**S1- S4 Estimating model parameters to understand endemic persistence of FMDVs in African buffalo**

**S5 Modelling FMDVs in African Buffalo populations**

Anna Jolles1,2, Erin Gorsich1,3, Simon Gubbins4, Brianna Beechler1, Peter Buss5, Nick Juleff6, Lin-Mari de Klerk-Lorist7, Francois Maree8,9, Eva Perez-Martin4, OL van Schalkwyk7, Katherine Scott8, Jan Medlock1, Bryan Charleston4

1 Department of Biomedical Sciences, Oregon State University, Corvallis, OR, 97331, USA

2 Department of Integrative Biology, Oregon State University, Corvallis, OR, 97331, USA

3 Department of Biology, Colorado State University, Fort Collins, CO, 80521, USA

4 Pirbright Institute, Ash Road, Woking, Surrey, GU24 0NF, United Kingdom

5 SANParks, Veterinary Wildlife Services, Kruger National Park, 1350 Skukuza, South

Africa

6 Bill & Melinda Gates Foundation, Livestock Program, Seattle, WA, USA

7 Office of the State Veterinarian, Department of Agriculture, Forestry and Fisheries,

Skukuza, 1350, South Africa

8 Transboundary Animal Disease Programme, ARC-Onderstepoort Veterinary Institute,

Private Bag X05, Onderstepoort 0110, South Africa

9 South Africa Department of Microbiology and Plant Pathology, University of Pretoria,

Pretoria, South Africa

**S1 Duration of maternal antibodies against FMDV in African buffalo**

The duration of maternal immunity was assumed to follow a gamma distribution. The likelihood for the distribution parameters (mean and shape parameter) is given by,



where *j* is an index identifying each calf, *s* is the serotype, *a*0 is the age at which the last protective titre was observed, *a*1 is the age at which the first non-protective titre was observed for each calf (see Fig. 1b in the main paper) and *f* is the probability density function for the gamma distribution with (serotype-specific) shape parameter and mean, *ks* and *µs*, respectively. (Note: the gamma-distribution is parameterised as,



where *k*>0 is the shape parameter, *µ*>0 is the mean and Γ(*k*) is a gamma function; this parameterisation for the gamma distribution, , is used for all analyses in the study.)

Two possibilities for the distribution parameters were considered. In the first, the parameters were assumed to be common to all serotypes, while in the second they were assumed to differ amongst the serotypes. In the second model, the parameters for each serotype were assumed to be drawn from higher-level gamma distributions (i.e. there is hierarchical structure in the parameters). Data from duration of maternal antibodies from an earlier experiment (Bengis et al. 1986) were used to construct a mildly informative prior for the mean duration (exponential with mean 0.5 years). Non-informative priors (exponential with mean 100) were assumed for all other parameters.

Samples from the joint posterior distribution 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 100,000 iterations were run, with the preceding 100,000 iterations discarded to allow for burn-in of the chain. The chains were then thinned (taking every 10th sample) to reduce autocorrelation amongst the samples. Convergence of the scheme was assessed visually and by examining various criteria provided in the coda package (Plummer et al. 2006) in R (R Core Team 2016). The two possibilities for the distribution parameters (i.e. common to serotypes or different amongst serotypes) were compared using the deviance information criterion (DIC) (Spiegelhalter et al. 2002).

**S2 Epidemiological parameters for African buffalo acutely infected with FMDV**

***S2.1 Parameter estimation***

To quantify transmission of FMDVs during acute infection the results of the experimental study were analysed 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,



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 (see Extended Data Table 2). 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 (see Extended Data Table 2).

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 section S2.2 below) 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 distribution 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).

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 analysis. 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.)

***S2.2 Constructing the prior distribution for the acute transmission rate***

Two previous experiments have considered transmission of FMDV from acutely-infected African 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 S1). 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 S1). The posterior distribution for the transmission rate, *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 S1.** Outcome of transmission experiments for FMDV in 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

**S3 Transmission rates for FMDV carrier African buffalo**

***S3.1 Data***

To quantify transmission from carrier buffalo to naive hosts the results of the experimental study were analysed up to the first time point at which transmission was observed to have occurred (i.e. an in-contact buffalo had a positive PCR result). Where transmission of the strain did not occur, results were analysed up to the last time point when samples for both carriers of the serotype were positive. In both cases this simplified the analysis by allowing us to assume the force of infection was constant during the study period. The results are summarised in Table S2 and were analysed independently for each strain, with carriers infected with other strains treated as in-contact animals.

