This is an electronic appendix to the paper by Streftaris & Gibson 2004 Bayesian analysis of experimental epidemics of foot-and-mouth disease. *Proc. R. Soc. Lond.* B **271**, 1111–1117. (DOI 10.1098/rspb.2004.2715.)

Electronic appendices are refereed with the text. However, no attempt is made to impose a uniform editorial style on the electronic appendices.

Electronic Appendix A

MCMC implementation

Markov chain Monte Carlo (MCMC) methods are used for stochastic simulation and integration in situations where the distribution of interest cannot be easily approximated and marginal densities or moment estimates are difficult to obtain. They rely on the construction of a Markov chain whose stationary distribution is the distribution under consideration, and the use of the transition probabilities of the chain for determining suitable acceptance probabilities for updating the model parameters in an iterative scheme. A suitable sample taken from the converged chain can then be used for Monte Carlo inference regarding the distribution of interest. Here we utilise a single-component Metropolis–Hastings algorithm, in a manner similar to that described in Streftaris & Gibson (2004). Each model parameter is updated separately in a single step by first generating a candidate value y from a proposal distribution $q(y|x)$, where x is the current parameter value. The new value y is then accepted with a probability given as

$$\min\left\{1, \frac{p(y|\mathrm{w})/q(y|x,\mathrm{w})}{p(x|\mathrm{w})/q(x|y,\mathrm{w})}\right\}, \tag{A.1}$$

where the vector w consists of the remaining parameters at their latest value, and the data. The acceptance probability involves ratios of the so-called full conditional distribution $p(\cdot)$ of a parameter given all other parameters and the data, and the proposal distribution $q(\cdot)$.

Here the distribution of interest is the posterior distribution of the model parameters, as defined in Section 3. The prior setting is as follows. For the parameters of the Weibull distributions of latent and infectious periods we assume gamma priors: $\gamma_1 \sim \mathrm{Ga}(c_1, d_1); \delta_1 \sim \mathrm{Ga}(m_1, \phi_1); \gamma_2 \sim \mathrm{Ga}(c_2, d_2); \delta_2 \sim \mathrm{Ga}(m_2, \phi_2); \nu \sim \mathrm{Ga}(c_3, d_3); \lambda \sim \mathrm{Ga}(m_3, \phi_3)$. The gamma prior parameters are determined in a way such that the resulting distributions reflect existing knowledge about the epidemic characteristics of FMD in sheep under similar conditions. Thus the prior parameters were set so that the expected values of (γ_1, δ_1) resulted in a Weibull distribution with mean equal to one day, and $97.5th$ percentile $q_{0.975} = 3.0$ days; the expected

values of (γ_2, δ_2) gave a latent period from a Weibull distribution with a mean of 3 days, and $q_{0.975} = 8.0$ days; and for the infectious period the Weibull distribution has mean equal to 2 days and $q_{0.975} = 6.0$ days. The transmission parameter β is assigned a non-informative $\mathrm{Ga}(a, b)$ distribution with $a = 1, b = 0.001$, while for the power parameter α we assume a vague exponential prior distribution $\mathrm{Expon}(\theta)$ with $\theta = 0.001$ (i.e. again $\mathrm{Ga}(1, 0.001)$). Updating the above priors using the information in the data given in the likelihood function (3.2), we obtain the joint posterior density (3.4). The full conditional distributions are given by

$$p(\alpha|\theta, \beta, \mathbf{e}, \mathbf{r}, \mathbf{s}) \propto \prod_{j \in \mathcal{E}} \left[\beta \sum_{l=1}^{n} \{v_l^{\alpha} i_l(G_j, e_j)\} \right] \exp\left\{ -\theta\alpha - \beta \int_0^T C(t)dt \right\} \tag{A.2}$$

$$\beta|a, b, \mathbf{e}, \mathbf{s}, \mathbf{r} \sim \mathrm{Ga}\left(n_e + a,\ b + \int_0^T C(t)dt \right) \tag{A.3}$$

$$p(\gamma_1|\delta_1, c_1, d_1, \mathbf{e}, \mathbf{r}, \mathbf{s}) \propto \gamma_1^{n_1+c_1-1} \prod_{j \in \mathcal{I}_1} (s_j - e_j)^{\gamma_1 - 1} \times \exp\left\{ -d_1\gamma_1 - \delta_1 \sum_{j \in \mathcal{I}_1} (s_j - e_j)^{\gamma_1} \right\} \tag{A.4}$$

