Supplementary Material to Bayesian inference for within-herd prevalence of *Leptospira interrogans* serovar Hardjo using bulk milk antibody testing

FRASER I. LEWIS∗, GEORGE J. GUNN

*Epidemiology Research Unit, Scottish Agricultural College, King’s Buildings, West Mains Road, Edinburgh, EH9 3JG*

fraser.lewis@sac.ac.uk

IAIN J. MCKENDRICK

*Biomathematics & Statistics Scotland, King’s Buildings, West Mains Road, Edinburgh, EH9 3JZ*

FIONA M. MURRAY

*Veterinary Services, Scottish Agricultural College, King’s Buildings, West Mains Road, Edinburgh, EH9 3JG*

APPENDIX

A. Farm selection and empirical distribution of bulk milk PP

Figure 1 shows a comparison of the empirical distribution of bulk milk PP values amongst the original pilot study of 34 farms, where only bulk milk rather than matched bulk milk and individual blood samples were taken, and the subset of 12 farms chosen from these initial 34 for whole herd testing. The 34 farms were stratified into three distinct blocks which are apparent in panel (a): low bulk milk PP (less than 20%); medium bulk milk PP (between 40% and 80%); and high bulk milk PP (greater than 90%). Farms were recruited within each block according to practical considerations such as geographical location and
perceived willingness of the farmer to take part in the study.

Fig. 1. Comparison of empirical distribution of bulk milk PP values in the pilot study of 34 farms (a) and the subset of 12 farms (b) chosen for whole herd serum testing. Note that the bulk milk PP values shown in (b) are not a direct subset of those in (a), as on each of the follow-up farms new bulk milk samples were taken to directly correspond with the serum samples taken from that herd on the day of sampling.
B. MULTIPLE SOLUTIONS

Figure 2 shows marginal output from two typical chains simulating from the general model. The uppermost chain was given an initial starting value close to one solution and the bottom chain a starting value close to the second solution. This behaviour was similarly present in output from all models, where the particular solution sampled depended on the initial region of parameter space from which the chain was initialised. Jumping between solutions was never observed.

Fig. 2. Trace plots of the gradient term in the general model (bulk milk term, overdispersion scaled to gradient term, logistic link). The uppermost trace is from a chain which was initialised close to the solution where the intercept was negative and the gradient positive, and the lower trace vice versa. The sampler does not jump between solutions and the marginal distributions are identical (not shown) except for sign.
C. Comparison of scaled and unscaled random effect parameterisations

Two different parameterisations of our general model defined in eqns (2.1) - (2.4) in the main manuscript were considered; firstly, the standard form for a linear mixed model of \( f(p_i) = \theta + \beta \text{bulk milk}_i + \phi_i \), and secondly, \( f(p_i) = \theta + \beta (\text{bulk milk}_i + \phi_i) \). As mentioned in the main manuscript these two parameterisations are mathematically equivalent, however, in practice we find that the latter parameterisation has rather better mixing. To examine this we simulated a range of data sets of varying size, both in terms of the total number of farms and the number of animals sampled on each farm, from each parameterisation. Figure 3 shows typical trace output for simulated data sets of the same size as the observed data with parameter estimates: \( \theta = -9.6, \beta = 0.09, S = 0.86, C = 0.95, \sigma_f^2 = 0.5 \) where \( \sigma_f^2 \) is defined for the standard form (and the variance of the random effects is matched in the alternative form using \( \sigma_f^2 = 0.5/(0.09^2) \)).

Two different data sets were generated, one from each parameterisation, and then both model parameterisations were fitted to each for comparison. It is clear that the parameterisation where \( \phi_i \) is scaled by \( \beta \) has better mixing in both data sets. The trace of the log likelihood function is not shown, however, we find that the higher variance in samples from the unscaled parameterisation appears to be due to the associated chains being much more prone to wandering into areas of lower likelihood and remaining there for some time. This results in parameter estimates with higher variances and also a poorer log marginal likelihood. For example, for the simulated data used in panels (a) and (b) in Figure 3 we find that the log marginal likelihoods are -356.1 and -355.3 for the unscaled and scaled parameterisations respectively. Similarly, for the data in panels (c) and (d) we find log marginal likelihoods of -363.7 and -362.3 for the unscaled and scaled parameterisations respectively. The differences in log marginal likelihood, while very modest, consistently result in an improved goodness of fit for the scaled parameterisation; more importantly, however, is that the scaled parameterisation results in lower variance parameter estimates. Simulations using
a range of parameter values similar to those obtained from the observed data gave very similar results to those already discussed.

Fig. 3. Comparison of scaled and unscaled random effects parameterisation using simulated data with the same sample sizes as the observed data. param (i) is $f(p_i) = \theta + \beta_{\text{bulkmilk}}i + \phi_i$ and param (ii) is $f(p_i) = \theta + \beta (\text{bulkmilk}_i + \phi_i)$, both use a logistic link function.

