Quantification of *Mycobacterium bovis* transmission in a badger vaccine field trial

Supplementary material

- 1. Initial derivations and model building

The derivations to support the estimations of the transmission parameters have already been published (12). In summary, and before considering vaccination, we showed how the number of new infections among all susceptible badgers (C) at the end of the time interval between two sweeps can be modelled using a binomial distribution with parameters S and p (where S is the number of susceptible badgers at the beginning of the time interval and $p = (1 - e^{-\beta * I * \frac{\Delta t}{N}})$ is the probability that each of these badgers will become infected during that time interval). The expected number of new cases is then equal to:

$$E(C) = S * (1 - e^{-\beta * I * \frac{\Delta t}{N}})$$

Then, as the number of new cases among susceptibles (C) can be obtained through serology, the transmission parameter β could be estimated by using a Generalized Linear Model (GLM) with a complementary-log-log link function, a binomial error function, with binomial total S and offset $ln\left(I * \frac{\Delta t}{N}\right)$.

We followed then by deriving new formulae that took into account vaccination, showing how the number of new infections among susceptibles at the end of each sweep could be estimated separately for vaccinated and non-vaccinated badgers as:

$$E(C_V) = S * (1 - e^{-e^{k_0 + k_1 * F_1} * Prev * \Delta t})$$

(where β_{UV} =Exp[k_0] and β_{VV}=Exp[k_0+k_1]) and

$$E(C_U) = S * (1 - e^{-e^{k_0 + k_1 * F_1} * Prev * \Delta t})$$

(where β_{UU}=Exp[k_0] and β_{VU}=Exp[k_0+k_1]), respectively.

Prev being the prevalence (I/N) and Fi the fraction of infected vaccinated badgers in the zone where the badger was trapped at the beginning of the time interval. Here, as C_V and C_U are known, all four betas can be calculated by using a GLM similar to the one described above.

In this paper, by assuming separable mixing (and as the vaccination status of each individual badger is available), we can then modify the GLM model introducing vaccination status of the recipient also as

one of the explanatory variables allowing us to estimate all the transmission rate parameters from the total population:

$$cloglog\ (C) = \beta_0 + \beta_{1,B} Z_B + \beta_{1,C} Z_C + \beta_2 Vs + \beta_3 Fi$$

A manual model building was carried out to select the final model; as the number of predictors in the maximum model was small, all possible combination of predictors were examined. The model including all three predictors: zone, vaccination status and fraction of infected vaccinated badgers was selected based on the lowest AIC (AIC = 490.2).

- 2. Descriptive results

Table S1. Number and percentage of positive trappings to the Enfer chemiluminescent multiplex ELISA at each sweep, overall and by zone.

	Zone						Overall	
	A		B		C			
Sweep	No	% Positive	No	% Positive	No	% Positive	No	% Positive
1	120	31.67	63	19.05	115	23.48	298	25.84
2	40	12.5	22	22.73	40	30	102	21.57
3	101	34.65	46	23.91	85	31.76	232	31.47
4	99	25.25	49	36.73	60	35	208	30.77
5	105	26.67	46	26.09	98	16.33	249	22.49
6	90	37.78	67	35.82	98	30.61	255	34.51
7	117	27.35	65	36.92	80	27.5	262	29.77
8	59	37.29	30	33.33	63	32.81	153	34.64