$$\delta_1|\gamma_1, m_1, \phi_1, \mathbf{e}, \mathbf{s}, \mathbf{r} \sim \mathrm{Ga}\left(n_1 + m_1,\ \phi_1 + \sum_{j \in \mathcal{I}_1} (s_j - e_j)^{\gamma_1} \right) \tag{A.5}$$

$$p(\gamma_2|\delta_2, c_2, d_2, \mathbf{e}, \mathbf{r}, \mathbf{s}) \propto \gamma_2^{n_e+c_2-1} \prod_{j \in \mathcal{I}_{2,3,4}} (s_j - e_j)^{\gamma_2 - 1} \times \exp\left\{ -d_2\gamma_2 - \delta_2 \sum_{j \in \mathcal{I}_{2,3,4}} (s_j - e_j)^{\gamma_2} \right\} \tag{A.6}$$

$$\delta_2|\gamma_2, m_2, \phi_2, \mathbf{e}, \mathbf{s}, \mathbf{r} \sim \mathrm{Ga}\left(n_e + m_2,\ \phi_2 + \sum_{j \in \mathcal{I}_{2,3,4}} (s_j - e_j)^{\gamma_2} \right). \tag{A.7}$$

In the above, n_1 and n_e denote the number of infected animals in G1 and the number of exposed animals in G2-G4 respectively, and $C(t)$ is defined as in (3.3). We note here that, as the infectious period was not estimated in the posterior analysis, the observed lengths of the viraemic periods were used (assuming that viraemic and infectious periods coincide and ignoring the uncertainty in the recordings) to evaluate the parameters of the relevant Weibull distribution involved in (3.2) by maximum likelihood estimation ($\nu = 4.443, \lambda = 0.006$).

When the full conditional distribution of a parameter is given in closed form and is available for sampling, it can serve as the proposal distribution $q(\cdot)$ leading to a Gibbs sampling step with acceptance probability equal to one. This is the case with the gamma full conditionals in (A.3), (A.5) and (A.7). For the remaining parameters suitable proposal distributions must be selected. We note here that the construction of the Markov chain in the MCMC methodology ensures that any proposal distribution will theoretically lead to convergence to the distribution of interest. However, the performance of the algorithm, and more specifically its convergence rate and its ability to explore the parameter space efficiently, will depend on the proposal choice. Here, for the shape Weibull parameters γ_1 and γ_2 we draw candidate values from suitably tuned gamma distributions, whose moments are matched to the moments of the corresponding full conditionals. This leads to an independence-type Metropolis–Hastings scheme, where the close resemblance of the proposal and full conditional distributions results in high acceptance rates ($88\% - 96\%$). The method is described in more detail in Streftaris & Gibson (2004). For parameter α, in order to avoid the computational cost involved in deriving a good approximation to (A.2), we employ a Metropolis algorithm, in which a uniform candidate distribution (centred at the current value of the parameter) provides a symmetrical random-walk step, simplifying (A.1) significantly.

Updating the unobserved vectors of event times $(\mathbf{e}, \mathbf{s}, \mathbf{r})$ is simpler than in cases where the epidemic modelling must take into account an unknown number of infections in the outbreak. Here the nature of the experiments implies that the size of the complete epidemic is known. However, the non-homogeneous and alternating patterns in the population mixing makes the updating of the hidden events problematic, as the available time windows for these events to give rise to plausible likelihood values are limited. We use a normal distribution, with mean equal to the current value, to propose new exposure times for the G2-G4 sheep. This again results in a symmetrical Metropolis sub-algorithm, with the variance of the normal proposal tuned (after an initial test algorithm run) to provide a balance between acceptance of new values and mixing of the chain. The times of onset of the infectious periods

(s), and recovery (r) were updated uniformly within the permissible 24-hour time windows.

The algorithm was run for 10^6 iterations, after an initial number of 20 000 'burn-in' samples were discarded to ensure that the chain had reached stationarity. The resulting chain was thinned (recording every $250th$ iteration) in order to reduce the autocorrelation within the simulated sequences. The vector of exposure times (e) was updated more frequently than the model parameters to allow the algorithm to explore the likelihood support more efficiently. The convergence of the algorithm was assessed by examining various criteria used in the software *Bayesian Output Analysis Program* (BOA, version 1.0.0, http://www.public-health.uiowa.edu).