For small sample sizes, similar to those from the observed data set, our simulations suggest that the
numerical properties of the scaled and unscaled parameterisations differ. As the amount of available data is increased, however, then the two parameterisations behave identically, as would be expected given that they are mathematically equivalent. Figure 4 is similar to Figure 3 but uses two simulated data sets where the sample sizes are 28 farms and 200 observations within each farm. It is clear from the figure that with this increased amount of data the mixing in the two parameterisations is much more similar than in Figure 3. Indeed, we now find that the log marginal likelihood for each parameterisation applied to the same data set are identical (to two decimal places). Other simulations using different parameter estimates and sample sizes gave similar results.
Fig. 4. Comparison of scaled and unscaled random effects parameterisation using simulated data for 28 farms and 200 observations within each farm. param (i) is $f(p_i) = \theta + \beta \text{bulk milk}_i + \phi_i$ and param (ii) is $f(p_i) = \theta + \beta (\text{bulk milk}_i + \phi_i)$, both use a logistic link function.
D. Comparisons with Alternative Models

Figures 5 and 6 compare output from the models with random effects denoted in Table 2 in the main manuscript. As mentioned these models all have similar log marginal likelihood values. The variance in the predictions from our favoured model, $\theta + \beta(bulkmilk_i + \phi_i)$ with logistic link, is rather less than that in the other models containing random effects. In terms of differences in estimates of ELISA specificity ($C$) and sensitivity ($S$) we find that the former is virtually identical in each of the models. In contrast, estimates of $S$ are more variable between models, with the estimates from our favoured model having a rather higher median value for $S$. There is a suggestion of bi-modality in $S$ in two of the models, particularly in $\theta + \beta(bulkmilk_i + \phi_i)$ with a complementary log link and to a lesser extent in the same model with logistic link. This is somewhat unexpected and appears to occur when sampled values for the intercept term are highly negative, which can be clearly seen in Figure 7. One possible explanation is that there are two competing trajectories which both fit the data well but have different combinations of intercept and sensitivity. This feature appears to affect the parameterisation $\theta + \beta(bulkmilk_i + \phi_i)$ to a much greater extent than that used in our favoured model, e.g. panels (a) and (b) in Figure 7 do not exhibit such high variance in the intercept term as that in panels (c) and (d).
Fig. 5. Comparison of posterior distributions for mean prevalence in models with comparable goodness of fit (see Table 2 in main manuscript). Panel (a) $\theta + \beta_{\text{bulkmilk}} i + \phi_i$ light blue - logistic, blue - cloglog; (b) $\theta + \beta_{\text{bulkmilk}} i + \phi_i$ red - logistic, orange - cloglog; (c) Comparison of medians for each of the four models, A-D are the medians from panels (a)-(b) with the logistic link followed by the cloglog link respectively in each panel.
Fig. 6. Comparison of posterior distributions for $C$ and $S$ in models with comparable goodness of fit (see Table 2 in main manuscript). In order A-D are: $\theta + \beta(bulkmilk_i + \phi_i)$ light blue - logistic, blue - cloglog; $\theta + \beta(bulkmilk_i + \phi_i)$ red - logistic, orange - cloglog;
Fig. 7. Comparison of posterior distributions for the intercept term and $S$ in selected models. $\theta + \beta (\text{bulk milk}_i + \phi_i)$ logistic (a), cloglog (b); $\theta + \beta \text{bulk milk}_i + \phi_i$ logistic (c), cloglog (d). The intercept term in the models in panels (c) and (d) is much more variable than those in panels (a) or (b).
E. JACKKNIFE RESULTS

To assess the robustness of our chosen model, $\theta + \beta(bulkmilk_i + \phi_i)$ with logistic link, given the small number of individual farms in the study we fit this model to jackknife samples from our data (Efron and Tibshirani, 1993). Given a data set $x = (x_1, x_2, \ldots, x_n)$, the jackknife sample $x_{(i)}$, is defined to be $x$ with the $i$th data point removed, $x_{(i)} = (x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ for $i = 1, \ldots, n$. Figure 8 suggests that overall, our posterior estimates of mean within herd prevalence conditional on bulk milk PP (panels a and b) and test specificity $C$ (panel c) are reasonably robust to the individual choice of farms. However it is clear that the observations from farm 12 have a substantial influence on the estimation of test sensitivity $S$ (and to a lesser extent on mean within herd prevalence, see panel a. Note also that these parameters are not independent). That farm 12 should have the greatest influence of all the individual data points is perhaps not surprising given its relative position in Figure 1 (in the main manuscript). Farm 12 has by far the largest bulk milk PP value and also the highest mean within herd prevalence, and therefore has a strong influence on the shape of the response function. When this data point is removed the shape of the response function is most strongly influenced by data from farms 8 - 11, which while having similar bulk milk PP values are highly variable in terms of mean within herd prevalence. For example removing farm 12 greatly increases the variance in our estimates of $\theta$ and $\beta$, where the standard deviation in each of these typically doubles compared to the other jackknife samples (from typically 6 to over 14 in the case of $\theta$ and from 0.4 to over 1.0 in the case of $\beta$).
Fig. 8. Jackknife modelling results. The observations for each farm can be found in Table 1 in the main manuscript. The key denotes which farm is excluded. The black lines are for exclusion of each of the farms 1 through 5 individually, and are coloured similarly to ease clarity. Comparison of posterior distributions for mean prevalence (a) and (b); test sensitivity $S$ (c); and test specificity $C$ (d).
REFERENCES

Efron, B., Tibshirani, R. J., 1993. An Introduction to the Bootstrap. CRC.