in order for an infected social group to be correctly identified as infected, StatPak could achieve this but only when a large proportion of the group were sampled and prevalence was high (Figure 3a). For example, if prevalence was 20%, then the entire social group would need to be sampled in order to be able to achieve the required number of badgers testing positive. The sample size required reduces as prevalence increases so at 30% prevalence, two thirds of the group would need to be tested, at 40% prevalence, half the group would need to be tested and at 50% prevalence, 40% of the group would need to be tested. Where prevalence was less than 20%, Stat-Pak was unable to correctly identify an infected social group (Figure 3a).

Diagnostic ability was improved by combining StatPak with IFNg. In this scenario, both tests were run on every sampled animal and if either or both tests gave a positive result then it would be considered positive. The same requirement was used that two or more badgers must test positive in order for a social group to be correctly identified as infected. The combination of INFg and Stat-Pak was able to correctly identify social group-level infection status at any prevalence level, but only when at least 90% of badgers in a social group were tested. The main advantage of using both tests together over using StatPak alone was that a social group could be correctly identified as infected at lower (but not very low) prevalence levels (Figure 3b). Hence, whereas StatPak on its own was unable to correctly identify an infected social group where the background prevalence was less than 20%, even if the entire social

group was tested, the addition of IFNg meant that an infected group could be detected even when the prevalence was as low as 10%. Furthermore, using this combination of tests an infected group could be correctly identified when prevalence was 20% even when only half of the group were tested (compared to the requirement to test the entire group if using StatPak alone). At 30% prevalence, one third of the group would need to be tested (compared to two thirds of the group with StatPak alone), at 40% prevalence, one quarter of the group would need to be tested (compared to half of the group with StatPak alone), and at 50% prevalence, 20% of the group would need to be tested (compared to 40% of the group with StatPak alone). However, if prevalence dropped below 10%, then the entire social group would need to be sampled in order to be able to achieve the required number of badgers testing positive when using StatPak and IFNg in combination (Figure 3b).

Figure 2. The comparative ability of three diagnostic tests, when used singly and in combination (parallel interpretation) to detect TB infected badger social groups. In this example, there are three truly infected animals in a social group of 15 badgers (20% prevalence). The IFNg test when used in isolation identified the correct number of infected animals if 70% or more of the social group were tested. Combined use of StatPak and IFNg tests identified the social group as infected if 10% or more of the group were tested, and identified the correct number of infected badgers if 30% or more were tested. Under these assumptions, none of the tests when used in isolation was able to correctly identify all infected animals in the social group. In contrast, when StatPak and IFNg test results are interpreted in parallel at the social group level, a group may be correctly identified as infected if only 50% of the animals were tested. The addition of culture adds very little to the diagnostic accuracy. Results shown are theoretical simulations based on empirical data for StatPak, IFNg and culture test results from 2,022 sampling events involving 541 individual badgers trapped at Woodchester Park from July 2006 to October 2013.

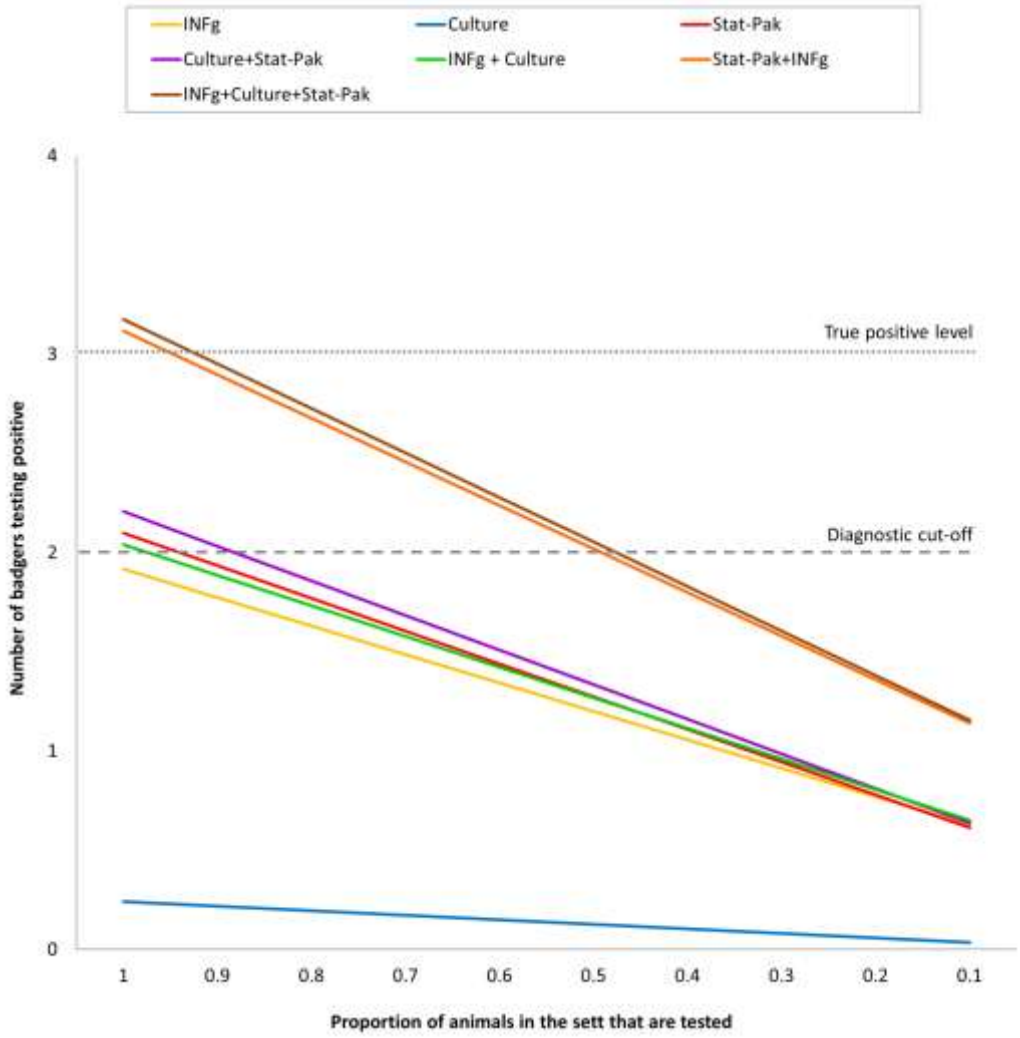
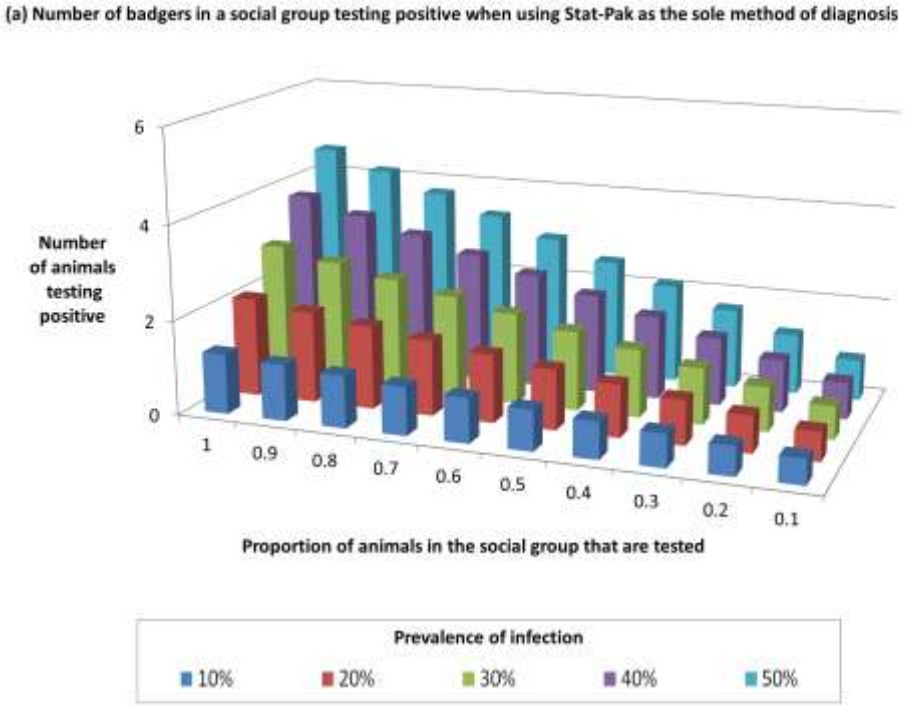
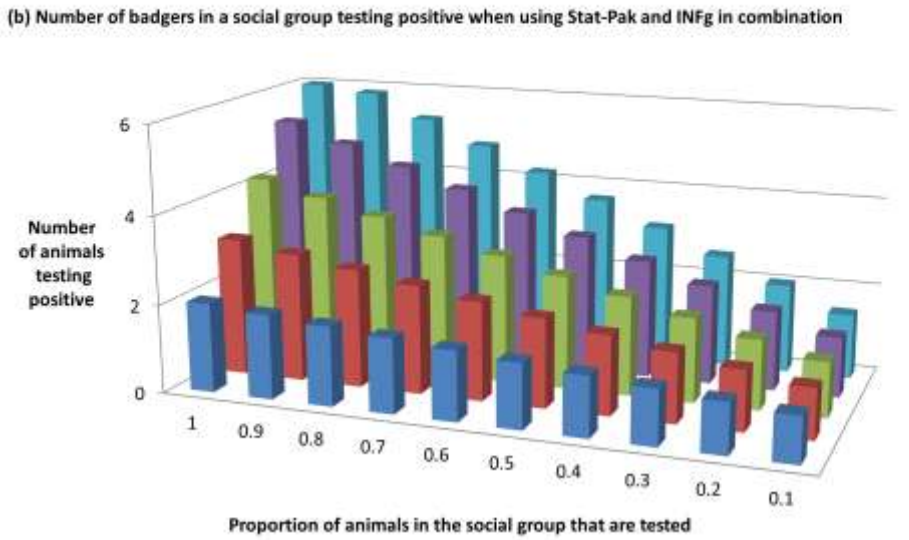


Figure 3. The influence of TB prevalence and the proportion of a badger group that is sampled, on the ability of diagnostic tests to identify infected badger groups. Graphs show the number of badgers identified as test positive across different values of background TB prevalence, using (a) StatPak in isolation, and (b) StatPak and INFg tests in combination (where both tests are run on every sampled animal and a positive diagnosis is made if one or both tests gives a positive result). Two animals testing positive was the threshold used for identifying infection in a social group of 15 animals. The combination of INFg and StatPak was able to correctly identify group-level infection status at any prevalence level but only when at least 90% of badgers in a social group were tested. In contrast, StatPak alone was unable to correctly identify an infected social group where the background prevalence was less than 20%, even if the entire group was tested. The addition of IFNg meant that infection could be detected where prevalence was as low as 10%. Results shown are theoretical simulations based on empirical data on StatPak, IFNg and culture test results from 2,022 sampling events involving 541 individual badgers trapped at Woodchester Park from July 2006 to October 2013.





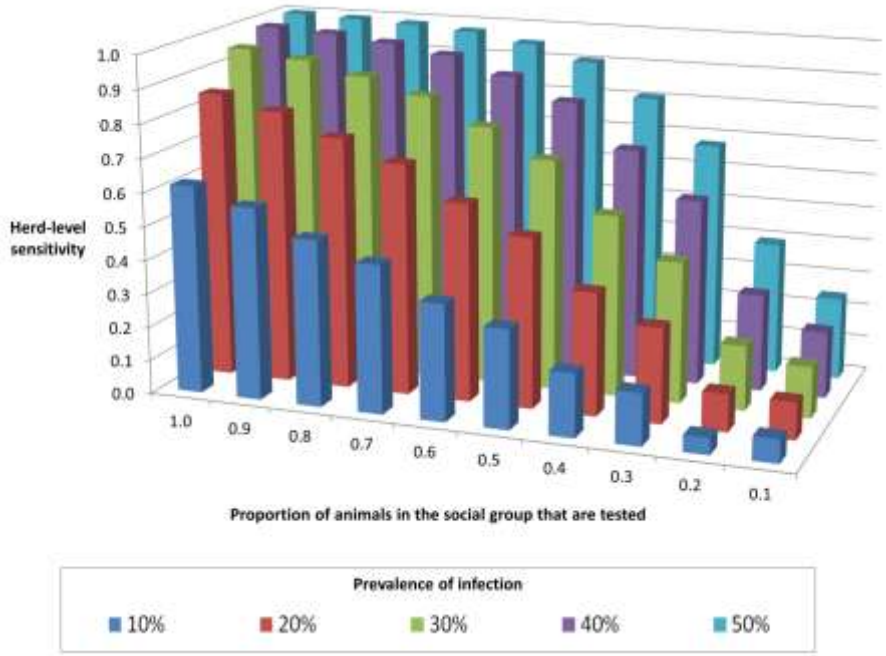
Impact of false positive results

It is important to note that because of the imperfect specificity of the tests some test positives are likely to in reality be uninfected false positives, and the impact of this potential problem increases as both (1) the prevalence decreases (resulting in a reduction in the positive predictive value, defined as the proportion of positive test results that are true positives) and (2) the proportion of the social group that is sampled decreases. For example, at a relatively high prevalence level of 50%, if 100% of a social group was tested, only 1 in 20 badgers that test positive would be false positives. At 20% prevalence the false positive rate rises to 1 in 5 test-positive badgers, and when prevalence is 10% and below, the false positive rate is 1 in 3 test-positive badgers. The impact of false positive results increased as the proportion of the social group that was tested decreased, such that with a prevalence level of 20% the false positive rate would be 1 in 4 test-positive badgers if 70% of the group were tested, 1 in 3 test-positives if 50% were tested and 1 in 2 test-positives where only 30% of the group was tested.

Herd-level sensitivity

The highest values of HSe for StatPak and INFg when used singly or combined in parallel were observed for the highest values of prevalence and for the highest proportion of badgers tested (Figure 4).

Figure 4. Effects of infection prevalence and the proportion of a badger group that is sampled on test sensitivity at the herd-level. In this example, a diagnosis of infection is based on the combined use of StatPak and IFNg in parallel with a threshold of two animals required to test positive for a social group to be considered infected. Results shown are theoretical simulations based on empirical data on StatPak, IFNg and culture test results from 2,022 sampling events involving 541 individual badgers trapped at Woodchester Park from July 2006 to October 2013.



The highest HSe values were obtained when a single badger was required to test positive, but this was at the expense of reduced specificity (i.e. there was an increased risk of incorrectly declaring an uninfected social group as infected). Increasing the threshold for a positive diagnosis at the group level (i.e. more badgers are required to test positive before a group is considered infected) reduced the chance of false positives but also led to lower herd-level sensitivity (Table 5).

Table 5. Variation in herd-level sensitivity values with different combinations of true infection prevalence, the proportion of animals tested, and selected cut-offs for concluding that a social group is infected, calculated using the combined use of StatPak and IFNg with their results interpreted in parallel.

*Figures quoted are herd-level **sensitivity** values*		**Proportion of animals in the social group that are tested**									
		1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
Threshold	**Prevalence (%)**										
1 positive animal required	50	1.00	1.00	1.00	1.00	0.99	0.98	0.96	0.92	0.81	0.57
	40	1.00	1.00	0.99	0.99	0.98	0.96	0.93	0.86	0.73	0.48
	30	0.99	0.99	0.98	0.97	0.95	0.92	0.86	0.77	0.63	0.39
	20	0.97	0.96	0.94	0.91	0.88	0.83	0.75	0.65	0.50	0.30
	10	0.89	0.86	0.82	0.78	0.73	0.66	0.58	0.48	0.35	0.19
2 positive animals required	50	1.00	0.99	0.99	0.98	0.95	0.91	0.81	0.68	0.39	0.24
	40	0.99	0.98	0.96	0.93	0.88	0.82	0.69	0.55	0.29	0.20
	30	0.95	0.93	0.89	0.85	0.77	0.69	0.54	0.42	0.19	0.15
	20	0.85	0.81	0.75	0.69	0.59	0.51	0.36	0.28	0.11	0.11
	10	0.62	0.57	0.49	0.44	0.35	0.29	0.19	0.15	0.05	0.07
3 positive animals required	50	0.98	0.97	0.94	0.90	0.82	0.73	0.51	0.40	0.08	0.00
	40	0.94	0.92	0.86	0.80	0.69	0.60	0.40	0.36	0.13	0.13
	30	0.84	0.79	0.70	0.63	0.49	0.42	0.22	0.21	0.02	0.06
	20	0.63	0.58	0.47	0.42	0.28	0.25	0.11	0.14	0.01	0.06
	10	0.33	0.30	0.21	0.20	0.11	0.12	0.04	0.08	0.00	0.05

Herd-level sensitivity was higher when StatPak and IFNg were interpreted in parallel, than when either of these tests was used individually. This difference was most pronounced at lower levels of *M. bovis* prevalence (Figure 5).