Simulated epidemics

In order to assess and validate the power of the presented methodology to address the questions of the relation between viraemia and infectivity and the decrease of viraemia over the chain of virus transmission, we simulated data from epidemic outbreaks resembling those in the studied experiments. More specifically, following the design of the experiments, we considered populations of 32 animals, divided into four groups, with the G1 animals already exposed to the disease at the start of the simulation process. Under the non-homogeneous population mixing in the experiments, described in Section 1, times of exposure to the virus for the remaining animals (G2-G4), and times of onset and end of viraemic period for all susceptible animals were generated using stochastic simulation involving the probability distributions assumed in the model (Section 3). Peak viraemic levels were also simulated from appropriate models, as described later. Values of the various model parameters were chosen in a way such that they represent epidemic determinants in three scenarios corresponding to the following interactions between viraemic level, infectivity and length of infection chain: (i) no association between viraemia and infective challenge ($\alpha = 0$), levels of viraemia unaffected by number of infectious contacts; (ii) viraemia

related to individual infectiousness, viraemic levels unaffected by number of infectious contacts; and (iii) viraemia related to infectiousness, viraemic levels follow a decreasing trend as the length of chain of infection increases. The parameter values employed in the simulation process were the corresponding estimates using the data in experiment 1, with the appropriate model adjustments for $\alpha = 0$ for scenario (i), and $\alpha = 1$ for scenarios (ii) and (iii). To represent lack of relation between the length of the infection chain and the viral blood load of infected animals (scenarios (i) and (ii)), viraemic levels were generated as random values from a log-normal distribution fitted to the peak viraemic values in experiment 1. For scenario (iii) a regression model of viraemic level on length of infection chain (as estimated in both experiments) was employed to generate the appropriate decreasing values. The generated times of events associated with nine outbreaks, three outbreaks corresponding to each epidemic scenario described above, are shown in Figure 1.

The analysis of the simulated data verifies the capacity of the suggested methodology to identify the complex interactions involved in the epidemic process and distinguish between the different infection process dynamics in the three cases. When the histograms of the posterior distribution of parameter α in Figure 2 are compared, a clear shift of the distribution towards zero is revealed in the case where no relation between viraemia and infective challenge is assumed (scenario (i)). We also note here that a discrete model for α resulted in a very similar posterior distribution (Figure 3).

Finally, the effect of the length of the infection chain on the viraemic levels of infected animals, is shown by the posterior distributions of the p-values corresponding to ANOVA tests (based on partitions of the population derived using information from the entire epidemic process under the assumed Bayesian framework), as described in Section 3. Again, as shown in Figure 4, there is an apparent distinction between the two involved scenarios. In cases (i) and (ii) the histograms (large p-values) suggest unaffected levels of viraemia as the length of infection chain increases, while in case (iii) the model correctly identifies significant differences in the viraemic level of groups of animals categorised according to length of infection chain.

