**Inferring transmission parameters for foot-and-mouth disease virus in acutely-infected African buffalo**

Simon Gubbins

*The Pirbright Institute, Pirbright, U.K.*

**1 Data**

In the transmission experiments four African buffalo (*Syncerus caffer*) were inoculated with one of three strains of foot-and-mouth disease virus (FMDV). One strain each was selected from the three Southern African Territories (SAT) serotypes of FMDV, so that the strains are denoted SAT-1, SAT-2 or SAT-3 hereafter. After two days, four naïve in-contact buffalo were introduced to the pen containing the inoculated animals and allowed to mix. Blood samples were taken from all buffalo in a pen at approximately 2-3 day intervals for detection of FMDV RNA by qPCR.

**2 Parameter estimation**

***2.1 Modelling approach***

The transmission experiments were modelled using either *SIR* (susceptible (i.e. uninfected)-infected (and infectious)-recovered) or *SEIR* (susceptible-exposed (i.e. infected but not yet infectious)-infectious-recovered) models (Keeling & Rohani 2008). In both models a buffalo was assumed to be infectious if FMDV RNA was detected in its blood.

The force of infection for a buffalo is given by,



where *s* is the strain used in the transmission experiment (i.e. SAT-1, SAT-2 or SAT-3), *I*(*t*) and *N*(*t*) are the number of infectious buffalo and total number of buffalo in the pen at time *t*, respectively, and *βs* is the strain-specific transmission rate. Where included, the latent period (i.e. the time from infection to becoming infectious) for each buffalo was assumed to be drawn from a gamma distribution with strain-specific mean and shape parameter Finally, the duration of the infectious period was assumed to follow a gamma distribution with strain-specific mean and shape parameter  The basic reproduction number for strain *s* can be computed as,



***2.2 Bayesian methods***

Parameters (transmission rates, latent and infectious period parameters) were estimated in a Bayesian framework, including a data augmentation step, such that the unobserved infection times and partially-observed latent and infectious periods are included in the analysis as nuisance parameters (Hu et al. 2017).

The likelihood for the data, *L*, can be split into two components, one relating to the inoculated buffalo, *LI*, and one related to the in-contact buffalo, *LC*, so that,



where **φ** is a vector of model parameters and  **E**={*Ej*} and **I**={*Ij*} are vectors of infection times, latent periods and infectious periods, respectively. The likelihood for the *SIR* model is identical except that terms related to the latent periods do not need to be included.

Inoculated buffalo were assumed to be infected at *t*=0 and all were positive at the first observation time after inoculation (*t*=2). Accordingly, no inferences were drawn about the latent period for inoculated buffalo. However, the resulting uncertainty in the infectious period was incorporated in the likelihood by assuming the latent period could be between zero and two days, so that,



where *sj* is the strain with which animal *j* was inoculated, *fI* is the probability density function (PDF) for the gamma-distributed infectious periods.

The likelihood for the in-contact buffalo is given by,



The first term in the likelihood, , is the probability that buffalo *j* becomes infected at time  (conditional on not having been infected previously; *t*0 is the time at which the in-contact buffalo were put in the pen). The second term corresponds to the gamma-distributed latent period for buffalo *j*, *Ej* (with PDF, *fE*). The third term corresponds to the gamma-distributed infectious period for buffalo *j* (with PDF, *fI*, and cumulative density function (CDF), *FI*), where *cj* is a variable indicating whether (*cj*=1) or not (*cj*=0) animal *j* was still viral RNA-positive at the final observation time.

The infection times, latent periods and infectious periods for the animals were constrained so that they were consistent with the virus isolation data. In the case of the *SIR* model, the appropriate constraints are,



while for the *SEIR* model, they are,



where *LN*, *FP*, *LP* and *FN* denote the last negative, first positive, last positive and first negative blood sample, respectively.

The results for all strains were analysed together, considering several possibilities for variation amongst strains in the transmission rate and latent and infectious period parameters (Table 1). Where they varied amongst strains, the parameters for each strain were assumed to be drawn from higher-level exponential distributions (i.e. there is hierarchical structure in the parameters). In this case,



where *μ*p is the mean for the hierarchical distribution for parameter p (=*β*, *kE*, *μE*, *kI* or *μI*). In total, ten models were fitted to the data: four *SIR* models and six *SEIR* models (Table 1).