Herd-level specificity

Values of HSp provide information on how often a typical social group of badgers will incorrectly be declared infected when in fact it is disease-free, using diagnostic test(s) with a given HSe. Herd-level specificity was calculated using the same scenarios as for HSe, but this time assuming that infection is absent (HSp does not vary with prevalence).

Values of HSp increased as the threshold number of badgers required to test positive increased. For example, when interpreting StatPak and IFNg in parallel (when 50% of the social group was tested), the group would be incorrectly declared as infected 38% of the time when using a threshold of just one badger required to test positive, but only 9% of the time if at least two positive animals were required (Table 6). Conversely, HSp decreased as the proportion of the social group that was tested increased (recall that HSp is calculated assuming the absence of infection, hence any positive results are considered false positives and the frequency with which they occur increases with sample size). High values of HSp (greater than 95%) were obtained when 40% of the social group were tested and a threshold of two positive badgers was used (Table 6 and Figure 6).

Figure 5. Effects of three different methods of diagnosis on the herd-level sensitivity across a range of infection prevalence values. The scenario shown is based on 50% of badgers in a social group being tested, with a threshold of two animals required to test positive for the group to be considered infected. Results shown are theoretical simulations based on empirical data on StatPak, IFNg and culture test results from 2,022 sampling events involving 541 individual badgers trapped at Woodchester Park from July 2006 to October 2013.

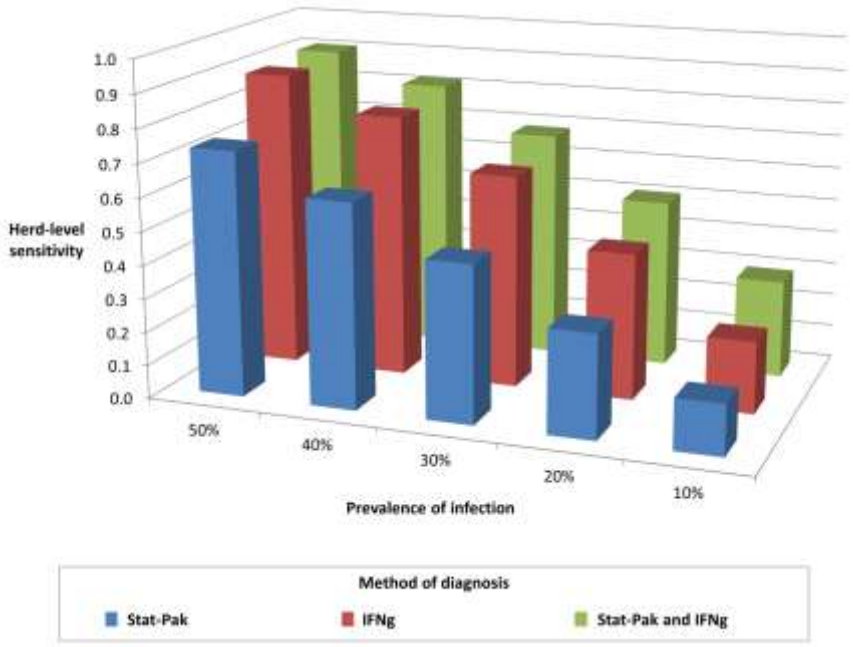
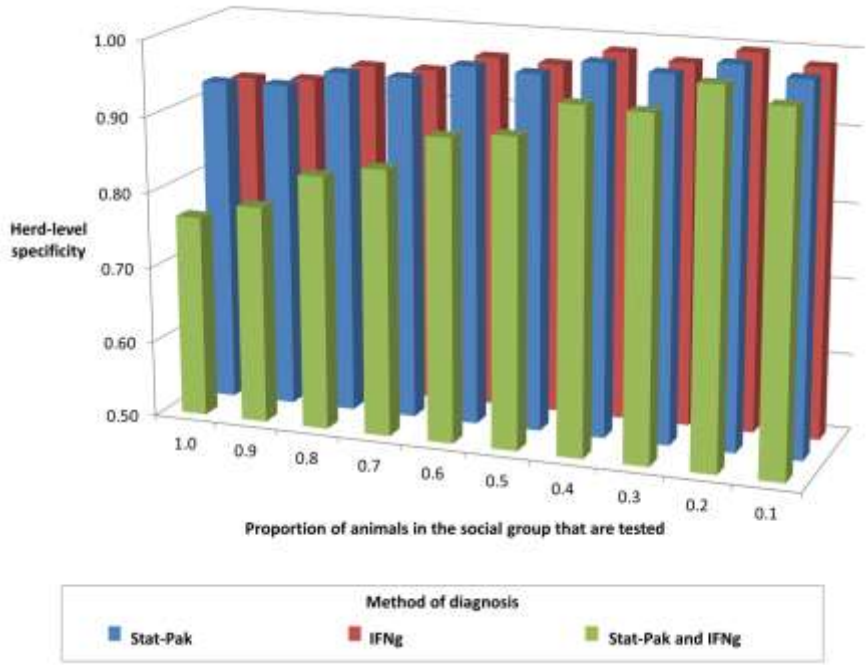


Table 6. Variation in herd-level specificity values with different combinations of the proportion of animals tested and selected thresholds for concluding that a social group is infected, calculated using StatPak and IFNg results interpreted in parallel. Note that these values were consistent across all levels of infection prevalence.

*Figures quoted are herd-level **specificity** values*	**Proportion of animals in the social group that are tested**									
	1.0	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
Threshold										
1 positive animal required	0.39	0.43	0.47	0.51	0.57	0.62	0.68	0.75	0.83	0.91
2 positive animals required	0.77	0.79	0.83	0.85	0.90	0.91	0.95	0.95	0.99	0.97
3 positive animals required	0.94	0.93	0.97	0.95	0.99	0.96	1.00	0.97	1.00	0.97

Figure 6. The influence of the proportion of a badger group that is sampled and the choice of test(s) on herd-level specificity. In this example, a threshold of two animals testing positive is required for a social group to be considered infected. Results shown are theoretical simulations based on empirical data on StatPak, IFNg and culture test results from 2,022 sampling events involving 541 individual badgers trapped at Woodchester Park from July 2006 to October 2013.



The HSp achieved when using StatPak and IFNg tests together and interpreting results in parallel was lower than the HSp for either test when used on its own at any threshold value (Figure 6). The opposite was true if the two tests were used together but the results were interpreted in series (i.e. both tests needing to be positive for an animal to be considered infected) due to the perfect specificity of this diagnostic approach (Table 4). However, the absence of false positives comes at the expense of a high probability of false negative results (i.e. reduced sensitivity resulting in missing cases of true infection: Table 4).

Discussion

On the basis of the calculations presented here, the sensitivity of StatPak, IFNg and culture of clinical samples was low when each test was used in isolation. For culture of clinical samples collected from live badgers, the sensitivity was estimated to be very low (less than 10%), for StatPak it was 58% and for IFNg it was 51%. Thus even the most sensitive test (StatPak) would be expected to miss 4 out of every 10 infected badgers. This level of false negative results would be expected to seriously limit the effectiveness of any disease control programme which aimed to use the StatPak as the sole means of detecting infection in live badgers.

It should be noted that the method of estimating the sensitivity and specificity of each diagnostic test (Bayesian latent class analysis: Branscum *et al.*, 2005) did not require knowledge of the true infection status. The figures quoted in the present study can be considered an update on the estimates previously published by Drewe *et al*. (2010) which were based on the same methods and used the same model priors. There are two notable differences in the estimates produced in the current study from those reported previously by Drewe *et al*. (2010) and Chambers *et al.* (2009), who calculated sensitivity and specificity by comparing test results to culture of *M. bovis* from necropsy tissues. Firstly, in the current analyses the StatPak is estimated to be slightly more sensitive than previously calculated (i.e. 58% in the current analyses vs 50% in Drewe *et al*. (2010) and 50% (adults) and 56% (cubs) in Chambers *et al.* (2009)). Secondly, the sensitivity of the IFNg test in the present study is estimated to be markedly lower than previously calculated (i.e. 52% in the current analyses vs 80% in Drewe *et al*. (2010) and 85% (adults) and 57% (cubs) in Chambers *et al.* (2009)). The likely explanation for differences between the findings of Drewe *et al*. (2010) and those of the current study is that the former

was based on test results from 875 capture events of 305 badgers caught over two years, wheras the latter involved results from 2022 capture (sampling) events involving 541 individual badgers caught over a seven-year period. Differences Also, the method used by Chambers *et al*. (2009) of estimating sensitivity and specificity by comparing the results of StatPak and IFNg tests with tissue culture is likely to overestimate test sensitivity because culture is itself of limited sensitivity, even when performed on necropsy tissues (Crawshaw *et al.*, 2008). This is a possible explanation why the estimates of IFNg sensitivity differ between the present study and Chambers *et al*. (2009). The estimates produced in the current study may therefore be considered more accurate and reliable.

There was little difference between the specificity of the StatPak, IFNg test and the culture of clinical samples, as all were within the range of 97-100%, and are comparable to the estimates given by Chambers *et al.* (2009). This suggests that when used individually, no test would be expected to have a false positive rate greater than 3%.

Parallel interpretation of the results of tests used in combination was adopted because this improves sensitivity, by multiplication of individual tests sensitivities. In contrast, the specificity of a combination of tests is lower than that of individual tests. Serial test interpretation was also investigated but although it improved the specificity of tests, this was at the cost of markedly lower sensitivity (Table 4) and consequently the risk of missing cases of infection was unacceptably high.

The interpretation of INFg and StatPak test results in parallel would be advisable during the initial stages of a disease control programme when prevalence is high, because in this scenario the proportion of test positives that are true positives is highest and the proportion of false positives is at its lowest. At this stage, where detection of infection is important, a diagnostic approach with a high negative predictive value (i.e. the proportion of negative test results that are truly uninfected) is likely to be preferred. As the control programme progresses so higher specificity becomes more important, to minimise the false positive fraction by correctly identifying all negative animals, and a diagnostic approach with a high positive predictive value is likely to be preferred. As the prevalence of infection is reduced, as would hopefully be the case later during the disease control programme, then it becomes increasingly undesirable to have high numbers of false positives, particularly in relation to demonstrating freedom from infection. Cut-off values for the sensitivity and specificity of tests should therefore be chosen in relation to the objectives of intervention and the stage of the disease control strategy.

The present study indicates that the combined use of several tests can improve diagnostic accuracy at the badger social group level when compared to the use of a single test. Although this was anticipated, in the present study we have been able to quantify the added benefit of combined approaches so that they can be assessed against the practical challenges of employing combinations of tests in the field. Two combinations [(StatPak + INFg) and (StatPak + INFg + Culture)] produced almost identical results, regardless of infection prevalence. Previous studies have shown the culture of clinical samples from live badgers to have very low sensitivity (Crawshaw *et al.*, 2008; Drewe *et al*., 2010), which is confirmed by the present analysis which suggests that it may be as low as 8%. Therefore, culture did not improve diagnostic accuracy when used in combination with the StatPak and/or INFg tests. Furthermore, the culture of clinical samples is expensive and time consuming, and hence is not a practical option for targeted disease control interventions in badger populations. It should however be noted that for the purposes of epidemiological research, clinical sampling can provide useful information on sources of infection and the progression of disease, particularly when used in combination with other tests (e.g. Delahay *et al*., 2000a & 2013; Tomlinson, in press).

Importantly, sensitivity analyses suggested that for the combination of INFg and StatPak tests to provide accurate results at the social group level (where a social group consists of 15 badgers in either a single sett or a cluster of nearby setts), estimates of trapping efficiency derived from the RBCT of 35-70% (Smith and Cheeseman, 2007) would be sufficient when infection prevalence levels are moderate or high (i.e. as might be expected at the start of a disease control programme). As prevalence reduces to below 10%, a higher proportion of the group would need to be sampled in order to accurately detect infected social groups.

The estimates generated here for herd-level sensitivity and specificity were comparable to those obtained from the individual-level analyses. The herd-level approach provided additional insight into the accuracy of diagnosis at different prevalence and trapping efficiencies, and suggests that the chosen threshold of two badgers per group required to test positive in order for the group to be considered infected appears robust.

It should be noted that the formulae used for interpreting diagnostic tests in combination assumed that the chance of detecting an infected animal is constant over time, within the limits of test performance. In reality this is unlikely to be the case, particularly for diseases such as TB, where detectability may be low just after the point of becoming infected and increase as the immune response develops (Chambers *et al.*, 2008). Consequently, diagnostic sensitivity in the field may differ from the results presented in this report. However, such effects may be mitigated by the combined use of a test to detect cell-mediated immunity (i.e. the IFNg test) and one which detects humoral immunity (i.e. the StatPak) which we would expect to detect infection in badgers at varying stages of disease progression.

Our results suggest some practical implications for the diagnosis of TB in badger populations in the field. In particular, the combined use of StatPak and INFg appear to provide practical opportunities to identify infected setts or social groups without the need to capture all individuals. Encouragingly, according to our results, levels of trapping efficiency that have been observed in previous field studies appear to be sufficient to usefully employ a combination of these tests to identify infected groups. The results of modelling empirical data from a 7-year period of the long-term WP badger study suggest that the combined use of INFg and StatPak can correctly identify infection at the level of the social group (or in a cluster of setts), using a threshold value of two badgers needing to test positive in order to consider a social group as infected, but the proportion of that badger social group that needs to be sampled in order to correctly detect infection increases as prevalence decreases. Under the same assumptions, StatPak used in isolation is able to detect infection at the group level only when a large proportion of the group are sampled and prevalence is high.

These findings help inform us of the desired characteristics that we may seek in novel diagnostic tests for use in selective management. In order to improve diagnostic ability at the social group level beyond that provided by existing tests, the sensitivity of any new test would need to be higher than 80% (the level achieved when using StatPak and IFNg together). Such a high level of sensitivity is likely to be difficult to achieve with a single test without compromising specificity, and hence the use of a combination of two (or even three) tests with slightly higher sensitivities than StatPak or IFNg has the potential to make a substantial practical difference in our ability to detect infection in badger groups. For example, if a diagnostic sensitivity of 90% could be achieved, this would allow a social group to be correctly identified as infected when only 10% or more of badgers were tested (under the model assumptions of 20% prevalence and a social group size of 15 badgers, and with the same threshold of two badgers required to test positive). The benefits of increased sensitivity include a reduction in the proportion of badgers that need to be tested and the ability to detect infection at lower prevalence.

It can be concluded that amongst the options investigated here, the most sensitive and specific diagnostic approach to detect *M. bovis* in badgers at the herd (social group) level is to use the StatPak and IFNg tests together, interpret their results in parallel, and use a threshold of two badgers required to test positive. However, there are considerable practical challenges to this approach related to the requirement for specialist laboratory facilities and time to run the IFNg test. In contrast, the StatPak is available in a rapid test format akin to a pregnancy test and can be conducted in about 30 minutes in the field (meaning badgers could be kept in the traps until the result was obtained). Keeping badgers in captivity for the 16-24 hours required to get a IFNg test result is likely to be impractical, and so animals would need to be released with the prospect of a subsequent capture event to cull on the basis of a positive test result (see modelled outputs below). It should be remembered however that if StatPak is used as the first (screening) test and two or more positive results are obtained, then the group would be considered infected and there would be no requirement for the IFNg test to be run in such circumstances.

(c). Modelling intervention strategies

Introduction

The purpose of these analyses was to investigate the potential practical outcomes of interventions including live testing of captured badgers compared to no control, vaccination and unselective culling approaches. Hence, we adjusted a current badger/cattle/bTB model which has already been used to consider a variety of disease control interventions (Smith *et al*., 2012).