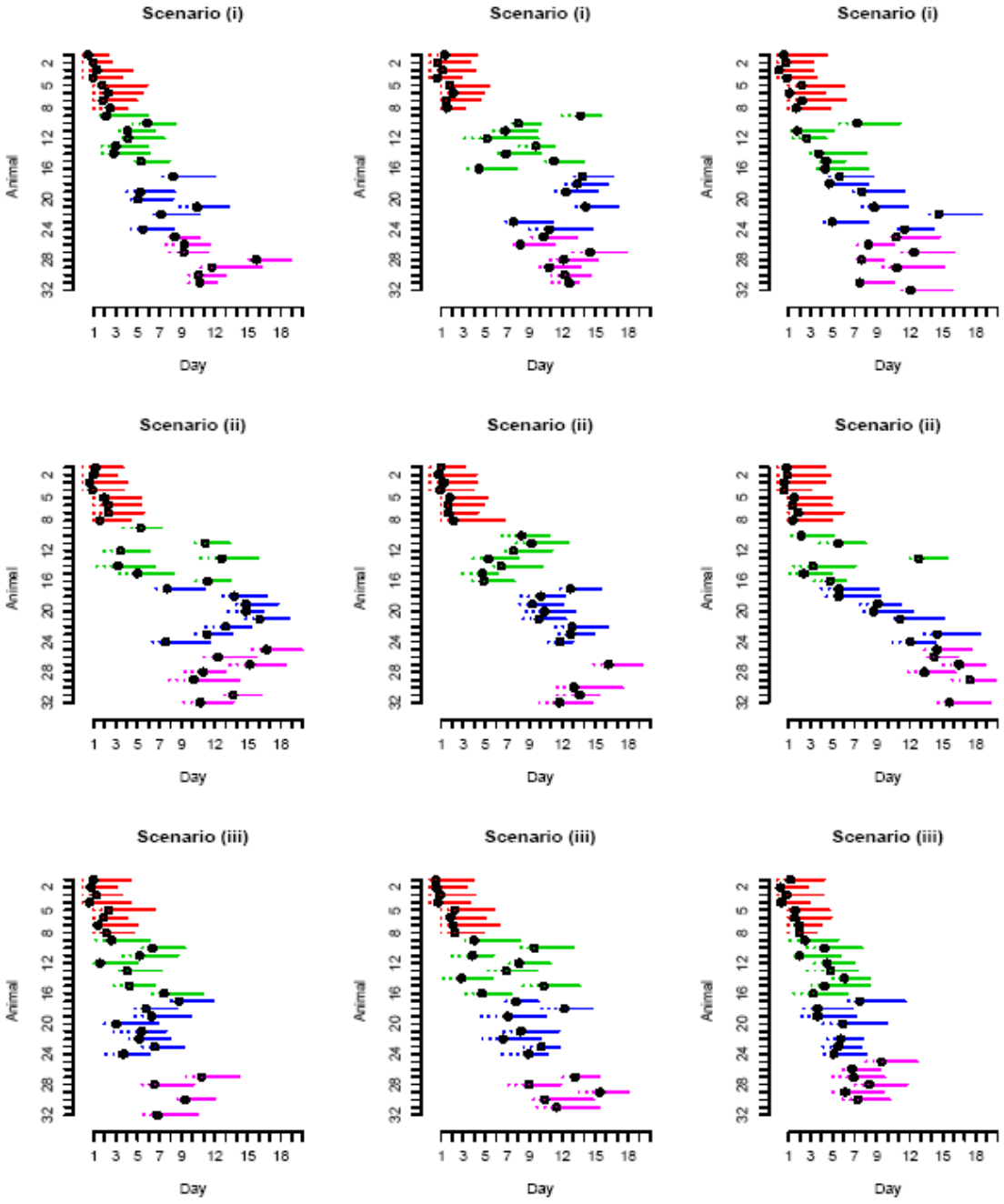

Figure 1: Latent periods (dotted lines) and infectious periods (solid lines) in simulated epidemic outbreaks. The circles show the onset of the infectious (viraemic) period. The different line colours correspond to the four groups of animals. The simulation process followed the design of the studied experiments. Each row of graphs corresponds to one of the three assumed epidemic scenarios: (i) $\alpha = 0$, no change in viraemic levels; (ii) $\alpha = 1$, no change in viraemic levels; (iii) $\alpha = 1$, decreasing viraemic levels.