Mildly-informative priors were constructed for the latent and infectious period parameters based on data for the acute phase of infection for the buffalo (Maree et al. 2016). Specifically, an exponential prior with mean 1 was used for the shape parameter for the latent period, a gamma prior with mean 2 and shape 2 was used for the mean latent period, an exponential prior with mean 1 was used for the shape parameter for the infectious period, a gamma prior with mean 6 and shape 2 was used for the mean latent period. For the transmission parameter, a gamma prior with mean 1.85 and shape 1.39 was constructed (see appendix) based on the outcome of previous transmission experiments (Gainaru et al. 1986; Vosloo et al. 1996). When between-strain variation (i.e. hierarchical structure) was included for a parameter, the hyperprior for the parameter mean (*μ*p) was the same as the prior for that parameter without hierarchical structure.

Samples from the joint posterior density were generated using an adaptive Metropolis scheme (Haario et al. 2001), modified so that the scaling factor was tuned during burn-in to ensure an acceptance rate of between 20% and 40% for more efficient sampling of the target distribution (Andrieu & Thoms 2008). Two chains of 10,000,000 iterations were run, with the preceding 10,000,000 iterations discarded to allow for burn-in of the chain. The chains were then thinned (taking every 1000th sample) to reduce autocorrelation amongst the samples. Convergence of the scheme was assessed visually and by examining the Gelman-Rubin statistic provided in the coda package (Plummer et al. 2006) in R (R Core Team 2018).

***2.3 Model comparison***

The most commonly used measure of model comparison, the deviance information criterion (DIC) (Spiegelhalter et al. 2002) is not uniquely defined for data augmentation models (Celeux et al. 2006), such as the one used in this study. In addition, DIC uses point estimates rather than the full posterior distribution (Gelman et al. 2014). Accordingly, the different models were compared using leave-one-out (LOO) cross-validation methods (Gelman et al. 2014; Vehtari et al. 2017). For computational efficiency, approximate LOO was implemented using Pareto-smoothed importance sampling (PSIS) from the MCMC chains as described in Vehtari et al. (2017). In the PSIS-LOO approach, we compute -2×expected log pointwise predictive density (elpd) for a model as a measure of its predictive accuracy. (Here, multiplication by -2 puts it on the conventional scale of deviance.)

**3 Results**

***3.1 Model comparison***

The *SIR* model with the lowest elpd is that in which all parameters vary amongst strains (model 4; Table 1). The model with the lowest elpd is the *SEIR* model in which all parameters vary amongst strains (Δelpd>4 for models 1-8; Table 1). However, the change in elpd compared with the *SEIR* model in which all parameters except the transmission rate differ amongst strains (model 9) is small (Δelpd=0.6; Table 1).

***3.2 Parameter estimates***

The marginal posterior distributions for the best-fit *SIR* and *SEIR* models (models 4 and 10, respectively) are shown in Figure 1 and summarised in Table 3. In both models the highest transmission rate (and also basic reproduction number) was that for SAT-1 and the lowest for SAT-3, with SAT-2 in between (Figure 1; Table 2). For the *SEIR* model, the mean latent period was highest for SAT-3 and lowest for SAT-1, with SAT-2 in between, while for the mean infectious period the order was reversed (Figure 1; Table 2).

***3.3 Inferred transmission dynamics***

The inferred transmission dynamics are shown in Figure 2 for the best-fit *SIR* and *SEIR* models (models 4 and 10, respectively). For SAT-1, there is little difference between the *SIR* and *SEIR* models, though the infection times for the in-contact buffalo are slightly earlier in the *SEIR* compared with the *SIR* model. However, in both cases, all four in-contact buffalo were most likely infected by the inoculated animals. For SAT-2, the infection times for the in-contact buffalo are earlier in the *SEIR* compared with the *SIR* model, markedly so for three of the animals. Moreover, in the *SEIR* model the in-contact buffalo are most likely infected by the inoculated animals, whereas in the *SIR* model, three of the in-contact buffalo could have been infected by an in-contact animal. Finally, for SAT-3, the transmission dynamics are quite different between the *SIR* and *SEIR* models. For the *SIR* model, there is a clear pattern of who infects whom: two in-contact buffalo are infected by the inoculated animals, these then infect the third in-contact buffalo, which subsequently infects the fourth in-contact animal. Although this pattern remains consistent with the infection times inferred using the *SEIR* model, other patterns are possible. This includes three of the in-contact buffalo being infected by the inoculated animals, one of which infects the fourth in-contact animal, as well as all four in-contact buffalo being infected by the inoculated animals.

**References**

Andrieu, C. & Thoms, J. 2008 A tutorial on adaptive MCMC. *Stat. Comput.* **18**, 343-373.