Methods

By default we assumed that 70% of a given population would be captured during interventions, but for some scenarios we also investigated outcomes for 50% and 30% trapping efficacy. The full set of scenarios we investigated was as follows (italics indicate how each scenario is referred to in the subsequent text):

- *No control*: business as usual.
- Trap and *vaccinate* all animals without testing.
- Trap and *cull* all animals without testing.
- Trap, test using StatPak and IFNg in combination and cull the whole group when either test result is positive: a whole group *two-test group cull*.
- Trap, test using only the StatPak and cull the whole group when a positive result occurs: a *one test group cull.*
- Trap, test using the StatPak and only cull test positive animals, but vaccinate and release all negative individuals: the *TVR* approach.

Strategies involving the whole group being culled on the basis of positive test results were investigated using thresholds of either one (trigger = 1) or two (trigger = 2) positive animals required to trigger culling of the group (Table 7).

In all simulations it was assumed that StatPak sensitivity was 50% and specificity was 96.9%, IFNg and StatPak combined sensitivity was 79.9% and specificity was 95%. We also assumed that the duration of control was for five years, the area subject to control was 150km^2, trapping was carried out in June, the mean social group size was 6.7 badgers, and all trapped animals could be correctly assigned to their social group. Each scenario was run for 100 simulations. We monitored the effect on the badger population within the core control area and also in a 2km ring around this as this is where the epidemiological effects of social perturbation have been identified during culling (Donnelly *et al*., 2006). For the group level culling we tested two options (a) all trapped animals are held and tested until the results are available and then the whole group is culled or released and (b) animals are trapped, released, and if results indicate a positive group, then a second trap up occurs for the purposes of culling and with the same probability of capture.

Results

In the absence of perturbation, the untargeted cull strategy (i.e. culling with no testing) removed a mean of 1432 badgers, whereas the test and cull strategies removed between 173 and 840 badgers (Table 7). Of the strategies simulated here, the largest reductions in numbers of infected badgers were observed for the untargeted cull strategy. The next largest reduction in the absolute number of infected badgers (assuming no perturbation) was achieved by the TVR approach; then the two-test cull (with no real difference between the trigger points); then the one-test group cull then vaccinate. The next largest reduction in the absolute number of infected badgers (assuming no perturbation) was achieved by the TVR approach; trigger=1 and two-test cull; whole group cull and vaccinate. If we assume that perturbation occurs whenever culling takes place then the untargeted cull strategy removed a mean of

1405 badgers, whereas the test and cull strategies removed between 215 and 924 badgers (Table 7). The simulations show an initial increase in numbers of infected badgers (owing to enhanced transmission arising through increased post-cull movements) and order of preference of the different scenarios changes (Figure 7). With perturbation the untargeted cull and two-test cull with trigger =1 give the greatest reduction in numbers of infected badgers, the whole group cull with a trigger =2 yields the smallest reduction and all the other approaches have similar intermediate outcomes.

Table 7. The number of badgers culled during each of the simulated scenarios.

Trap efficacy (%)	Cull	TVR	Cull trigger = two positives/grp		Cull trigger = one positive/grp	
			One test group cull	Two test group cull	One test group cull	Two test group cull
With perturbation						
30	795	119	49	206	233	361
50	1142	183	170	433	481	656
70	1405	215	284	598	680	924
Without perturbation						
30	807	85	33	138	174	291
50	1174	128	110	295	357	559
70	1432	173	219	494	560	840

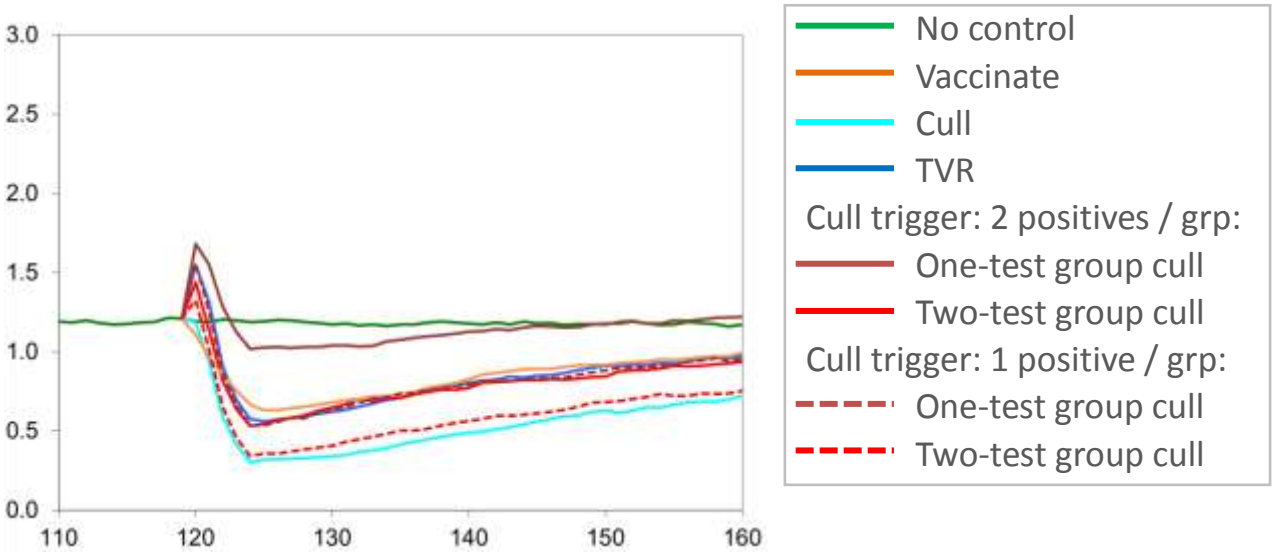


Figure 7. The simulated effect of the various strategies on the mean number of infected badgers per social group in the core control area, assuming that social perturbation occurs with any level of culling. Management occurs from year 120 to 124 and then the population recovers.

If no perturbation occurs then there is no negative effect (i.e. no increase in numbers of infected badgers) in the 2km ring outside the control area. However, if perturbation does occur, then in the 2km surrounding ring the number of infected badgers increases for a short time for all strategies except vaccination. When trapping efficacy was reduced to 50% and perturbation occurred then all strategies involving culling also led to a short-term increase in the number of infected badgers in the core area, with the whole group cull never returning to the no control level.

If we assume that all trapped badgers are vaccinated if they are not culled, then the difference between the strategies is much reduced (Figure 8). Note that outputs for the cull-only, vaccinate only and TVR strategies do not change, but the efficacy of all other scenarios (in terms of the reduction in the number of infected badgers) is improved.

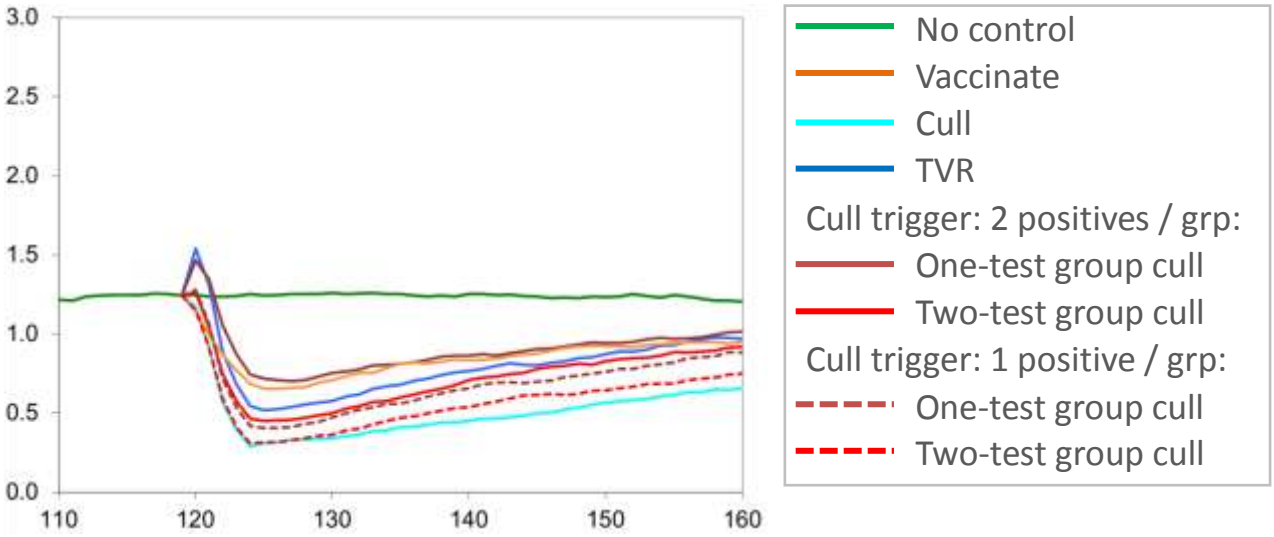


Figure 8. The simulated effect of the various strategies on the number of infected badgers in the core control area, assuming that social perturbation occurs with any level of culling and all badgers that are released are also vaccinated.

Discussion

As with previous work, the presence of social perturbation counters the effect of culling to some extent in the core, and particularly in the outer 2km ring. The use of the StatPak test at the group level with a threshold of two positive results required to trigger a group-wide cull was always the worst strategy. The combined StatPak and gamma test at the group level with a trigger of one positive result leading to a group level cull was very similar to an untargeted cull in terms of the reduction in the number of infected badgers. However, the selective approach resulted in about one third less badgers being killed. The remaining strategies gave similar levels of reduction in numbers of infected badgers but differed in respect to the number of badgers killed and would vary in cost. Most of the test and cull strategies investigated here were improved to some extent when test-negative groups were vaccinated, although we note that none of these strategies led to an overall improvement on the untargeted culling strategy.

Although these analyses simulate the effect of various intervention strategies on numbers of infected badgers rather than cattle herd breakdown (CHB) rates, it is expected that the relative benefits of the different approaches will be the same. However, it is predicted that CHB outputs will be more variable, and the absolute level of reduction will depend on the proportion of cases caused (directly or indirectly) by badgers.

Objective 2: Disease progression and test outcomes

Introduction

The purpose of these analyses was to investigate whether IFNg test results in badgers could be used as predictors of future positive results on other diagnostic tests. If this were the case then it might be possible to use IFNg titres as a predictor of future disease progression.

Microbiological culture of samples identifies the presence of viable *M. bovis* organisms, while the StatPak test identifies the presence of antibodies produced in response to exposure to *M. bovis*. Positive culture results are considered to represent infectiousness, although positive StatPak results have also been reported to indicate infectious badgers, due to the correlation between time of antibody production (or detection) and the onset of infectiousness (Chambers *et al*., 2010). The IFNg test provides a quantitative measure of cytokine production in blood, levels of which are raised following infection with *M. bovis* (Dalley *et al*., 2008). We hypothesise therefore that badgers with high values of IFNg may be more likely to produce a positive subsequent response to other TB tests (i.e. StatPak and culture of clinical samples) than are badgers with low IFNg titres. Should this be the case, then it would be possible to quantify the ability of IFNg titres for predicting future diagnostic test results, and as a result improve the certainty in the diagnosis obtained using the IFNg test. In the present study, we used a retrospective cohort analysis to compare the risk (probability) of subsequent positive StatPak and culture test results amongst badgers with different IFNg test results.

Materials and methods

Data source and description

Longitudinal data used in this analysis were collected from the WP badger population from July 2006 to October 2013. Data included IFNg, StatPak and culture test results on 550 captured badgers. Badgers were enrolled in the study on the date of their first IFNg test (usually the first time they were sampled within the study period) and were followed until the date of their last StatPak or culture test during that period. The median total observation period per badger was 10 months (range: one day to 86 months per badger). Badgers with one day observation time were trapped and tested just once, therefore, a true follow-up time period was not recorded for them. But because their infection status might have been different to those who were trapped more than once, it was decided to enter the test results of these badgers into the analysis by artificially increasing the time period between IFNg and subsequent tests by one day.

The IFNg titre (i.e. the quantity of IFNg produced from PPD-B stimulation minus that produced by PPD-A stimulation: Dalley *et al*., 2008) at the time of first testing was the 'risk factor' of interest and was included as the explanatory variable in the model. Therefore, one IFNg test result was used per badger. Gamma interferon titres (optical density values) on a continuous scale were grouped into five categories (see Table 8) for analyses. Categories 0 and 1 were both below the cut-off value for infection (currently set at 0.044; Dalley *et al*., 2008) and categories 2, 3 and 4 were all above this threshold (Table 8).

Table 8. Categorisation of IFNg test results in ascending order of optical density (OD values). Data represent the first IFNg result obtained from each of 550 badgers sampled at Woodchester Park from July 2006 to October 2013.

IFNg category	IFNg OD values (PPD-B minus PPD-A)	Number of observations	Percentage of observations
0	<0	181	33
1	0.000 - 0.043	277	50
2	0.044 - 0.366	66	12
3	0.367 - 0.696	15	3
4	0.697 - 1.920	11	2
Total	**0.000 - 1.920**	**550**	**100**

Multiple StatPak and culture test results were recorded from most badgers. There were 2,342 StatPak results for these 550 badgers, with a median of 3 per badger (range: 1 to 21 per badger). StatPak results were binary (either negative or positive). The distribution of Stat-Pak results by category of IFNg is presented in Table 9. There were 2,388 mycobacterial culture results for these 550 badgers, with a median of 3 per badger (range: 1 to 21 per badger). Culture test results were also binary (either negative or positive). The distribution of culture test results by category of IFNg is presented in Table 10.

Table 9. Frequency distribution of StatPak test results by category of IFNg OD values. Data are derived from 2,342 StatPak tests performed on 550 badgers at Woodchester Park from July 2006 to October 2013.

IFNg category	IFNg OD values (PPD-B minus PPD-A)	Number (%*) of StatPak test results		
		Negative	Positive	Total
0	<0	764 (84)	143 (16)	907
1	0.000 - 0.043	913 (81)	208 (19)	1121
2	0.044 - 0.366	134 (60)	91 (40)	225
3	0.367 - 0.696	24 (56)	19 (44)	43
4	0.697 - 1.920	24 (52)	22 (48)	46
Total	**0.000 - 1.920**	**1,859 (79)**	**483 (21)**	**2,342**

*Percentages of positive and negative Stat-Pak test results within each IFNg category.

Table 10. Frequency distribution of the results of mycobacterial culture of clinical samples in relation to different categories of IFNg OD result. Data from 2,388 culture tests performed on 550 badgers at Woodchester Park from July 2006 to October 2013.

IFNg category	IFNg OD values (PPD-B minus PPD-A)	Number (%*) of culture test results		
		Negative	Positive	Total
0	<0	898 (97)	24 (3)	922
1	0.000 - 0.043	1,117 (97)	30 (3)	1,147
2	0.044 - 0.366	219 (96)	10 (4)	229
3	0.367 - 0.696	40 (95)	2 (5)	42
4	0.697 - 1.920	45 (94)	3 (6)	48
Total	**0.000 - 1.920**	2,319 (97)	69 (3)	**2,388**

*Percentages of positive and negative culture test results within each IFNg category.

Data analysis

Results of the diagnostic tests were modelled using a Cox proportional hazards regression analysis to estimate the rates (probabilities) of positive StatPak or culture results relative to the different categories of explanatory variable (IFNg). Survival analysis was chosen because this method focuses on 'time-to–event' which permits the calculation of rate ratios, which are relative measures of association. The time between two different TB tests (i.e. IFNg and StatPak, or IFNg and culture) for each badger was the input for this analysis, to determine whether values of IFNg can be used as predictors of subsequent infectious individuals.

Data was formatted to meet the requirements of the Cox regression analysis. The time that elapsed between subsequent diagnostic tests was recorded for every animal. As a result, every badger had a number of time periods that corresponded to the number of times they were tested. The results of mycobacterial culture of different clinical samples (e.g., urine, faeces, tracheal aspirate) were pooled into one culture result (positive or negative) per badger per trapping event.

Preliminary data exploration employed Chi-square tests to assess associations between the categories of the independent variable (IFNg) and the dependent variables (StatPak, culture). A log-rank test was used to assess equality in survival function between categories of independent variable to determine whether the differences in survival between groups were more than would be expected by chance alone (Allison, 1984).