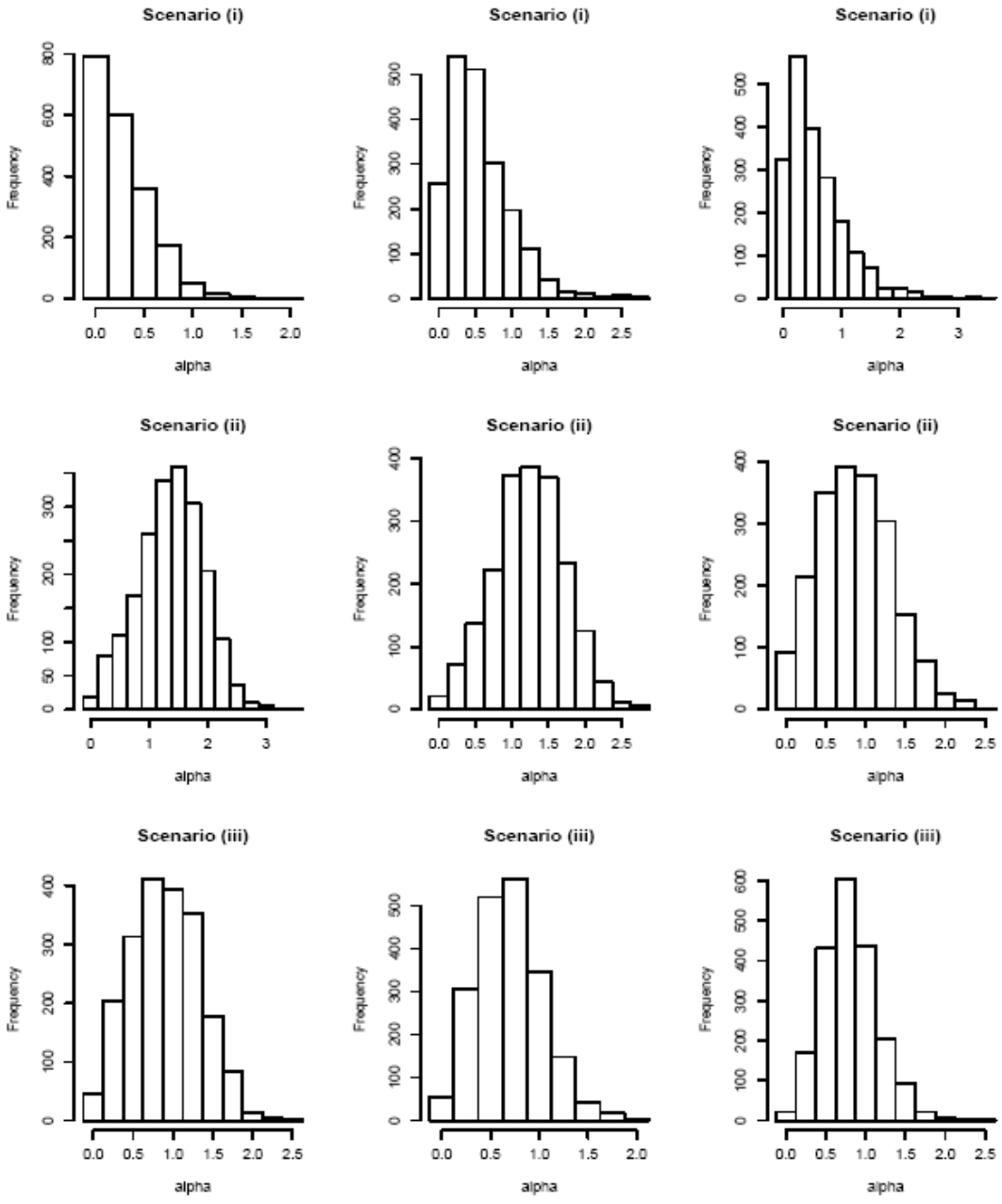

Figure 2: Posterior distribution of parameter α under the three assumed epidemic scenarios: (i) $\alpha = 0$, no change in viraemic levels; (ii) $\alpha = 1$, no change in viraemic levels; (iii) $\alpha = 1$, decreasing viraemic levels. A continuous model was employed for α, assuming an Exp(0.001) prior distribution.

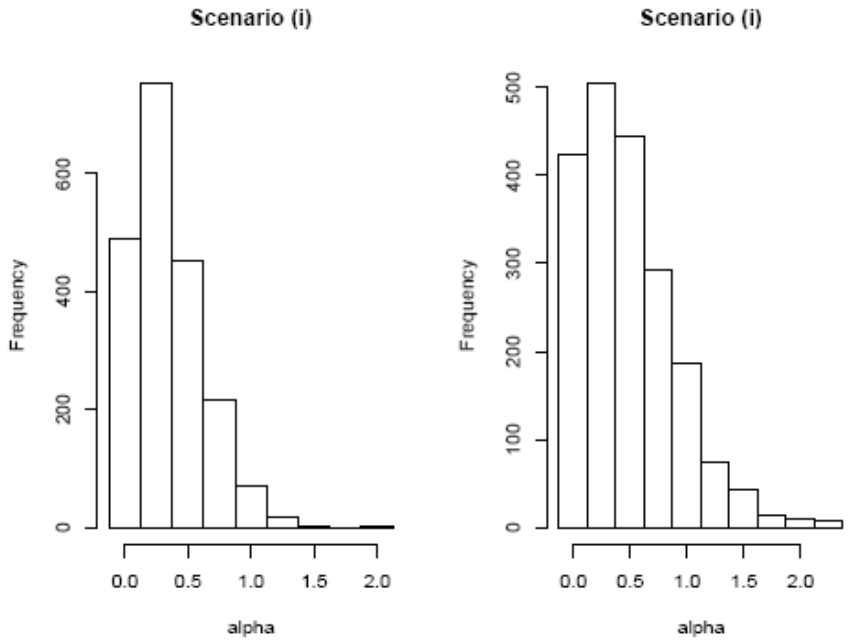

Figure 3: Posterior distribution of parameter α employing a discrete model, with α taking the values $(0.00, 0.25, 0.50, 0.75, 1.00, 1.25, 1.50, 1.75, 2.00, 2.25, 2.50)$ with equal probabilities. The graphs correspond to the two first simulated epidemics in Figure 2, under scenario (i): $\alpha = 0$, no change in viraemic levels.