**Table S2.** Outcome of experiments on the transmission for foot-and-mouth disease virus from carrier African buffalo.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| strain | group | *N* | *t*0 | *t*1 | *nU* | *nI* |
| SAT-1 | 1 | 12 | 0 | 14 | 3 | 7 |
|  | 2 | 11 | 14 | 44 | 3 | 6 |
| SAT-2 | 1 | 12 | - | 14 | 10 | 0 |
|  | 2 | 11 | - | 104 | 9 | 0 |
| SAT-3 | 1 | 12 | 0 | 14 | 7 | 3 |
|  | 2 | 11 | - | 44 | 9 | 0 |

*N* - number of buffalo in the group (i.e. carrier and in-contact)

*t*0 - time of last sampling at which all in-contact buffalo were PCR-negative (if transmission occurred)

*t*1 - time of first sampling at which in-contact buffalo were PCR-positive (if transmission occurred) or time of last sampling at which both carriers were positive (if transmission did not occur).

*nU* - number of uninfected in-contact buffalo at *t*1

*nI* - number of infected in-contact buffalo at *t*1

***S3.2 Parameter estimation***

The likelihood for each transmission experiment is given by



where *nU* and *nI* are the number of uninfected and infected in-contact buffalo at the end of the study period, respectively,



are the probabilities an in-contact buffalo remains uninfected or becomes infected during the study period, respectively, *λ*(*τ*) is the force of infection on day *τ* of the experiment, *t*0 is the time of last sampling at which all in-contact buffalo were PCR-negative and *t*1 is the time of first sampling at which in-contact buffalo were PCR-positive. The term in the likelihood, , relating to infection of in-contact buffalo allows for the fact not all in-contact buffalo may have been infected by carriers (i.e. once the first in-contact buffalo became infected, it could transmit to other in-contact animals). Here, the force of infection is given by



where *I* and *N* are the number of infectious (carrier) buffalo and total number of buffalo in the pen, respectively, and *β* is the transmission rate. In this case, the likelihood for those groups in which transmission occurred can be rewritten as,



while for those groups in which transmission did not occur, it can be rewritten as,



Finally, the likelihood was constructed by multiplying the appropriate likelihoods for groups 1 and 2 and strains SAT-1, SAT-2 and SAT-3.

Two possibilities for the transmission rates were considered. In the first, the rates were assumed to be common to all strains, while in the second they were assumed to differ amongst the strains. In the second model, the rates for each strain were assumed to be drawn from a higher-level exponential distribution (i.e. there is hierarchical structure in the parameters). An exponential prior with mean 1 was assumed for the (mean) transmission rate, based on that estimated for acutely-infected buffalo (Extended Data Table 3). Non-informative priors (exponential with a mean 100) were used for other parameters.

Samples from the joint posterior distribution 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 200,000 iterations were run, with the preceding 100,000 iterations discarded to allow for burn-in of the chain. The chains were then thinned (taking every 20th sample) to reduce autocorrelation amongst the samples. Convergence of the scheme was assessed visually and by examining various criteria provided in the coda package (Plummer et al. 2006) in R (R Core Team 2018).

The two possibilities for the transmission rates (i.e. common to serotypes or different amongst serotypes) were compared using the deviance information criterion (DIC) (Spiegelhalter et al. 2002). The model in which the transmission rates differed amongst strains resulted in a significantly better fit compared to that in which the transmission rate was common (DIC=16.8 for differing rates compared with DIC=19.8 for a common rate).

**S4 Duration of FMDV carrier status in African buffalo**

***S4.1 Data***

Data on virus isolation from tonsil swabs taken from eight African buffalo experimentally infected with the same three strains of FMDV (SAT-1, SAT-2 and SAT-3) as in the present study were used to estimate the duration of carrier status (Maree et al. 2016). Specifically, the data were used to determine the time post infection of the last positive sample for infectious FMDV and the first subsequent negative sample (Table S3). This assumes that the last positive sample was genuinely the last positive sample, but, because detection of FMDV was intermittent in all the buffalo, this may not be the case. Only buffalo which were confirmed carriers for the serotype (i.e. positive sample at 28 or more days post infection) were included in the analysis (seven for SAT-1, four for SAT-2 and five for SAT-3).