Kaplan-Meier and Nelson-Aalen curves were plotted for visual assessment of data distribution and to check if the proportional hazards assumption was upheld. Data were formally assessed using a plot of −log (−log) survival lines and a Schoenfeld residuals test (Grambsch and Therneau, 1994) which revealed that the proportional hazards assumption was not met. Therefore, data were corrected by splitting the observation time into three-month intervals (Table 11). A Schoenfeld test indicated that following this step the data no longer violated the proportional hazards assumption.

Table 11. The distribution of StatPak test results across the time periods used in the analysis. Data is derived from 2,342 StatPak tests performed on 550 badgers at Woodchester Park from July 2006 to October 2013 (total study period = 86 months).

Time period		**Number of StatPak test results obtained per time period**	**Number (%) of positive StatPak test results obtained per time period**
Short term	0 to 3 months	744	111 (15)
	0 to 6 months	1,013	144 (14)
	0 to 9 months	1,134	166 (14)
	0 to 12 months	1,370	215 (16)
Medium term	12 to 24 months	433	101 (23)
Long term	24 to 86 months	539	167 (30)
Total	**0 to 86 months**	**2,342**	**483 (20)**

A Cox regression model was fitted to the data, specifying the clustering of multiple observations per badger, using Stata version 11.2 (Statacorp LP, College Station, Texas, USA). Final models were checked for goodness of fit by using Cox-Snell residuals (Cox and Snell, 1968). The hazard function followed approximately the 45^{o} line and was exponentially distributed with a hazard ratio that approximated one. Therefore it was concluded that the data fitted the models adequately.

Estimates of rate ratios (the relative probabilities of subsequently obtaining a positive StatPak or culture result following a given IFNg result) were produced for inter-test periods of three, six, nine and 12 months, and for the period between 12 and 24 months. For time periods greater than 12 months, annual time categories were used (1 to 2 years, 2 to 3 years, etc.) until the end of the study (up to just over seven years).

Results

Chi-square tests produced evidence against the null hypothesis of no association between categories of IFNg and StatPak test results ($\chi^2 = 105.7$, $p < 0.001$). Associations between categories of IFNg and culture results were present for observation periods up to nine months only ($\chi^2 = 21$, $p < 0.001$). Log-rank tests indicated that the survival function of StatPak test results was not the same for all categories of IFNg in all time periods and for the culture test there was good evidence against equality for a follow-up period of less than a year. This suggests that a difference exists between badgers of different IFNg titres in relation to the probability of subsequently testing positive on StatPak or culture.

For follow-up periods in excess of 24 months, the proportional hazards assumption was violated for both StatPak and culture tests. Consequently, only observations made within two years of each badger's first IFNg result were included in subsequent analyses. This meant that any effect of IFNg titre on other diagnostic test results more than 24 months later could not be ascertained.

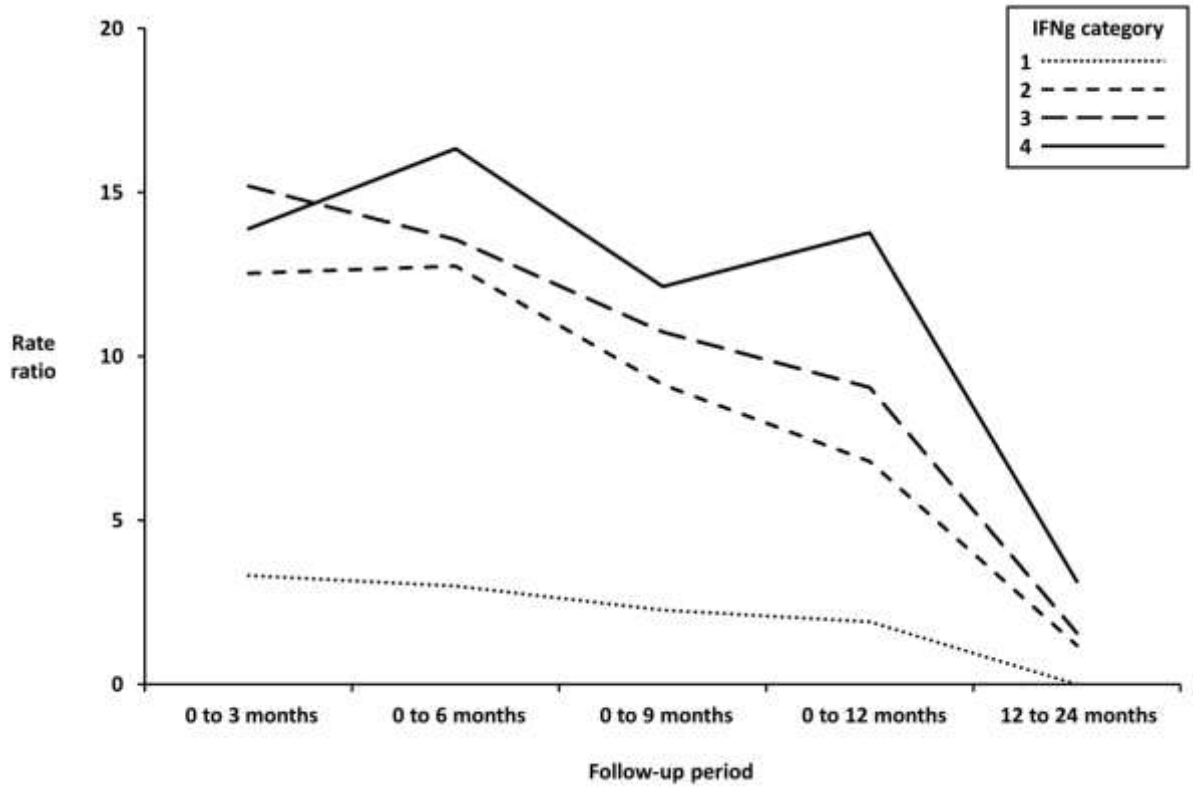
Can IFNg be used as a predictor of future StatPak test results?
The highest rates of StatPak positive test results were observed following the highest IFNg optical density values throughout all time periods up until two years (Table 12). Predictive ability gradually declined over a year and became inconclusive when follow-up time was more than 12 months (Table 12 and Figure 9). An exception was for badgers in the highest IFNg category, where the correlation was sustained over the longest time period (up to 24 months between IFNg and StatPak tests being conducted on the same badger), albeit with a rate ratio of only 3.1 (Table 12).

Table 12. Relative incidence of positive Stat-Pak results in badgers over different time periods subsequent to a IFNg titre result. Data are from five Cox regression models, each of which was run for a different time period (defined as the interval between the IFNg test being conducted and a subsequent StatPak test on the same badger). Rate ratios were calculated by comparing the incidence of positive StatPak test results for badgers in each IFNg category to a baseline rate (category zero in Table 8), which was allowed to vary by time period (proportional hazard assumption). Significant differences from baseline are shaded in grey. As an example, to determine the relative chance of a badger with an IFNg titre of 0.50 subsequently testing StatPak positive 12 months later, compared to a badger with an initial IFNg titre of zero, first determine the category of IFNg using Table 8: in this example it would be category 3 (because the IFNg OD value falls within the range of 0.367 - 0.696). The rate ratio of 9.05 can be interpreted as badgers with an IFNg OD value of 0.50 being nine times as likely to test StatPak positive up to a year later than are badgers with an IFNg OD value of zero.

Time period	IFNg category*	Rate ratio	SE	z	P >|z|	95% CI
0 to 3 months	1	3.32	1.19	3.35	0.001	1.64 - 6.70
	2	12.53	4.54	6.99	0.000	6.17 - 25.48
	3	15.19	6.86	6.03	0.000	6.27 - 36.83
	4	13.90	6.22	5.89	0.000	5.79 - 33.42
0 to 6 months	1	3.00	0.94	3.53	0.000	1.63 - 5.54
	2	12.75	3.97	8.18	0.000	6.93 - 23.46
	3	13.56	5.12	6.91	0.000	6.47 - 28.41
	4	16.33	6.63	6.88	0.000	7.37 - 36.18
0 to 9 months	1	2.27	0.66	2.81	0.005	1.28 - 4.02
	2	9.14	2.72	7.43	0.000	5.10 - 16.38
	3	10.75	3.72	6.87	0.000	5.46 - 21.17
	4	12.13	4.81	6.30	0.000	5.58 - 26.38
0 to 12 months	1	1.91	0.56	2.40	0.016	1.13 - 3.43
	2	6.80	2.00	6.51	0.000	3.82 -12.12
	3	9.05	3.66	5.44	0.000	4.09 - 20.00
	4	13.77	5.05	7.16	0.000	6.72 - 28.24
12 to 24 months	1	0.96	0.29	-0.14	0.892	0.53 - 1.74
	2	1.18	0.45	0.43	0.669	0.56 - 2.50
	3	1.56	0.76	0.91	0.364	0.60 - 4.08
	4	3.14	1.69	2.12	0.034	1.09 - 9.02

*IFNg categories are detailed in Table 1.

Figure 9. Rate ratios for subsequently obtaining a positive StatPak test result in badgers after varying follow-up periods, in relation to their initial IFNg titre (category 1 = lowest IFNg titre; category 4 = highest IFNg titre: see Table 8 for details of categories) compared to badgers with a negative IFNg titre. A badger with a rate ratio of 15 has a 15 times higher chance of testing positive on StatPak within the indicated follow-up period than a badger with a negative IFNg titre. Data is derived from badgers sampled at Woodchester Park from July 2006 to October 2013.



Can IFNg be used as a predictor of future culture test results?
IFNg results were of less value in predicting future culture test results than they were at predicting subsequent StatPak results. Only badgers with the highest IFNg OD values (category 4), predicted a future positive culture result over every time interval up to two years (Table 13). Low numbers of positive culture test results explain the wide confidence intervals and why reliable estimates could not be produced for badgers with lower IFNg OD values. The ability of IFNg to predict subsequent positive culture test results declined as follow-up time increased (Figure 10).

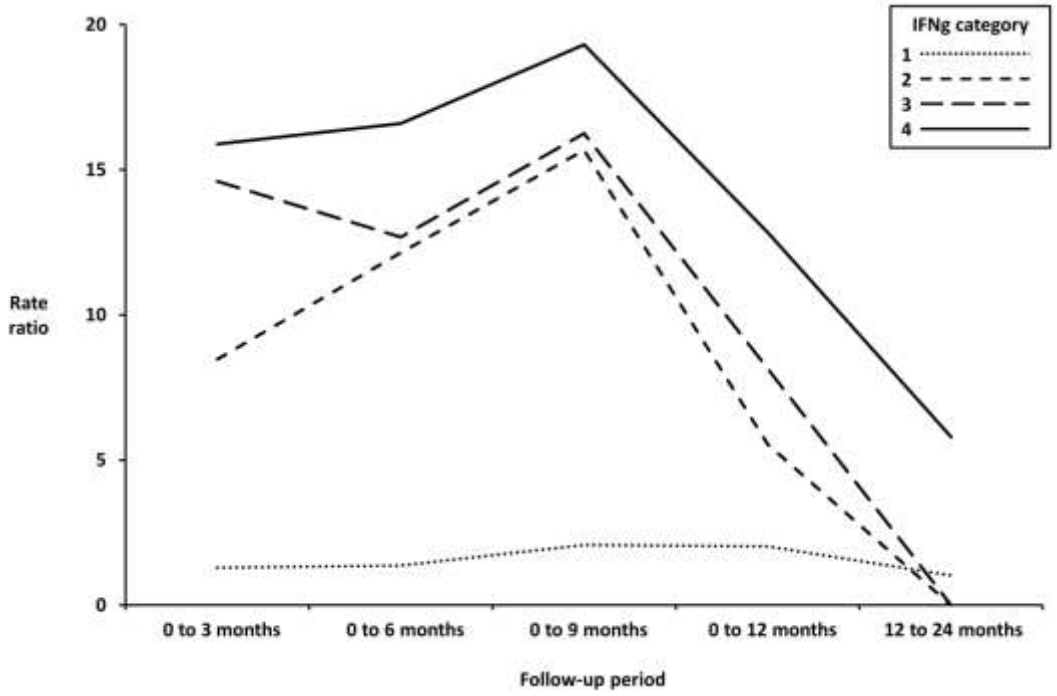
Putting these findings in context, the predictive ability of IFNg over one subsequent year can be seen by using as an example badgers with IFNg OD values equal to or greater than 0.044 (the current cut-off for a badger to be considered infected with *M. bovis*: Dalley *et al*., 2008). These badgers had at least a six times higher chance of subsequently testing positive on StatPak and culture within 12 months than did animals testing negative (i.e. IFNg OD < 0.044). IFNg results remained correlated with other test results two years later but the association was less pronounced: badgers with the highest INFg OD values (> 0.70) had at least a three times higher chance of subsequently testing positive on StatPak, and at least a five times higher chance of subsequently testing positive on culture, than animals with IFNg OD values of zero.

Table 13. The relative incidence of positive culture results in badgers in relation to previous IFNg titre results. Data are derived from five Cox regression models, each of which was run for a different time period (defined as the interval between the IFNg test being conducted and a subsequent StatPak test on the same badger). Rate ratios were calculated by comparing the incidence of positive StatPak test results for badgers in each IFNg category to a baseline rate (category zero in Table 8), which was allowed to vary by time period (proportional hazard assumption). Significant differences from baseline are shaded in grey. For method of interpretation, see legend to Table 12.

Time period	IFNg category*	Rate ratio	SE	z	P >|z|	95% CI
0 to 3 months	1	1.29	1.57	0.21	0.834	0.12 - 14.04
	2	8.47	9.80	1.85	0.065	0.88 - 81.71
	3	14.60	20.76	1.88	0.060	0.90 - 237.22
	4	15.89	22.61	1.94	0.052	0.98 - 258.61
0 to 6 months	1	1.37	1.69	0.26	0.796	0.12 - 15.26
	2	12.14	13.75	2.21	0.027	1.32 - 111.68
	3	12.69	18.13	1.78	0.075	0.77 - 208.90
	4	16.60	23.66	1.97	0.049	1.02 - 271.18
0 to 9 months	1	2.08	2.41	0.63	0.530	0.21 - 20.23
	2	15.68	17.27	2.50	0.012	1.81 - 135.70
	3	16.26	23.54	1.93	0.054	0.95 - 277.51
	4	19.32	27.87	2.05	0.040	1.14 - 326.78
0 to 12 months	1	2.03	1.35	1.07	0.284	0.55 - 7.47
	2	5.54	4.04	2.34	0.019	1.32 - 23.17
	3	8.15	9.68	1.77	0.077	0.80 - 83.49
	4	12.84	14.94	2.19	0.028	1.31 - 125.82
12 to 24 months	1	1.03	0.67	0.05	0.959	0.30 - 3.61
	2	0	0	-63.93	0.000	0 - 0
	3	0	0	-51.75	0.000	0 - 0
	4	5.80	3.99	2.56	0.011	1.508 - 22.33

*IFNg categories are detailed in Table 1.

Figure 10. Rate ratios for subsequently obtaining a positive culture test result after varying follow-up periods, in badgers categorised by their initial IFNg titre (category 1 = lowest IFNg titre; category 4 = highest IFNg titre: see Table 8 for details) compared to badgers with a negative IFNg titre. Significant differences from baseline were consistent over all time periods only in badgers within the highest IFNg OD value category (category 4, solid line). Data is derived from badgers sampled at Woodchester Park from July 2006 to October 2013.



Discussion

Badgers with the highest IFNg OD values were most likely to subsequently test positive on both StatPak and culture, and this effect was detectable for up to 24 months after the IFNg test. StatPak showed a positive trend in its dose-response relationships with IFNg, meaning that the higher the original IFNg OD value, the greater the chance that a badger would subsequently test positive on StatPak. The relationship between IFNg and culture was less clear, which may be a real effect but is likely to have been influenced by the low numbers of culture-positive badgers in the analysis.