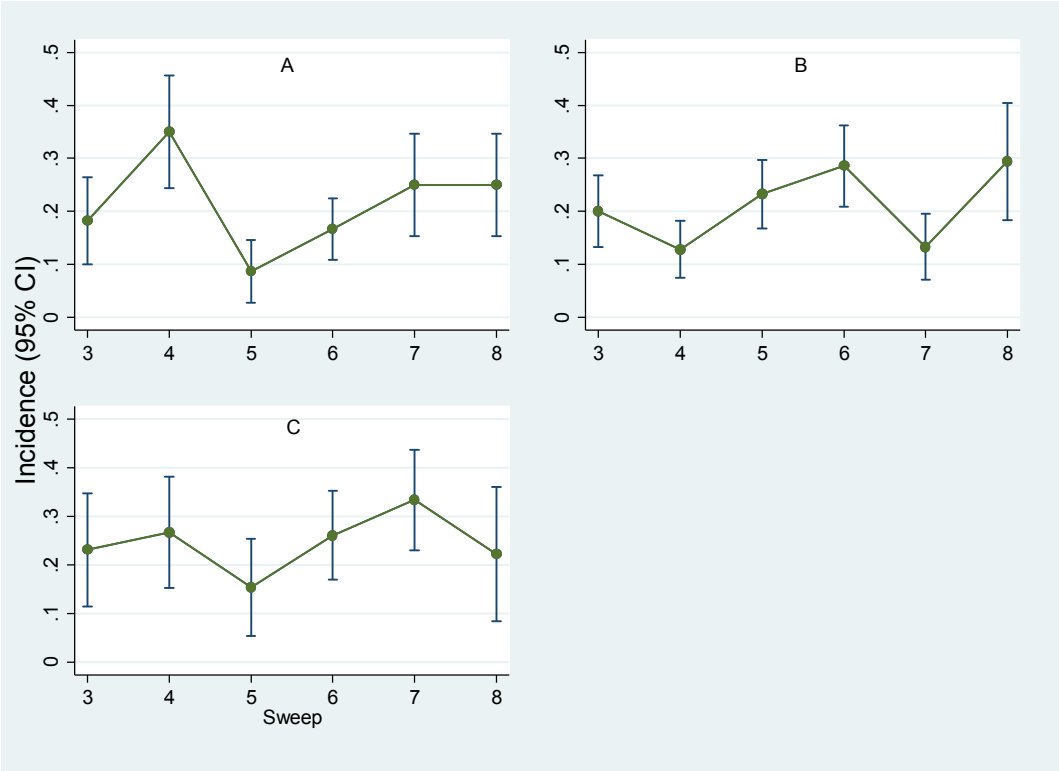

Figure S1. Incidence per sweep (estimated as number of new cases divided by the total number of susceptible badgers per sweep) and exact confident intervals (95% CI) presented by trial zone.

- 3. Supporting modelling exercises

Table S2. Generalized Linear Model results showing coefficient, p-value and 95% confidence interval for the explanatory variable (vaccination status) and constant when the model was fitted in data from zone B only.

Variable	Coef	p-Value	(95% CI)
Constant	-5.97	<0.001	-5.97 to -4.90
Vs	**-0.79**	**0.059**	**-1.60 to -0.03**

Table S3. Generalized Linear Model results showing coefficient, p-value and 95% confidence interval for all three explanatory variables (and constant): zone, vaccination status and the fraction of infected vaccinated badgers when the model was fitted in data from zones B and C only.

Variable	Coef	p-Value	(95% CI)
Constant	-5.55	<0.001	-6.09 to -5.00
Zone			
B		Reference	
C	0.78	0.832	-0.64 to 0.79
Vs	**-0.92**	**0.029**	**-1.76 to -0.09**
Fi	1.67	0.065	-0.11 to 3.45

- 4. Supplementary note on omitted data

On reconciliation of project databases at the completion of the trial, it was noted that 70 samples had not been tested by Enfer, due to a procedural error. These samples belonged to badgers that had been trapped in sweep 2 and were either trapped in sweep 1 or later on at subsequent sweeps. The samples were equally distributed between all three zones. The samples were also equally distributed in terms of vaccination status of the badger they belonged to. The missing samples represented between 9.0% and 10.7% of our entries for statistical analysis (depending on the infection status). The reduction in power was not considered to be a major constraint as the statistical analysis, in the absence of these 70 samples, showed significant results for VE_S. Additionally, even with the inclusion of these 70 samples, the estimates of VE_I would never be significantly above zero.