**Table S3.** Duration (in days post infection) of foot-and-mouth disease virus carrier state in African buffalo.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| reference | serotype | | | | | |
| SAT-1 | | SAT-2 | | SAT-3 | |
| *tlp*† | *tfn*‡ | *tlp* | *tfn* | *tlp* | *tfn* |
| Maree et al. 2016§ | 316 | 336 | 136 | 155 | 126 | 136 |
|  | >185 | - | >185 | - | >185 | - |
|  | >400 | - | nc¶ | nc | 95 | 109 |
|  | 200 | 214 | 35 | 80 | nc | nc |
|  | >185 | - | nc | nc | 109 | 126 |
|  | 168 | 185 | nc | nc | nc | nc |
|  | 214 | 231 | 35 | 80 | 35 | 80 |
| Dawe et al. 1994 | - | - | 126 | 154 | - | - |
|  | - | - | >188 | - | - | - |
| Vosloo et al. 1996 | - | - | 366 | 399 | - | - |
|  | - | - | nc | nc | - | - |

† *tlp* - time of last positive virus isolation (a > sign indicates the observation is right-censored)

‡ *tfn* - time of first subsequent negative virus isolation

¶ nc - not a carrier (i.e. no positive samples 28 days or more post infection)

§ animals were co-infected with all three serotypes

***S4.2 Statistical methods***

The duration of carrier status was assumed to follow a gamma distribution. The likelihood for the distribution parameters (mean *μ* and shape parameter *k*) is given by,



where *tp* is the time of the last positive sample, *tn* is the time of the first subsequent negative sample for the *j*th buffalo, *cj* is a variable indicating whether there was (*cj*=0) or was not (*cj*=1) a first negative sample (i.e. whether or not the observation for the animal is right-censored) and *f* is the probability density function for the gamma distribution and *sj* is the serotype with which the *j*th buffalo was infected.

Four possibilities for the distribution parameters were considered: (i) mean and shape parameter common to all serotypes; (ii) mean differs amongst serotypes and shape parameter common; (iii) mean common and shape parameter differs amongst serotypes; and (iv) mean and shape parameter differ amongst serotypes. Where they varied amongst serotypes, the parameters for each serotype were assumed to be drawn from higher-level gamma distributions (i.e. there is hierarchical structure in the parameters).

An informative prior was constructed for the mean duration of carrier status using data on the duration of carrier status in experimentally-infected African buffalo (Table S2) extracted from the published literature (Dawe et al. 1994; Vosloo et al. 1996). Fitting an exponential distribution to these data by maximum likelihood yielded an estimate for the distribution (or hierarchical) mean of 307. Non-informative priors (exponential with mean 100) were assumed for all remaining parameters.

Samples from the joint posterior distribution 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). For each serotype, two chains of 200,000 iterations were run, with the preceding 100,000 iterations discarded to allow for burn-in of the chain. Chains were subsequently thinned (by taking every 20th sample) to reduce autocorrelation amongst the samples. Convergence of the scheme was assessed visually and by examining various criteria provided in the coda package (Plummer et al. 2006) in R (R Core Team 2018).

The four possibilities for the distribution parameters (i.e. whether the mean or shape parameter are common to serotypes or different amongst serotypes) were compared using the deviance information criterion (DIC) (Spiegelhalter et al. 2002). The best-fit model (as judged by the DIC) was one in which the mean duration in carrier status varied amongst serotypes, but where the shape parameter was common to all serotypes (mean and shape common: DIC=76.3; mean differs, shape common: DIC=73.1; mean common, shape differs: DIC=76.2; mean and shape differ: DIC=73.3).

**References for S1-S4**

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

Bengis, R.G., Thomson, G.R., Hedger, R.S., de Vos, V. & Pini, A. 1986 Foot-and-mouth disease and the African buffalo (*Syncerus caffer*). 1. Carriers as a source of infection for cattle. *Ondesterpoort J. Vet. Res.* **53**, 69-73.

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

Dawe, P.S., Sorensen, K., Ferris, N.P., Barnett, I.T.P., Armstrong, R.M. & Knowles, N.J. 1994 Experimental transmission of foot-and-mouth disease virus from carrier African buffalo (*Syncerus caffer*) to cattle in Zimbabwe. *Vet. Record* **134**, 211-215.

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.

Gelman, A., Carlin, J.B., Stern, H.S. & Rubin, D.B. 2004 *Bayesian data analysis (2nd edition)*. Boca Raton, Florida, U.S.A.: Chapman Hall/CRC.

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.