Celeux, G., Forbes, F., Robert, C.P. & Titterington, D.M. 2006 Deviance information criteria for missing data models. *Bayesian Analysis* **1**, 651-674.

Gainaru, M.D., Thomson, G.R., Bengis, R.G., Esterhuysen, J.J., Bruce, W. & Pini, A. 1986 Foot-and-mouth disease and the African buffalo (*Syncerus caffer*). II. Virus excretion and transmission during acute infection. *Ondesterpoort J. Vet. Research* **53**, 75-85.

Gelman, A., Hwang, J. & Vehtari, A. 2014 Understanding predictive information criteria for Bayesian models. *Stat. Comput.* **24**, 997-1016.

Haario, H., Saksman, E. & Tamminen, J. 2001 An adaptive Metropolis algorithm. *Bernoulli* **7**, 223-242.

Hu, B., Gonzales, J.L. & Gubbins, S. 2017 Bayesian inference of epidemiological parameters from transmission experiments. *Sci. Reports* **7**, 16774.

Keeling, M.J. & Rohani, P. 2008 *Modelling infectious diseases in humans and animals*. Princeton, New Jersey, U.S.A.: Princeton University Press.

Maree, F., de Klerk-Lorist, L.-M., Gubbins, S., Zhang, F., Seago, J., Perez-Martin, E., Reid, E., Scott, K., van Schalkwyk, L., Bengis, R., Charleston, B. & Juleff, N. 2016 Differential persistence of foot-and-mouth disease virus in African buffalo is related to viral virulence. *J. Virol.* **90**, 5132-5140.

Plummer, M., Best, N., Cowles, K. & Vines, K. 2006 CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News* **6**, 7-11.

R Core Team 2018 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (<http://www.R-project.org/>).

Spiegelhalter, D.J., Best, N.G., Carlin, B.P. & van der Linde, A. 2002 Bayesian measures of model complexity and fit (with discussion). *J. R. Stat. Soc. B* **64**, 583-639.

Vehtari, A., Gelman, A. & Gabry, J. 2017 Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27**, 1413-1432.

Vosloo, W., Bastos, A.D., Kirkbride, E., Esterhuysen, J.J., Janse van Rensburg, D., Bengis, R.G., Keet, D.W. & Thomson, G.R. 1996 Persistent infection of African buffalo (*Syncerus caffer*) with SAT-type foot-and-mouth disease viruses: rate of fixation of mutations, antigenic change and interspecies transmission. *J. Gen. Virol.* **77**, 1457-1467.

**Appendix Constructing the prior distribution for the transmission parameter**

Two previous experiments have considered transmission of FMDV from acutely-infected buffalo (Gainaru et al. 1986; Vosloo et al. 1996). For each experiment we extracted the number of inoculated buffalo, the number of in-contact buffalo, the number of these which became infected and the duration of the challenge (i.e. the time from when the in-contact animals were introduced to the inoculated ones until virus was first isolated from an in-contact animal) (see Table A1). Although different serotypes/strains were used in the experiments, for simplicity, this information was not incorporated in the analysis.

The likelihood for the data can be written as,



where *I* is the number of inoculated buffalo, *t*C is the duration of challenge, *nU* the number of in-contact buffalo that remain uninfected at the end of the challenge period, *nI* is the number of in-contact buffalo that become infected during the challenge period and *N* is the total number of buffalo in the *j*th experiment (Table A.1). The posterior distribution for the transmission parameter, *p*(*β*), is given by,



where *π*(*β*) is the prior distribution for the transmission rate (we used a non-informative exponential prior with mean 100). Rather than use the full posterior, , as the prior for our analysis, we constructed a less informative prior by using a gamma distribution parameterised so that its median and interquartile range coincided with the median and 95% credible interval of the posterior, . In this case, the mean is 1.85 and the shape is 1.39.

**Table A.1.** Previous experiments for the transmission of FMDV from acutely-infected buffalo.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Gainaru et al. (1986) |  |  | Vosloo et al. (1996) |
| serotype\* | SAT-1 | SAT-2 | SAT-2 | SAT-2 |
| number of challenge buffalo | 2 | 2 | 2 | 2 |
| number of in-contact buffalo |  |  |  |  |
| infected | 4 | 4 | 4 | 1 |
| total | 4 | 4 | 4 | 2 |
| duration of challenge (days) | 4 | 4 | 4 | 7 |