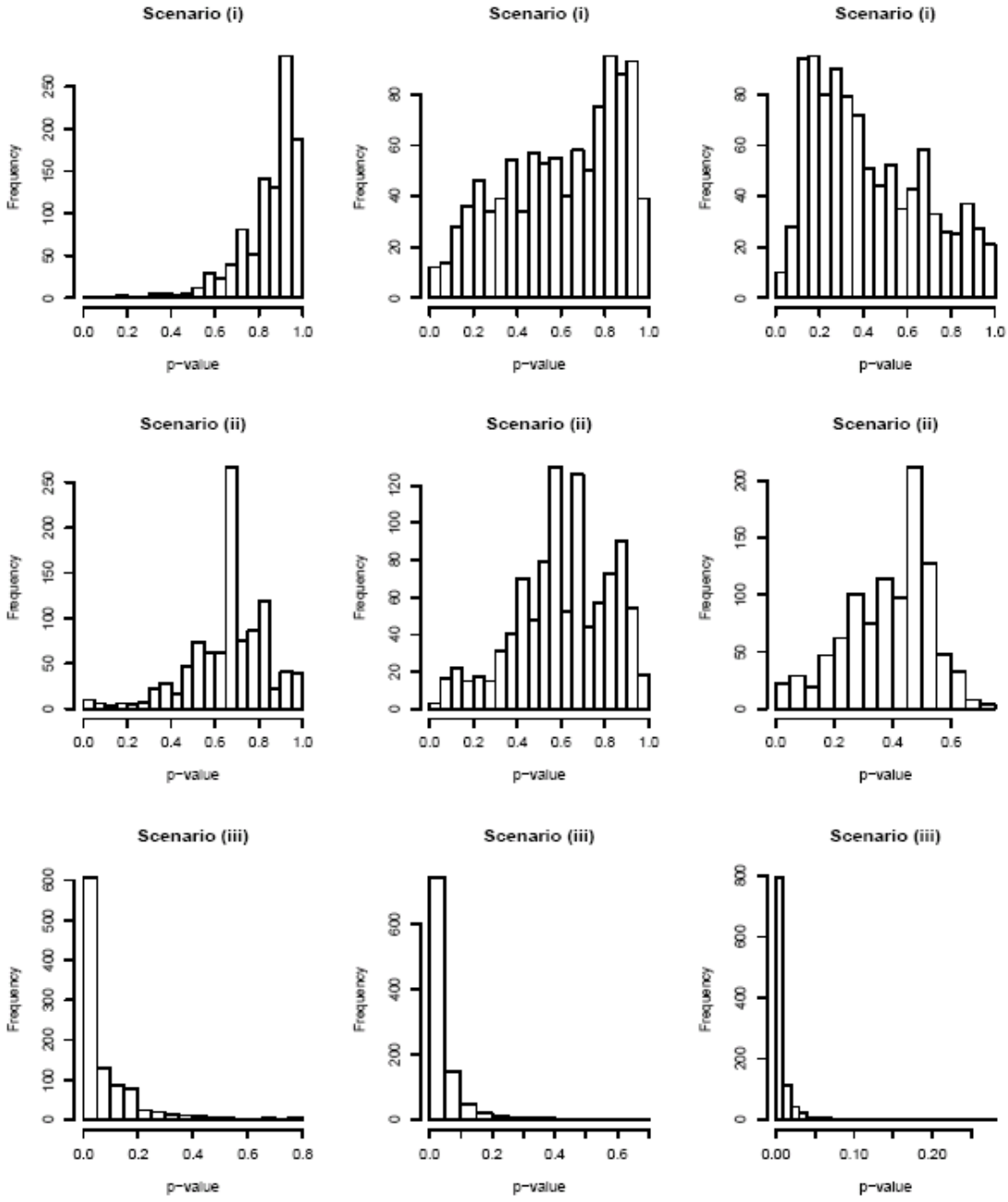

Figure 4: Posterior distributions of predictive p-values for testing the hypothesis of no differences among average peak viraemic levels of groups of sheep categorised according to length of infection chain. Each row of graphs corresponds to one of the three assumed epidemic scenarios: (i) $\alpha = 0$, no change in viraemic levels; (ii) $\alpha = 1$, no change in viraemic levels; (iii) $\alpha = 1$, decreasing viraemic levels.