Although IFNg test results are obtained on a continuous scale (OD values), the diagnosis of infection status is currently based on whether the value falls above or below the chosen cut-off (i.e. 0.044 at present). Hence, as the diversity in the range of OD values is not fully used for diagnosis, some information is lost. Results of our analyses indicate that by using the raw OD values it is possible to go beyond answering whether or not an animal is 'positive', and to potentially infer the stage of infection and the likelihood that it will subsequently test positive on other diagnostic tests. Those animals producing the highest values of IFNg (i.e. category 4 in the present analysis) were most likely to go on to also test positive on culture. This suggests that it may be possible to identify and target individuals most likely to go on to become infectious (where a positive culture result is interpreted as indicating shedding of *M. bovis* (Gavier-Widen *et al*., 2001)) by applying a cut-off for IFNg OD values of 0.697 (the lower boundary of our category 4) rather than the currently used cut-off of 0.044. Nonetheless, our results suggest that the current cut-off is useful, as badgers with an OD value greater than or equal to this cut-off are likely to go on to test positive on StatPak, which has also been shown to be a proxy for infectious badgers (Chambers *et al*., 2009).

This study suggests that by examining the quantitative value of an IFNg test result, rather than simply considering it as positive or negative, it is possible to predict the likelihood of subsequent StatPak, and to a lesser extent culture, test results within a 24-month time window. These findings are consistent with our previous identification of positive correlations between StatPak, culture and IFNg tests. From previous analyses it was possible to conclude that animals that test positive on culture or StatPak are

also likely to test positive on IFNg and those animals that are culture-positive on multiple different samples are likely to have higher IFNg OD values. However, it should be noted that these previous correlations related to contemporaneous test results (i.e. test results obtained at the same time) in contrast to the current analyses where there was a time lapse between test results.

Previous research has shown that badgers exhibit a range of immunological responses following experimental challenge with *M. bovis*, and that these responses appear to occur in a specific order: levels of IFNg steadily increasing to a maximum at 9 months after infection, followed by decrease at 10-12 months and a rapid rise to a high level at 14 months, showing fluctuation at a relatively high level thereafter (Mahmood *et al*., 1987). The same study concluded that the level of antibodies and number of bacilli detected by culture showed a statistically significant increase by 17 months. Therefore, there was a delay of at least eight months between maximum titres of IFNg and antibodies (Mahmood *et al*., 1987). This offers a potential explanation for the observation in the present study that badgers with the highest IFNg OD values produced positive test results on StatPak and culture at a significantly higher rate than did badgers in the lowest IFNg category, when the follow-up time was greater than a year. It is less clear why, in the present study, badgers with lower IFNg OD values did not produce conclusive predictive results on StatPak and culture tests when the follow-up time exceeded a year. A possible explanation is that badgers with low IFNg OD values may have a generally weaker immune response, or perhaps a degree of immunological anergy (Chaparas, 1982), and hence express low IFNg and antibody responses.

The analyses described above have some inherent limitations. For example, the proportional hazards approach assumes that the effect of the predictor variable (i.e. IFNg OD value) was constant for the duration of the study. However, this is clearly unlikely to be the case, as a badger's IFNg titre is expected to vary over time and with the course of infection (Mahmood *et al*., 1987). Nevertheless, the assumption was not violated for a follow-up period of two years (as indicated by the formal assessment of survival lines and a Schoenfeld residuals test: see Methods section) and hence the analyses and conclusions are valid. Other limitations included the uneven distribution of observations amongst categories of INFg responses. In order to address the limited number of observations in the highest categories of IFNg (due to few badgers giving very high IFNg OD readings) we focussed on interpreting the trends in outputs rather than individual values. The two highest IFNg categories (i.e. the third and fourth) accounted for only approximately 4% of observations for both StatPak and culture, which is likely to have resulted in low statistical power for the parameters estimated. Moreover, the mycobacterial culture test has limited sensitivity in live badgers (perhaps as low as 10 per cent in some cases: Drewe *et al*., 2010) which means that the true predictive ability of IFNg may be higher than that described here. Only 3% of culture test results were positive in comparison to 21% of Stat-Pak tests, thus any relationship between IFNg and culture may be masked by inaccurate data and/or a low sample size. These limitations could potentially be addressed in future studies by improving the sensitivity of the culture test, possibly by using an extended sampling protocol involving more types of samples or more frequent sampling although this is unlikely to be practical. A more practical alternative would be to repeat the analysis in the future when more data become available.

In conclusion, our findings suggest that insights into the likely time to progression to testing positive on other tests (which are likely to represent infectiousness) may be gleaned by examining IFNg test results on a quantitative scale rather than simply categorising animals into infected or not infected. Knowledge of the likely amount of time that will be taken for an infected badger (with a quantified IFNg titre) to subsequently test positive on StatPak and culture (indicators of infectiousness) is useful from a research perspective as it improves our understanding of the way immunological processes affect disease progression and the dynamics of *M. bovis* transmission and may help us to better parameterise models of *M. bovis* infection in badgers. Both of these advances improve our evidence base on which to make decisions and evaluate possible infection control options. Further, this approach may help identify which badgers (when trapped and sampled) are most likely to become infectious soonest and hence might be at a highest risk of transmitting infection. Consequently, IFNg test results could potentially be used to inform targeted removal at the group level and if the test were available in a format that allowed rapid trap-side diagnosis then it could be used to selectively remove captured individuals.

Objective 3: Novel approaches to interpreting diagnostic test results

Introduction

As demonstrated above (Objective 1b) the combined use of more than one diagnostic test can increase the accuracy of TB diagnosis in live badgers, particularly when interpreted at the social group level. But an important limitation of the analyses presented here and previously (e.g. Drewe *et al.*, 2010) is that a diagnosis is frequently based on just one sampling event. Hence any previous test results for the same animal (and there may be many of these over several years in the WP dataset) have been effectively excluded from the diagnosis. This represents a loss of potentially useful information particularly given the imperfect performance of the tests. Animals that test positive at one point in time frequently test negative (using the same test) at the subsequent test event. In the analyses below we describe an approach for interpreting potentially disparate results from animals sampled multiple times over a protracted period. Hence these analyses address the question of determining the probability that a badger is truly infected given its current and historical pattern of results to existing diagnostic tests.

When using imperfect diagnostic tests, the true state of infection is not, or is only partially, determinable. Clearly this creates problems in classifying an animal's infection status confidently. We hypothesise that the incorporation of historical test results will improve the ability to correctly interpret diagnostic test results in order to infer the underlying current true infection status of a badger (or group of badgers). Such estimations are possible by adapting models that are applied to understand population dynamics (Choquet *et al.*, 2013). Here, we use a state-dependent capture-mark-recapture analysis accounting for complexities in ecological epidemiology, to estimate the probability that a badger is truly infected. We hypothesise that interpreting a badger's historic diagnostic test results alongside any current results will provide a deeper and more accurate insight into the true infection status of individuals. Such information could then be used to improve the reliability of decisions made to manage this disease in badger populations in situations where several test results are available from the same animals over a period of time.

During the long-term study at WP traps are routinely deployed at badger setts throughout the year and individual animals can therefore be caught more than once a year and several times during their lifetime. We used data from this study to assess for each individual badger the probability of each possible infection status sequence given its capture and testing history. As a result, we were able to produce estimates of the likelihood that any given badger was truly infected at the time of sampling given any combination of test results (StatPak, IFNg and culture) and accounting for its capture history and diagnostic test result history. This represents a clear improvement on current methods where diagnosis of infection is often based on imperfect diagnostic tests applied to samples collected at a single sample point.

Materials and methods

Study area and study population

The data analysed in this study were derived from badger captures at WP from July 2006 to October 2013. During this period badger trapping took place at all main setts in the study area four times each year, usually once in each of the four seasons (dates refer to classification of seasons based on exact dates of sampling across all eight years of the study): spring (5 May to 22 June), summer (5 July to 2 October), autumn (3 October to 1 December) and winter (6 December to 31 January). Trapping was suspended each year from February to April inclusive to avoid capture of new-born cubs and separating lactating females from dependent cubs. Each sample point consisted of a cluster of three trapping sessions (A, B and C), each of which covered approximately one third of the study area. Hence the combined sampling effort across these three areas represented one round of sampling across the entire study area. The duration of each such sweep across the study area (i.e. from the start of trapping at A, through B to the end of trapping at C) was less than the length of time between clusters. For the purposes of the analyses, the few trapping sessions that straddled two seasons were allocated to whichever season they fell into most. As a result, sampling did not take place in some seasons, for example in summer 2009 there was no trapping but there were two trapping sessions in

the autumn of that year (Appendix 1). In the rare instances where the same badger was captured and sampled twice in the same trapping cluster, only the test results from the first capture were used in the analysis.

Trapping and testing of badgers

Badgers were trapped, anaesthetised and sampled as reported previously (Delahay *et al.,* 2000a & 2013). At each capture event samples were collected for StatPak, IFNg and the microbiological culture of clinical samples. The dataset subjected to analyses and modelling in the present study included results from these three diagnostic tests at each of 2022 sample points (each representing one badger trapped and sampled: see Appendix 1) collected from 541 individual badgers (a median of 3 captures per badger, range: 1 to 18) across 28 capture sessions from July 2006 to October 2013.

Multistate capture-recapture model

Capture-mark-recapture modelling is frequently used for estimating population parameters in ecological studies (Lebreton *et al.*, 1992). This type of analysis allows the estimation of transition probabilities between ecological or epidemiological states, for example as animals become infected and make the transition from susceptible to infected state. Consequently, the true state of infection in any given animal was modelled as a dynamic variable. It was assumed that once infected an individual could not revert to being uninfected, and hence the state transition was unidirectional from negative to positive.

Results of each of the three diagnostic tests were coded into one of eight possible combinations of diagnostic test results (Table 14) and assigned to each sampling point (hereafter referred to as a time step). If the animal was not captured it was recorded as zero for that time step.

Table 14. The eight possible combinations of diagnostic test results and their respective coding numbers (1-8) for input into the capture-mark-recapture model. Numbers in brackets are the cross-classified observed frequencies of these three test result combinations when used for detection of *M. bovis* infection in live badgers, obtained from 2,022 sampling occasions involving 541 badgers captured at Woodchester Park from July 2006 to October 2013.

	IFNg –		**IFNg +**	
	Stat-Pak –	**Stat-Pak +**	**Stat-Pak –**	**Stat-Pak +**
Culture –	1 (1437)	2 (208)	3 (164)	4 (165)
Culture +	5 (3)	6 (6)	7 (3)	8 (36)

– = negative test result
+ = positive test result

The probabilities of each of five events occurring at the individual badger level were estimated within the capture-mark-recapture model by means of maximum likelihood estimation (Choquet *et al.*, 2009). These were, for any given badger, the probabilities of:

- Being captured (trapped and sampled, C);
- Testing positive on its first capture event (apparent prevalence, P);
- Surviving from one time step to the next (S);
- Becoming infected (transitioning from uninfected to infected, T);
- Observing each diagnostic test combination as presented in Table 14, given that the animal was truly either (a) infected or (b) uninfected.

Dependencies were modelled between the first four of these probabilities and external parameters (sex, season, and year). For example, the probability of a badger which is caught at time step *t* whilst in state *f* (uninfected) being re-caught at time *t+1* in state *i* (infected), given the observed capture at time *t* in state *f*, will depend on: the survival probability from *t* to *t+1* (dependent or independent of state *i*), the state transition probability (between *f* and *i*) and the capture probability (both dependent on survival probability).

Calculation of true infection status probabilities

Inferences about the underlying true infection status (infected or uninfected) of any given badger were made by estimating the probability of infection being present given any of the eight possible combinations of test results (Table 14), for all possible combinations of a badger's capture and testing history. These probability calculations were carried out using R software (R Development Core Team, 2008). The probability of being truly infected at the previous trapping occasion, $p(Status^{+}|History)$, was calculated according to the formula:

$$p(S^{+}|H) = \sum p\left(H|S_i{}^{+}\right) * p\left(S_i{}^{+}\right) / \sum p\left(H|S_i{}^{+}\right) * p\left(S_i{}^{+}\right) + p(H|S^{-}) * p(S^{-})$$

Where:

$\sum p\left(H|S_i{}^{+}\right)$ is the sum of probabilities of observing a given history for every combination of i lines in the status matrix (Figure 11);

$p\left(S_i{}^{+}\right)$ is the probability of a badger being 'status positive' (infected), estimated as the probability of being infected on the first capture multiplied by the transition probabilities for every time step;

$p(H|S^{-})$ is the probability of obtaining the observed history given the badger is 'status negative' (uninfected); and

$p(S^{-})$ is the probability of a badger being 'status negative' (uninfected), calculated as $1 - p\left(S_i{}^{+}\right)$.

To estimate the probability (for each combination of test results) of a badger being truly infected at the end of the study period, a matrix was generated containing the estimated probabilities of the true infection status of any given badger across all 28 observed time steps (roughly equivalent to the 28 seasons of the study: see Appendix 1) plus one simulated extra time step. This final simulated time step was necessary because some badgers were caught only once and hence did not have a history of previous captures or test results. To account for this, the probabilities of being truly infected given any possible test combination at the current (last) sampling session were estimated using an additional hypothetical 29th time step. This enabled the probability of infection in these single-capture badgers to be estimated. The matrix was therefore 29 time steps wide by 30 lines long, where each of the 29 lines represented the possible range of probable true infection histories when infection was truly acquired at some point and the final line corresponded to a history of continuous true negative infection status. This enabled us to examine, for any given badger, how the results of any previous diagnostic tests performed on the same animal, influenced the interpretation of the most recent (current) diagnostic test results.

As an example, let us consider an individual whose capture and testing history is '– + 0 +' (indicating a timespan containing four possible trapping episodes, where: '–' means the badger was trapped and tested negative; '+' means the badger was trapped and tested positive; and '0' indicates that the animal was not captured) which was recoded as '1 8 0 8' (where 1 refers to all three tests giving a negative result and 8 refers to all three tests giving a positive result: see Table 14). Using the parameters defined above, we can express the probability of the observed history being seen given each possible true status sequence. Thus the probability of the history '1 8 0 8' given that the true status sequence is '– – – –' (meaning the badger was actually uninfected all along) can be expressed as:

$$\begin{aligned} p(history|S^{-}) &= [(1 - \text{prevalence}) * \text{p(capture)} * \text{p}(T_1^{-}T_2^{-}T_3^{-}|S^{-})] \\ &* [(1 - \text{p(transition)}) * \text{p(capture)} * \text{p}(T_1^{+}T_2^{+}T_3^{+}|S^{-})] \\ &* [(1 - \text{p(transition)}) * (1 - \text{p(capture)})] \\ &* [(1 - \text{p(transition)}) * \text{p(capture)} * \text{p}(T_1^{+}T_2^{+}T_3^{+}|S^{-})] \end{aligned}$$

Where:

$p(history|S^{-})$ is the probability of obtaining the observed pattern of trapping and test results;

$\mathrm{prevalence}$ is the apparent prevalence at that time step (equivalent to the probability that a trapped animal tests positive);

$\mathrm{p(capture)}$ is the probability of capturing the badger;

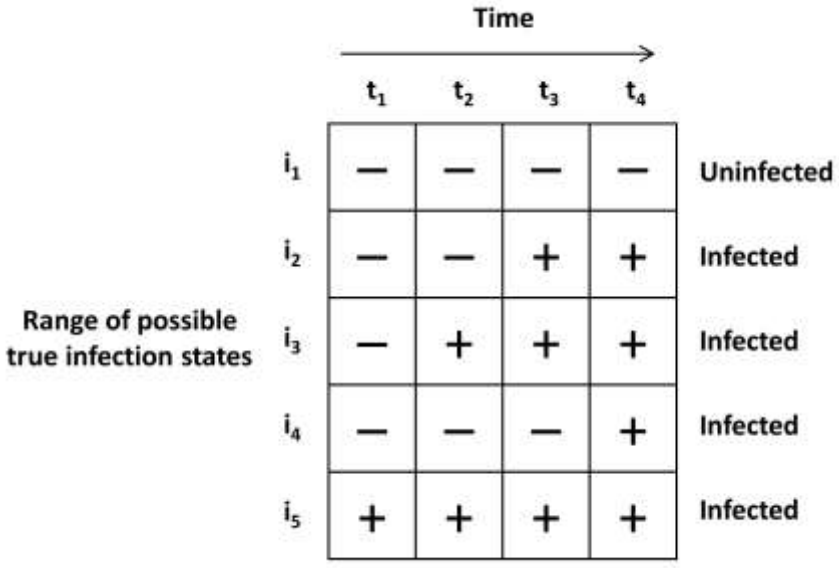
$\mathrm{p}(T_1^- T_2^- T_3^- | S^-)$ is the probability that all three diagnostic tests (StatPak, IFNg and culture) would produce a negative result given that the animal is truly uninfected (status negative);

$\mathrm{transition}$ is the probability of an animal becoming infected between two sampling sessions; and

$\mathrm{p}(T_1^+ T_2^+ T_3^+ | S^-)$ is the probability that all three diagnostic tests (StatPak, IFNg and culture) would produce a positive result given that the animal is truly uninfected.