\* note: these are not the same strains as used in the present study

**Table 1.** Comparison of models for the transmission of foot-and-mouth disease virus strains in acutely-infected African buffalo.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| model | parameters† | | | | | -2×elpd‡ |
| *kE* | *μE* | *kI* | *μI* | *β* |
| *SIR* |  |  |  |  |  |  |
| 1 | - | - | C | C | C | 139.6 |
| 2 | - | - | C | C | V | 138.1 |
| 3 | - | - | V | V | C | 135.2 |
| 4 | - | - | V | V | V | 133.4 |
| *SEIR* |  |  |  |  |  |  |
| 5 | C | C | C | C | C | 138.5 |
| 6 | C | C | C | C | V | 134.1 |
| 7 | V | V | C | C | C | 130.2 |
| 8 | C | C | V | V | C | 135.4 |
| 9 | V | V | V | V | C | 126.7 |
| 10 | V | V | V | V | V | 126.1 |

† C: common to all strains; V: varies amongst strains

‡ elpd: expected log pointwise predictive density

**Table 2.** Transmission parameters for three strains of foot-and-mouth disease virus in acutely-infected African buffalo.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| parameter |  | SAT-1 | | SAT-2 | | SAT-3 | |
| estimate\* | 95% CI† | estimate\* | 95% CI† | estimate\* | 95% CI† |
| *SIR* model |  |  |  |  |  |  |  |
| infectious period |  |  |  |  |  |  |  |
| shape parameter | *kI* | 14.7 | (4.3, 41.0) | 7.4 | (2.2, 21.0) | 14.7 | (4.0, 44.3) |
| mean (days) | *μI* | 6.3 | (5.1, 8.0) | 5.3 | (3.9, 7.1) | 5.1 | (4.0, 6.5) |
| transmission rate | *β* | 1.9 | (0.5, 7.4) | 0.8 | (0.3, 1.7) | 0.5 | (0.2, 1.2) |
| basic reproduction number | *R*0 | 12.1 | (3.2, 49.0) | 4.0 | (1.3, 9.7) | 2.7 | (0.9, 6.2) |
| *SEIR* model |  |  |  |  |  |  |  |
| latent period |  |  |  |  |  |  |  |
| shape parameter | *kE* | 1.2 | (0.1, 8.7) | 1.6 | (0.2, 9.2) | 1.6 | (0.2, 8.3) |
| mean (days) | *µE* | 0.5 | (0.02, 2.4) | 1.3 | (0.1, 3.5) | 2.8 | (0.5, 7.0) |
| infectious period |  |  |  |  |  |  |  |
| shape parameter | *kI* | 11.8 | (3.5, 33.5) | 8.7 | (2.4, 27.0) | 11.8 | (3.3, 35.3) |
| mean (days) | *μI* | 5.7 | (4.4, 7.4) | 4.6 | (3.5, 6.3) | 4.2 | (3.2, 5.8) |
| transmission rate | *β* | 2.8 | (0.8, 11.3) | 1.6 | (0.4, 9.0) | 1.2 | (0.3, 7.8) |
| basic reproduction number | *R*0 | 15.8 | (4.1, 65.6) | 7.5 | (1.9, 41.5) | 5.2 | (1.3, 34.1) |

\* posterior median

† CI: credible interval



**Figure 1.** Transmission parameters for three strains of foot-and-mouth disease virus (SAT-1, SAT-2 and SAT-3) in acutely-infected African buffalo: (*a*) shape parameter for the latent period (*kE*); (*b*) mean latent period (in days) (*μE*); (*c*) shape parameter for the infectious period (*kI*); (*d*) mean infectious period (in days) (*μI*); (*e*) transmission rate (per day) (*β*); and (*f*) basic reproduction number (*R*0). Each plot shows the marginal posterior density (shape), the posterior median (circle) and 25th and 75th percentiles (line) for the parameter. The colour of the shape indicates the FMDV strain and the model used: SAT-1, *SIR* model (blue), SAT-1, *SEIR* model (green), SAT-2, *SIR* model (red), SAT-2, *SEIR* model (magenta), SAT-3, *SIR* model (cyan) and SAT-3, *SEIR* model (yellow).



**Figure 2.** Inferred transmission dynamics for three strains of foot-and-mouth disease virus (SAT-1, SAT-2 and SAT-3) in African buffalo during acute infection. Results are shown for the best-fit *SIR* and *SEIR* models. Each panel shows a violin plot of the inferred infection times for each in-contact buffalo, with the red shape indicating the posterior density for the infection time and the black circle and line indicating the posterior median and posterior interquartile range for the infection times. The blue shapes show the inferred times at which each buffalo was infectious, with the width of each shape indicating the proportion of samples for which the buffalo was infectious at that time. An "I" denotes an inoculated buffalo and a "C" denotes an in-contact buffalo.