The probabilities of the capture and testing histories, $p(History|Status^+)$, for each of the other possible true infection state sequences can be similarly determined. The sum of the probabilities that correspond to all the possible true positive infection states represents the overall probability that a badger is truly infected (Figure 11).

Figure 11. An Example matrix showing the range of five possible true infection states that a single badger observed for four time steps could have followed. The first line (i_1) represents the status of being uninfected for the entire period of four time steps; the second line (i_2) represents infection occurring during the third time step; and the fifth line (i_5) represents the status of being infected for the whole period under observation. A probability was assigned to each line of the matrix (see main text for details). The sum of the probabilities that correspond to true positive infection status (lines 2, 3, 4 and 5) represent the overall probability that the badger is truly infected. For the purposes of this example, only four time steps and five possible infection states are shown, although the full model contained 29 time steps and 30 possible infection states for each badger.



In order to be able to interpret a badger's current diagnostic test results in the context of its trapping and test result history, we assessed the probabilities of a badger being truly infected given any possible combination of captures and test results across all time steps (seasons) through the study period, for each possible test result combination at the last trapping occasion. This enabled us to generate estimates of the probabilities of a badger being truly infected given any possible diagnostic test result combination at its most recent sampling occasion. Ultimately, we estimated the probability of being truly infected at the end of the observational period for every badger in the study, given every possible combination of test results that a badger might exhibit at the last sampling point.

Dependencies between capture-mark-recapture model parameter estimates

We treated the estimated parameters (probability of capture, apparent prevalence, probability of transition from being uninfected to infected, and probability of survival from one time step to the next)

as homogeneous across the study population, varying only by these specific factors: season, year, sex of badger, and infection status. We tested the effect of these covariates on the above probabilities, as well as their interdependencies.

The potential impact of variation in the timing of the first trapping occasion on apparent prevalence in different seasons across all eight years of the study was investigated for male and female badgers. Also, infection status at the time of the previous capture event and the effect of badger sex on the probability of survival were examined because they had previously been found to influence survival in badgers (Graham *et al.*, 2013). A dependency between the probability of becoming infected (transition in state from uninfected to infected) and season for male and female badgers across all eight years was included to reflect seasonal variations in behaviour which may differ between sexes (Graham *et al.*, 2013). The probability that a badger would change its status from uninfected to infected in the model was not considered to depend on it's true infection status, because we assumed that once infected it would not recover from infection, and hence no reversion to the uninfected state was possible (see Delahay *et al*., 2000a; Vicente *et al.*, 2007 for justification). Seasonal and annual variations were included in the estimation of capture probabilities for each sex, as these factors are known to vary within this badger population (Graham *et al.*, 2013, Delahay, 2010).

All plausible combinations of parameters were examined and a consensus model was selected by removing one parameter at a time. In this way, the best fitting model was selected by assessing the Akaike Information Criterion adjusted for over-dispersion (QAICc; see Table 15). This consensus model was then used to perform neighbourhood investigation by removing one parameter at a time and assessing its effect on model fit (Grosbois & Tavecchia, 2003). Therefore, the compared models were nested because they were each a simpler version of the initial full model.

Table 15. The process used to obtain the consensus capture-mark-recapture model. Models were systematically selected by removing one variable at a time from the full model, which included dependencies between the four probabilities of: infection at first capture (P); survival from one time period to the next (S); transition (from an uninfected to infected state, T); and capture (C). Models incorporated plausible combinations of: calendar year (year), season (seas), sex (g), infection status ('from.to' indicating the one-way transition from uninfected to infected), and probabilities of obtaining each diagnostic test result (F). The consensus model had the lowest QAICc (Akaike Information Criterion adjusted for over-dispersion) for all of the four parameters that were specified only by significant dependencies from other variables.

Model ID	Parameters included in the model	QAICc
Full model	$P_{g+seas+year}, S_{from+g+seas+year}, T_{g+seas+year}, C_{from+g+seas+year}, F_{from.to}$	9269.78
M1	$P_{g+year}, S_{from+g+seas+year}, T_{g+seas+year}, C_{from+g+seas+year}, F_{from.to}$	9264.12
M2	$P_{g+seas}, S_{from+g+seas+year}, T_{g+seas+year}, C_{from+g+seas+year}, F_{from.to}$	9272.68
M3	$P_{seas+year}, S_{from+g+seas+year}, T_{g+seas+year}, C_{from+g+seas+year}, F_{from.to}$	9268.13
M4	$P_{g+seas+year}, S_{from+g+year}, T_{g+seas+year}, C_{from+g+seas+year}, F_{from.to}$	9283.21
M5	$P_{g+seas+year}, S_{from+g+seas}, T_{g+seas+year}, C_{from+g+seas+year}, F_{from.to}$	9279.65
M6	$P_{g+seas+year}, S_{from+seas+year}, T_{g+seas+year}, C_{from+g+seas+year}, F_{from.to}$	9275.96
M7	$P_{g+seas+year}, S_{g+seas+year}, T_{g+seas+year}, C_{from+g+seas+year}, F_{from.to}$	9268.21
M8	$P_{g+seas+year}, S_{from+g+seas+year}, T_{g+year}, C_{from+g+seas+year}, F_{from.to}$	9308.78
M9	$P_{g+seas+year}, S_{from+g+seas+year}, T_{g+seas}, C_{from+g+seas+year}, F_{from.to}$	9276.74
M10	$P_{g+seas+year}, S_{from+g+seas+year}, T_{seas+year}, C_{from+g+seas+year}, F_{from.to}$	9268.26
M11	$P_{g+seas+year}, S_{from+g+seas+year}, T_{g+seas+year}, C_{from+g+year}, F_{from.to}$	9360.63
M12	$P_{g+seas+year}, S_{from+g+seas+year}, T_{g+seas+year}, C_{from+g+seas}, F_{from.to}$	9267.24
M13	$P_{g+seas+year}, S_{from+g+seas+year}, T_{g+seas+year}, C_{from+seas+year}, F_{from.to}$	9343.45
M14	$P_{g+seas+year}, S_{from+g+seas+year}, T_{g+seas+year}, C_{g+seas+year}, F_{from.to}$	9351.15
Consensus model	$\boldsymbol{P_{year}, S_{g+seas+year}, T_{seas+year}, C_{from+g+seas}, F_{from.to}}$	**9256.85**

Results

Capture probabilities and diagnostic test results

Badgers of both sexes were most likely to be caught in summer (July to September) and least likely to be caught in winter (December to January), although males were twice as likely as females to be caught in winter (Table 16). It should be noted that this sex difference in winter trapping rates is likely to be artificial due to the policy of releasing any captured females that are suspected to be lactating at this time of year. The majority of badgers (69% of males and 73% of females) tested negative on all three diagnostic tests (Table 17). There was high agreement between StatPak and IFNg test results, and although very few badgers were culture positive those that were tended to also be positive on StatPak and IFNg (Table 17).

Table 16. The frequency of badger capture by season for 541 badgers (247 male and 294 female) from July 2006 to October 2013 at Woodchester Park. Badgers were caught multiple times (median: 3 times per badger, range: 1 to 18 times) giving a total of 2,022 sampling events during the study period.

	Number (%) of captures		
Season	**Males**	**Females**	**All**
Winter (6 Dec - 31 Jan)	151 (16)	86 (8)	237 (12)
Spring (5 May - 22 June)	259 (27)	358 (34)	617 (31)
Summer (5 July - 2 Oct)	327 (34)	385 (37)	712 (35)
Autumn (3 Oct - 1 Dec)	232 (24)	224 (21)	456 (23)
Total	969 (100)	1053 (100)	2022 (100)

Table 17. The distribution of diagnostic test result combinations for captured badgers relative to their sex, for 2,022 sampling events involving 541 individuals (247 male and 294 female) from July 2006 to October 2013 at Woodchester Park. Numbers indicate the percentage of male [female] badgers giving each of the diagnostic test combinations.

	IFNg –		**IFNg +**	
	Stat-Pak –	**Stat-Pak +**	**Stat-Pak –**	**Stat-Pak +**
Culture –	69 [73]	12 [8]	10 [6]	7 [10]
Culture +	0.1 [0.2]	0.4 [0.2]	0.2 [0.1]	1 [3]

– = negative test result
+ = positive test result

Prevalence of infection

The consensus capture-mark-recapture model identified substantial annual variation in the probability that a badger would be infected on its first capture occasion (Figure 12). This measure is equivalent to the apparent prevalence of infection within the captured population of badgers. Male badgers were generally more likely to be trapped than females, and the probability of trapping uninfected badgers was twice that of infected badgers in every season for both sexes (Figure 13). Therefore, the true prevalence was higher than the apparent prevalence (Figure 12). True prevalence was calculated by adjusting apparent prevalence by the capture probabilities which varied by season, infection status and the sex of the badger. Hence, true prevalence estimates included annual, seasonal and sex-related variation (Figure 12). Using this method, the true prevalence of *M. bovis* infection within the WP badger population was estimated to be approximately 50% (Figure 12).

Figure 12. Trends in the apparent and true prevalence of *M. bovis* infection in the badger population at Woodchester Park throughout the study period. True prevalence in male badgers (black line) and female badgers (grey line) represent apparent prevalence corrected for imperfect performance of the diagnostic tests and seasonal variations in capture probabilities for males and females. True prevalence varied relative to sex and season. Apparent prevalence (dashed line) varied by year but not by season (1 = spring, 2 = summer, 3 = autumn, 4 = winter) (see Table 16 for details of season dates).

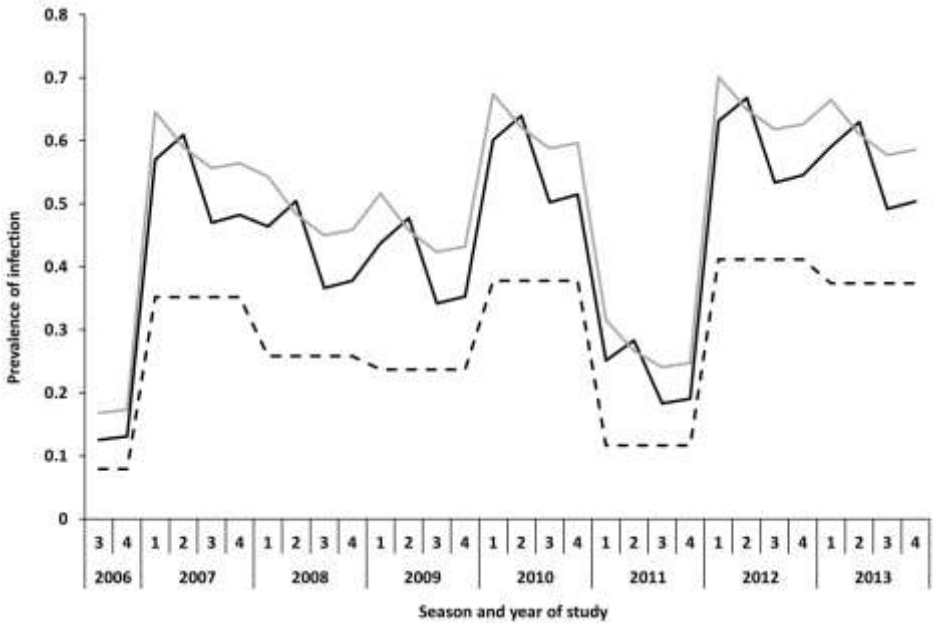
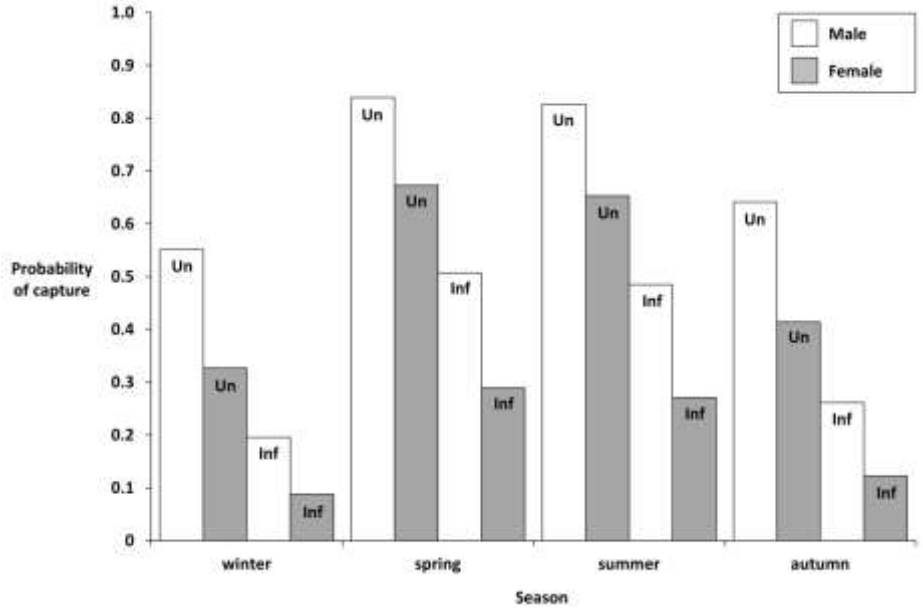


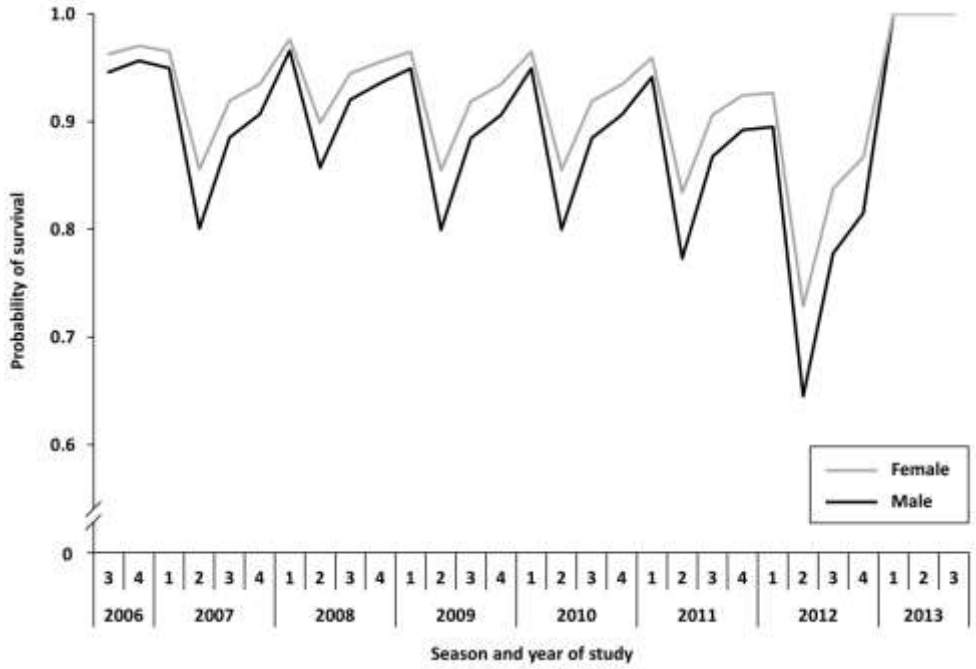
Figure 13. The probability of trapping badgers varied by season (see Table 16 for dates), infection status (Un = uninfected; Inf = infected) and sex. Males were more likely to be caught than females in all seasons, although values for winter would have been artificially skewed by the policy of releasing lactating females without sampling, and uninfected badgers of both sexes were more likely to be caught than infected badgers. There was no annual variation in these estimates which were generated from a multistate capture-mark-recapture model of empirical data collected from the badger population at Woodchester Park from July 2006 to October 2013.



Survival and transition probabilities

The probability that a badger survived from one time step to the next varied by season, year and sex, with females surviving slightly longer than males throughout the study (Figure 14). The best-fitting model did not include any dependency between a badger's *M. bovis* infection status and its probability of survival, and so survival was estimated homogeneously for infected and uninfected badgers. The probability that a badger underwent a transition from uninfected to infected status varied by season and year. The highest probability of this transition was observed in the spring of all years.

Figure 14. Survival probabilities for badgers at Woodchester Park from July 2006 to October 2013. The probability that individual badgers survived from one season to the next varied by sex, season and year (1 = spring, 2 = summer, 3 = autumn, 4 = winter; see Table 16 for details of season dates).



Probabilities of obtaining each diagnostic test result combination

The probabilities of obtaining each diagnostic test result combination given that a badger is truly (a) uninfected or (b) infected are presented in Table 18. By combining these probability estimates that correspond to a positive or negative infection status, it was possible to calculate individual test sensitivities and specificities respectively. The probability of each diagnostic test combination given a true positive or a true negative infection status was calculated from individual test sensitivities and specificities using the general formula of conditional probability for independent events (Bayes & Price, 1763):

$$p(A \cap B \cap C) = p(A) * p(B) * p(C)$$

For example, if we tested a badger using StatPak, IFNg and culture and all three tests gave a positive result (test result combination number 8: Table 14), the probability of a badger being truly infected can be calculated as follows (this is the sensitivity of the three tests combined, used together and interpreted in parallel):

$$p(T^{+}_{culture}\, T^{+}_{IFN}\, T^{+}_{Stat-Pak} | Status^{+}) = SeT_{culture} * SeT_{IFN} * SeT_{Stat-Pak}$$

Where:

T^+ represents a positive diagnostic test result;

$Status^+$ indicates that the badger is truly infected;

Se is the individual diagnostic test sensitivity; and

T represents one of the three tests used to diagnose *M. bovis* infection.

In a similar manner, it is possible to calculate the probability of every diagnostic test result combination presented in Table 14. In this way we estimated the sensitivity and specificity of each diagnostic test and obtained comparable results to those estimated using a Bayesian approach (Objective 2b above) and therefore concluded that these estimates of test performance were robust.

Table 18. The probabilities of observing each of the eight possible combinations of diagnostic test results given that a badger is either (a) truly uninfected, or (b) truly infected with *M. bovis*. For example, an uninfected badger has a 94.3% probability of testing negative on all three tests, whereas an infected badger has only a 5.1% chance of testing positive on all three tests. The sensitivity (or specificity) of each diagnostic test can be estimated by summing the probabilities that correspond to positive (or negative) test results given that a badger is infected (or uninfected). Hence, the sensitivity of culture to detect *M. bovis* infection in live badgers would be estimated by summing the four probabilities in Table (b) below that correspond to a positive culture status: 0.004 + 0.009 + 0.002 + 0.051= 0.066 (or about 7%). Probabilities were estimated using a capture-mark-recapture model in E-surge with empirical data collected from 541 badgers sampled a total of 2,022 times at Woodchester Park from July 2006 to October 2013.

(a) Uninfected badgers	**IFNg –**		**IFNg +**	
	Stat-Pak –	**Stat-Pak +**	**Stat-Pak –**	**Stat-Pak +**
Culture –	0.943	0.019	0.034	0.003
Culture +	0.000	0.000	0.001	0.000

(b) Infected badgers	**IFNg –**		**IFNg +**	
	Stat-Pak –	**Stat-Pak +**	**Stat-Pak –**	**Stat-Pak +**
Culture –	0.275	0.260	0.170	0.229
Culture +	0.004	0.009	0.002	0.051

– = negative test result
+ = positive test result

Probability that a badger is truly infected given any combination of diagnostic test results

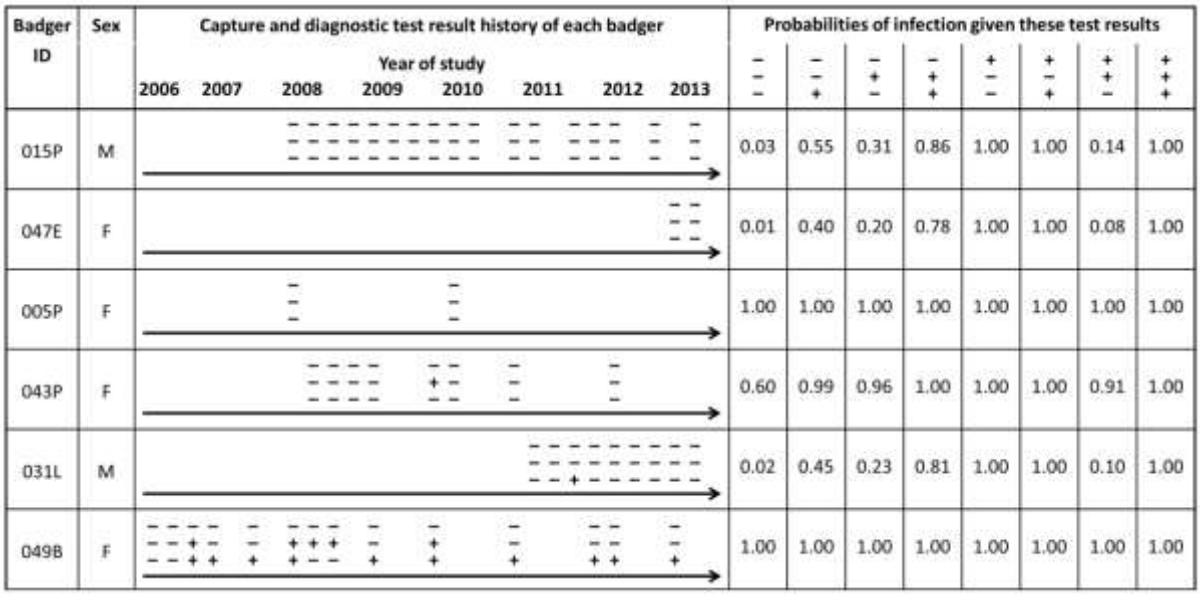
Of course, we do not usually know a badger's true infection status: rather, we have a set of diagnostic test results and from these we wish to infer true infection status. We can do this by rearranging the test result probabilities (Table 18) in the manner presented earlier (*Calculation of true infection status probabilities*) to produce estimates of the probability of a badger being truly infected given the individual's sex and capture and testing history. The probability that a badger was infected at the end of the observation period (time step 29) was found to be dependent on the observed diagnostic test result combinations from previous captures, the duration of the follow-up period, the frequency of recapture and the diagnostic test results on the last capture occasion. All of these factors were incorporated into the model to produce a separate estimate of the probability of infection for a badger given each of the eight possible diagnostic test result combinations (ranging from all three tests being negative through to all three tests being positive: Table 14).

Figure 15 shows the capture and diagnostic test result histories for six badgers, and illustrates the effect of sex, capture history and current and historical diagnostic test results on the probability of a being truly infected given any diagnostic test result. Badgers that were caught many times over several

years and consistently tested negative (e.g. badger 015P, Figure 15) were unlikely to be truly infected. Lower probabilities of infection arose only in badgers that were trapped for the first time late in the study. For example badger 047E (Figure 15) was trapped only twice, but both captures occurred near the end of the study and produced negative results on all three diagnostic tests. The lower probabilities of infection in badger 047E than in badger 015P are due not to the diagnostic test results (both badgers always tested negative on all three tests), but because of the lower amount of time that the former was likely to have been exposed to infection. The perhaps unexpectedly low probability of infection (8%) given negative StatPak but positive culture and IFNg results in this animal was probably due to the rarity of this result (only 0.1% of badgers sampled gave this diagnostic test result combination: Table 14).

A badger that tests negative may of course be infected, given the limited sensitivity of the diagnostic tests. An example of this is badger 005P (see Figure 5), which showed the maximum probability of being truly infected, despite producing negative results to all diagnostic tests on both occasions when it was trapped. The likely explanation for this observation is the high prevalence of infection during the long period when the badger was not trapped which increased the likelihood of infection during a time when it was not trapped, or on the occasions when it was trapped it incorrectly tested negative. This illustrates the value of the approach presented in the present study. The current interpretation of a badger that tests negative and has never tested positive is that it is uninfected, but these analyses show the limitation of any interpretation which is based solely on apparent test results.

Figure 15. The probability of a badger being truly infected with *M. bovis* varied with its sex, capture history and current and historical diagnostic test results. This diagram shows a range of examples from six badgers, each with a different trapping and diagnostic test result history. From July 2006 to October 2013 there were 28 time steps when trapping and sampling of badgers occurred. A set of three diagnostic test results appears on a badger's timeline if it was captured and tested at that time step, with the uppermost test result in each triad representing culture, the middle test result is for the IFNg test and the lowermost test result for the StatPak. Hence badger 015P was caught and sampled 17 times during the study and on every one of these occasions it tested negative on all three tests. If this badger had been sampled again immediately after the end of the study in October 2013 and on that occasion had tested negative on all three tests, the probability of it being truly infected would have been 0.03 (indicating that despite consistently testing negative over a long time period, the insensitivity of the testing regime and the background prevalence of infection in the population meant this badger nevertheless had a 3% chance of being infected). However, if this same badger had tested positive on StatPak (but negative on the other two tests) at the end of the study, the probability that it was truly infected would have increased to 55%.



M = male, F = female, – = negative test result, + = positive test result

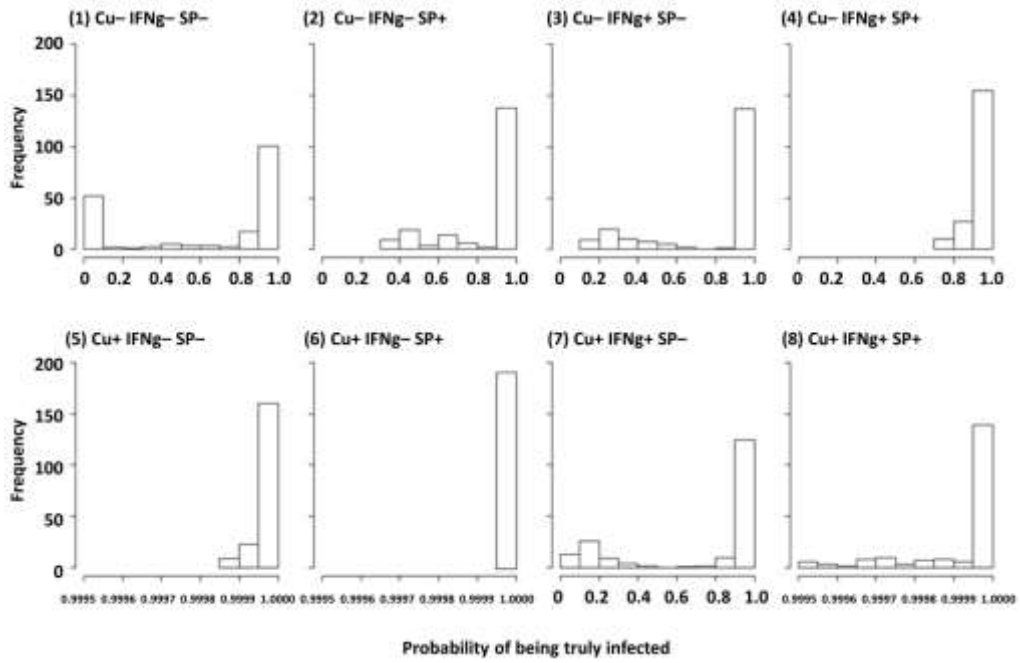
The potential impact of a single positive diagnostic test result is illustrated by badger 043P (see Figure 15), which was trapped and sampled eight times during the study, and tested positive on IFNg in early

2010. Despite testing negative on all tests on all seven other sampling occasions, the probability that this badger was infected at the end of the study was a minimum of 60% (and this was if all diagnostic tests were negative at the last sampling occasion). Had this same badger produced another positive IFNg result at the end of the study, the probability that it was truly infected would have risen to 96%. Thus a single positive test result may be sufficient to 'ruin a badger's reputation' by vastly increasing the chances that it was truly infected. However, a single positive diagnostic test result may of course be a false positive. This is illustrated by badger 031L (see Figure 15), where a single positive result is followed by a long run of negative test results. These repeated negative test results appear to nullify the single positive result with the consequence that this badger only has a 2% chance of infection. Thus this badger's results further illustrate the value of the analytical approach presented here over the method of interpretation presently used, which considers that once a badger tests positive it is infected. The present approach on the other hand allows and accounts for the possibility of false positive diagnostic test results.

Badgers with multiple positive diagnostic test results were, unsurprisingly, highly likely to be truly infected. As an example, badger 049B (see Figure 15) was trapped 14 times across the entire duration of the study (7 years) and although it never tested culture positive it was frequently StatPak positive and occasionally IFNg positive. This run of repeatedly testing positive explains why the maximum probability of infection was estimated given any combination of diagnostic test results.

The probabilities of being truly infected given any combination of diagnostic test results are expressed as distributions at the population level in Figure 16. Every test combination included badgers that would have been considered truly infected, even badgers testing negative on all three tests, because the model incorporated sex, capture and testing history, the duration of the follow-up period, the frequency of recapture, and the diagnostic test results on the last capture occasion. This again illustrates the added value of this approach over the current system of interpreting test results for diagnosing *M. bovis* infection in live badgers.

Figure 16. Population-level distributions showing the probabilities of badgers being truly infected with *M. bovis* given any one of the eight possible diagnostic test combinations (see Table 14) being observed on the 29th (simulated) trapping occasion. Probabilities were estimated using a capture-mark-recapture model in E-surge with empirical data collected from 541 badgers sampled a total of 2,022 times at Woodchester Park from July 2006 to October 2013. Only the 192 badgers that had a greater than 50% probability of being alive at the end of the study are included in the probability distributions shown. Numbers in brackets correspond to the diagnostic test combinations given in Table 14.



Cu = culture test, IFNg = Interferon gamma assay, SP = Stat-Pak test, – = negative test result, + = positive test result

Discussion

In these analyses we related longitudinal results of imperfect diagnostic tests to a true infection status. We modelled non-exhaustive capture-recapture data to estimate the dynamics of uncertain biological and behavioural processes (Williams *et al.*, 2002) which were used to estimate probabilities of the true underlying infection status of live badgers. The findings of our modelling support the hypothesis that it is possible to gain a much deeper insight into the true infection status of badgers by interpreting longitudinal diagnostic test results and accounting for uncertain biological and behavioural processes. These processes included the population level of infection, the probability of badgers becoming infected, the probability of badgers surviving from one trapping session to the next, and the probability of badgers being captured, with seasonal as well as annual variations in these parameters. Hence our method of interpreting diagnostic test results is becoming multidimensional. In doing so, the use of state-dependent capture-mark-recapture modelling is enabling us to uncover some of the underlying biological complexities which should not be ignored when attempting to evaluate epidemiological parameters of diseases in populations (Delahay *et al*., 2013).

There is a high level of uncertainty associated with conventional methods of interpreting diagnostic test results in live badgers, due to the tests' inherent insensitivity and imperfect specificity. Current methods of interpretation often assume that a badger is infected from the point at which it produces its first positive diagnostic test result. This method is likely to miss several cases of infection due to the insensitivity of the tests. Accuracy of diagnosis can be improved by interpreting several tests together (Drewe *et al.*, 2010) but this method loses any information gained from previous testing of the same animal. The probability of a badger being infected depends upon, among other things, the duration of follow-up, the availability of previous test results and the probability that the tests used will correctly give a positive or negative diagnosis. It therefore makes sense to incorporate as much of this information as we can into our decision making process. By including this historic diagnostic information in our contemporary interpretation of an animal's infection status, and accounting for variation in factors such as the trappability of individual animals and the prevalence of infection at the population level, we can improve the reliability of our diagnostic interpretation.

Moreover, our model gives us a different interpretation for each possible combination of diagnostic test results and hence our diagnosis can be fine-tuned to improve accuracy. In doing so we have moved diagnosis away from the traditional binary classification of an animal's infection status (either infected or not infected) towards a probability-based interpretation where the probabilities are updated each time a badger is re-captured. It may take a little time to work out what it means to say that a badger has a '90% chance of being infected' but this interpretation is inherently more useful than a simple diagnosis of 'infected' because it acknowledges and quantifies the small chance that the badger may in fact be uninfected. Similarly, whereas a badger that has tested negative on all tests would currently be considered uninfected, our model suggests that due to the population-level pressure of infection (prevalence), there is a reasonable chance that some of these badgers will nevertheless be truly infected. Hence our model allows us to accommodate rather than ignore the possibility of false positive and false negative test results.

Our findings suggest that the probability of a badger being truly infected with *M. bovis* depends more on the number, timing and results generated from previous diagnostic tests conducted on the same animal, and less so on the last set of observed test results. This may seem intuitive given the inaccuracy (particularly the insensitivity) of tests for *M. bovis* in live animals, yet current methods of interpretation do not utilise all this historic information. It could be suggested that in the absence of an accurate gold standard test for *M. bovis,* conclusions on the true infection status of badgers should account for the capture and testing history of the animal to minimise the chance of misclassification. This approach is clearly highly relevant to long term studies, but including prior results in the interpretation of infection status will also be of value in shorter studies where animals are for example sampled only twice in close succession. A key advantage of this method over conventional interpretation of test results is the reduced chance of incorrectly accepting a false positive or false negative result.

Uninfected badgers were more likely to be trapped than infected badgers, and male badgers were more likely to be caught than were females in all seasons, although values for winter would have been

artificially skewed by the policy of releasing lactating females without sampling. We are currently running a sensitivity analysis to determine the likely impact of these released females on the model output. Whilst minor changes to the numerical results may occur as a result of this, the conclusions are expected to remain largely the same as presented here. Accounting for these variations provided additional information for interpreting observed test results. Badgers that were captured more frequently exhibited a lower chance of being infected than those that were caught fewer times. This can be explained by the 'prove you are healthy principle' whereby untrapped badgers are not only more likely to be infected but also the longer a badger remains at large in the population the higher the chance it will be exposed to infection due to the high background prevalence.

Our model output corroborated previous findings that capture probabilities tend to vary by season (Graham *et al*., 2013, Delahay, 2010). This is likely to be due to variations in badger behaviour throughout the year due to fluctuations in weather conditions and hence food availability (Kruuk & Parish, 1987, Tolhurst *et al.*, 2009), and the timing of cubs becoming independent in the summer (Page *et al.*, 1994). Trapping effort at WP is constant from one year to the next (Delahay, 2010) and therefore human factors were assumed not to have influenced these estimates. Our model did not include a dependency between survival and infection status and hence differed in this respect from the model developed by Graham *et al*. (2013). However, it was possible to confirm that there was differential survival for male and female badgers in the study which might be partially explained by behavioural differences between the sexes, for example the increase in roaming behaviour of males in the spring (Delahay *et al.*, 2006).

The true prevalence of *M. bovis* infection within the WP badger population was estimated to be approximately 50% throughout this study, which is higher than previous estimates for the year 2005 of approximately 30% (Delahay *et al*., 2013). It should be noted however that annual prevalence estimates were observed to rise from 2000 onwards in the previous study (Delahay *et al*., 2013) and hence the present study's estimate may not be dissimilar. Also, previous studies have acknowledged that the limitations in the performance of diagnostic tests for *M. bovis* infection in live badgers will have resulted in underestimates of the true prevalence, whilst acknowledging that only an index of infection may be necessary in order to reveal spatio-temporal variations in infection (e.g. Delahay *et al*., 2013). In addition, previous prevalence estimates referred only to the sampled population whereas estimates in the present study relate to the entire study population (because the present method also accounted for the probability of capture). The implications are that methods for estimating prevalence that do not take into account diagnostic test performance limitations and unequal capture probabilities are likely to underestimate the true infection prevalence, quite possibly by a significant amount.

In conclusion, we have developed a multidimensional approach to the interpretation of diagnostic test results for *M. bovis* infection in live badgers, based on probability theory. This method enables us to uncover some of the underlying biological complexities which influence several epidemiological parameters, as well as our measurement of these parameters, of this infection in badger populations.

References

Allison, P. D. (1984). *Event History Analysis: Regression for Longitudinal Event Data*, Sage.

Bayes, M. & Price, M. (1763). An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr Bayes, FRS, communicated by Mr Price, in a letter to John Canton, AMFRS. *Philosophical Transactions,* **53,** 370-418.

Branscum, A. J., Gardner, I. A. &Johnson, W. O. (2005). Estimation of diagnostic-test sensitivity and specificity through Bayesian modelling. *Preventive Veterinary Medicine* 68, 145-163.

Chambers, M. A., Crawshaw T., Waterhouse, S., Delahay, R., Hewinson, R. G. & Lyashchenko, K. P. (2008). Validation of the BrockTB Stat-Pak assay for detection of tuberculosis in Eurasian badgers (*Meles meles*) and influence of disease severity on diagnostic accuracy. *Journal of Clinical Microbiology* **46**, 1498-1500.

Chambers, M. A., Lyashchenko, K. P., Greenwald, R., Esfandiari, J., James, E., Barker, l., Jones, J., Watkins, G. & Rolfe, S. (2010). Evaluation of a rapid serological test for the determination of *Mycobacterium bovis* infection in badgers (*Meles meles*) found dead. *Clinical and Vaccine Immunology*, 17, 408-411.

Chambers, M. A., Rogers, F., Delahay, R. J., Lesellier, S., Ashford, R., Dalley, D., Gowtage, S., Davé, D., Palmer, S., Brewer, J., Crawshaw, T., Clifton-Hadley, R., Carter, S., Cheeseman, C., Hanks, C., Murray, A., Palphramand, K., Pietravalle, S., Smith, G. C., Tomlinson, A., Walker, N. J., Wilson, G. J., Corner, L. A. L., Rushton, S. P., Shirley, M. D. F., Gettinby, G., McDonald, R. A. & Hewinson, R. G. (2010). Bacillus Calmette-Guérin vaccination reduces the severity and progression of tuberculosis in badgers. *Proc. R. Soc. B.* **278**, 1913-1920.

Chambers, M. A., Waterhouse, S., Lyashchenko, K., Delahay, R., Sayers, R. & Hewinson, R. G. (2009). Performance of TB immunodiagnostic tests in Eurasian badgers (*Meles meles*) of different ages and the influence of duration of infection on serological sensitivity. *BMC Vet Res*, 5, 42.

Chaparas, S. D. (1982). The immunology of mycobacterial infections. *Crit Rev Microbiol*, 9, 139-97.

Cheeseman, C. L. & Harris, S. (1982). Methods of marking badgers (*Meles meles*). *Journal of Zoology*, 197, 289-292.

Choquet, R., Carrie, T. Chambert & Boulinieri, T. (2013). Estimating transitions between states using measurements with imperfect detection: application to serological data. *Ecology,* **94,** 2160-2165.

Choquet, R., L. Rouan & Pradel, R. (2009). Program E-SURGE: a software application for fitting multievent models. *Environmental and Ecological Statistics* **3,** 845-865.

Clifton-Hadley, R. S., Wilesmith, J. W. & Stuart, F. A. (1993). *Mycobacterium bovis* in the European badger (*Meles meles*): Epidemiological findings in tuberculous badgers from a naturally infected population. *Epidemiology and Infection*, 111 (1), 9-19.

Corner, L. A. L., Buddle, B. M., Pfeiffer, D. U. & Morris, R.S. (2001). Aerosol vaccination of the brushtail possum (*Trichosurus vulpecula*) with the bacille Calmette-Guérin: the duration of protection. *Veterinary Microbiology* **81**, 181-191.

Corner L. A. L., Costello, E., Lesellier, S., O'Meara, D., Sleeman, D. P. & Gormley, E. (2007). Experimental tuberculosis in the European badger (*Meles meles*) after endobronchial inoculation of *Mycobacterium bovis*: I. Pathology and bacteriology. *Research in Veterinary Science* **83**, 53-62.

Cousins, D. V. & Florisson, N. (2005). A review of tests available for use in the diagnosis of tuberculosis in non-bovine species. *Revue Scientifique et Technique* **24**, 1039-1059.

Cox, D. R. & Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society. Series b (methodological)*, 248-275.

Crawshaw, T. R., Griffiths, I. B., Clifton-Hadley, R. S. (2008). Comparison of a standard and a detailed post-mortem protocol for detecting *Mycobacterium bovis* in badgers. *Veterinary Record* 163: 473–477.

Dalley, D., Dave, D., Lesellier, S., Palmer, S., Crawshaw, T., Hewinson, R.G. & Chambers, M. (2008) Development and evaluation of a gamma-interferon assay for tuberculosis in badgers (*Meles meles*). *Tuberculosis (Edinb)* **88**, 235-243.

de Leeuw, A. N., Forrester, G., Spyvee, P. D., Brash, M. & Delahay, R. J. (2004). Experimental comparison of ketamine with a combination of ketamine, butorphanol and medetomidine for general anaesthesia of the Eurasian badger (*Meles meles* L.). *Veterinary Journal* 167, 186-193.

Delahay, R. J. (2010). *Annual report to Defra on research conducted at Woodchester Park: Project SE3265*.

Delahay, R. J., Langton, S., Smith, G. C., Clifton-Hadley, R. S. & Cheeseman, C. L. (2000a). The spatio-temporal distribution of *Mycobacterium bovis* (Bovine Tuberculosis) infection in a high density badger (*Meles meles*) population. *Journal of Animal Ecology* 69, 1-15.

Delahay, R. J., Brown, J. A., Mallinson, P. J., Spyvee, P. D., Handoll, D., Rogers, L. M. & Cheeseman, C. L. (2000b) The use of marked bait in studies of the territorial organisation of the European badger (*Meles meles*). *Mammal Review* 30(2), 73-87.

Delahay, R. J., Walker, N., Smith, G. S., Wilkinson, D., Clifton-Hadley, R. S., Cheeseman, C. L., Tomlinson, A. J. & Chambers, M. A. (2013). Long-term temporal trends and estimated transmission rates for *Mycobacterium bovis* infection in an undisturbed high density badger (*Meles meles*) population. *Epidemiology & Infection* 141, 1445–1456. doi:10.1017/S0950268813000721.

Delahay, R., N. Walker, G. Forrester, B. Harmsen, P. Riordan, D. Macdonald, C. Newman & Cheeseman, C. (2006). Demographic correlates of bite wounding in Eurasian badgers,Meles meles L., in stable and perturbed populations. *Animal Behaviour,* **71,** 1047-1055.

Dohoo, I., Martin, W. & Stryhn, H. (2009). Screening and diagnostic tests. In: *Veterinary Epidemiologic Research*, 2nd edition. VER Inc, Charlottetown, Canada. Page 111.

Donnelly, C. A., Woodroffe, R., Cox, D. R., Bourne, 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.

Drewe, J. A., Tomlinson, A. J., Walker, N. J. & Delahay, R. J. (2010). Diagnostic accuracy and optimal use of three tests for tuberculosis in live badgers. *PLoS One*, **5**, e11196.

Gavier-Widen, D., Chambers, M. A., Palmer, N., Newell, D. G. & Hewinson, R. G. (2001). Pathology of natural *mycobacterium bovis* infection in European badgers (*Meles meles*) and its relationship with bacterial excretion. *Vet Rec,* 148**,** 299-304.

Graham, J., Smith, G. C., Delahay, R. J., Bailey, T., McDonald, R. A. & Hodgson, D. (2013). Multi-state modelling reveals sex-dependent transmission, progression and severity of tuberculosis in wild badgers. *Epidemiology & Infection,* **141,** 1429-1436.

Grambsch, P. M. & Therneau, T. M. (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika,* 81**,** 515-526.

Grosbois, V. & Tavecchia, G. (2003). Modeling dispersal with capture-recapture data: disentangling decisions of leaving and settlement. *Ecology,* **84,** 1225-1236.

Krebs, J. (1997). *Bovine tuberculosis in cattle and badgers. Report to The Rt Hon Dr Jack Cunningham MP*. MAFF.

Kruuk, H. & Parish, T. (1987). Changes in the size of groups and ranges of the European badger (*Meles meles* L.) in an area in Scotland. *The Journal of Animal Ecology***,** 351-364.

Lebreton, J.-D., Burnham, K. P. Clobert, J. & Anderson, D. R. (1992). Modeling survival and testing biological hypotheses using marked animals: a unified approach with case studies. *Ecological monographs,* **62,** 67-118.

Mahmood, K., Rook, G., Stanford, J., Stuart, F. A. & Pritchard, D. (1987). The immunological consequences of challenge with bovine tubercle bacilli in badgers (*Meles meles*). *Epidemiology and Infection*, 98, 155-163.

Page, R., Ross, J. & Langton, S. (1994). Seasonality of reproduction in the European badger *Meles meles* in southwest England. *Journal of Zoology,* **233,** 69-91.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. URL http://www.R-project.org.

Smith, G. C. & Cheeseman, C. L. (2007). Efficacy of trapping during the initial proactive culls in the randomised badger culling trial. *Veterinary Record*, **160**, 723-726.

Smith, G. C., McDonald, R. A. & Wilkinson, D. (2012). Comparing Badger (*Meles meles*) Management Strategies for Reducing Tuberculosis Incidence in Cattle. *PLoS ONE,* **7,** e39250.

Spiegelhalter, D., Thomas, A., Best, N. & Gilks, W. (1996). *BUGS 0.5: Bayesian inference Using Gibbs Sampling - Manual* (version ii), Medical Research Council Biostatistics Unit, Cambridge.

Tolhurst, B. A., Delahay, R. J., Walker, N. J., Ward, A. I. & Roper, T. J. (2009). Behaviour of badgers (Meles meles) in farm buildings: Opportunities for the transmission of Mycobacterium bovis to cattle? *Applied animal behaviour science,* **117,** 103-113.

Tomlinson, A. J., Chambers, M. A., McDonald, R. A. & Delahay, R. J. (submitted). Association of quantitative interferon-gamma responses with progression of naturally acquired *Mycobacterium bovis* infection in wild European badgers (*Meles meles*). *Immunology*.

Vicente, J., Delahay, R. J., Walker, N. & Cheeseman, C. L. (2007). Social organization and movement influence the incidence of bovine tuberculosis in an undisturbed high density badger *Meles meles* population. *Journal of Animal Ecology* 76, 348-360.

Williams, B., Nichols, J. D. & Conroy, M. J. (2002). *Analysis and management of animal populations: Modeling, estimation and decision making*. Academic Press, London.

Woodroffe, R., Bourne, F. J., Cheeseman, C. L., Cox, D. R., Donnelly, C. A., Gettinby, G., McInerney, J. P. & Morrison, W. I. (2005). Welfare of badgers (*Meles meles*) subjected to culling: development and evaluation of a closed season. *Animal Welfare* **14**, 19-25.

Woodroffe, R., Frost, S. D. W. & Clifton-Hadley, R. S. (1999) Attempts to control tuberculosis in cattle by removing infected badgers: constraints imposed by live test sensitivity. *Journal of Applied Ecology*, **36**, 494-501.

Appendix 1.

The number of observations (badger trapping and sampling events) at Woodchester Park across each season and year of the study from July 2006 to October 2013.

Year	Season	Time step*	Number of sampling events by season	Number of sampling events by year
2006	summer	1	99	156
	autumn	2	57	
2007	winter	3	41	335
	spring	4	116	
	summer	5	101	
	autumn	6	77	
2008	winter	7	39	350
	spring	8	102	
	summer	9	131	
	autumn	10	78	
2009	winter	11	60	290
	spring	12	103	
	summer	no capture[#]	0	
	autumn	13, 14	71 + 56 = 127	
2010	winter	no capture[#]	0	241
	spring	15	125	
	summer	16	81	
	autumn	17	35	
2011	winter	18	40	186
	spring	no capture[#]	0	
	summer	19, 20	76 + 70 = 146	
	autumn	no capture[#]	0	
2012	winter	21	38	246
	spring	22	102	
	summer	23	69	
	autumn	24	37	
2013	winter	25	19	218
	spring	26	69	
	summer	27	85	
	autumn	28	45	
Total	30 seasons	28 time steps	2022	2022

*Time step represents one sampling point, each of which consisted of three trapping clusters in order to target badgers across the three regions (A, B and C) of the study population.

[#]Although sampling was conducted four times per year throughout this study, the timing of trapping sessions was not always equally spaced. Hence, some sessions fell into the previous or next season depending on exact timing. Season dates are given in Table 16.

References to published material

9. This section should be used to record links (hypertext links where possible) or references to other published material generated by, or relating to this project.