Web Appendices for "Bayesian Hierarchical Regression on Clearance Rates in the Presence of 'Lag' and 'Tail' Phases with an Application to Malaria Parasites"

Colin B. Fogarty[1]*, Michael P. Fay[2], Jennifer A. Flegg[3] [4], Kasia Stepniewska[3] [5], Rick M. Fairhurst[2],

and Dylan S. Small[1]

[1]Department of Statistics, The Wharton School, University of Pennsylvania,

Philadelphia, Pennsylvania 19104, USA

[2] National Institute of Allergy and Infectious Diseases, National Institutes of Health

Bethesda, Maryland 20892, USA

[3]WorldWide Antimalarial Resistance Network (WWARN) and Centre for Tropical Medicine,

Nuffield Department of Clinical Medicine, University of Oxford, Oxford OX3 7LJ, UK

[4] School of Mathematical Sciences and Monash Academy for Cross and Interdisciplinary Mathematical Applications,

Monash University, Melbourne, Australia

[5]Centre for Tropical Medicine, Nuffield Department of Clinical Medicine, University of Oxford,

Oxford OX3 7LJ, UK

*email: cfogarty@wharton.upenn.edu

This paper has been submitted for consideration for publication in *Biometrics*

Web Appendix A

Coverage of Posterior Predictive Intervals

In the first posterior predictive check, we create pointwise 95% posterior predictive ranges for the log parasite profiles for each individual at each of that individual's measurement times. We then assess what fraction of the actually observed log parasite counts fall within these ranges. If this number deviates substantially from 0.95, it would be an indication that our model does not adequately fit the data.

Our data set has a grand total of 1394 uncensored parasite counts. We created posterior predictive intervals at each of these 1394 observations based on our posterior samples. Of these 1394 uncensored parasite counts, 1330 (or 95.4%) of the posterior predictive intervals created with nominal coverage of 95% actually captured the observed data points. This shows that our predictive intervals come very close to attaining their nominal coverage rates, providing assurance that the model based inference presented in this paper is not misguided.

Posterior Predictive Check for Serial Correlation

Our model assumes that our observations are not serially correlated within an individual's decay profile. To test the validity of our assumption with respect to this data set, we perform a posterior predictive check on the cross moments of our residuals, $\mathbb{E}[r_t r_s], t \neq s$ for each observed combination of measurement time, where this expectation is over individuals. After obtaining our posterior predictive distribution, we calculate the residuals from each of our simulated data sets relative to the posterior mean fits, $\mathbf{r}_i = \tilde{\mathbf{y}}_i - (\hat{\mathbf{y}}_i | \hat{\beta}_i, \hat{\alpha}_i, \hat{\delta}_i^{\ell}, \hat{\delta}_i^{\tau})$. We calculate residuals in this manner because this is also how we define the residuals for our observed data, thus ensuring that the predictive data sets and the actual data can be fairly compared.

For each posterior predictive data set and for each individual within those data sets, we calculate cross products of residuals within an individual, with each cross product corresponding to a pair of times $\{s, t\}$. We then look across individuals and group the residuals

into sets corresponding to the same pairs of observation times, and take the average for each of these groups. This results in values $\overline{r_{(t)}r_{(s)}}$ for each combination of time values for each of our data sets. When we consider the values of $\overline{r_{(t)}r_{(s)}}$ across posterior predictive data sets, we are left with a "null" distribution for each pair of time values. We then compute the actual residuals and residual cross moments from our true data, and see where the cross moments from our data set lie relative to their corresponding null distribution.

In our data set, we had 171 distinct pairs of observation times, hence resulting in 171 hypothesis tests. We found that 165 of them failed to reject the null hypothesis of model correctness. Though 6 rejections is not far from what would be expected even if the null were true due to the number of hypothesis tests being performed, four of them occurred in pairs of measurement times that were very close to one another (time pairs $\{0, 12\}, \{0, 18\}, \{6, 24\}$). It is our belief that these correlations arise mainly from our approximations of the transition between lag and decay phases by a piecewise linear function, whereas the true transition is probably somewhat smoother.. This belief is strengthened by the fact that this serial correlation did not persist beyond the beginning of the decay profiles; that is, at later time points once the lag phases have completed (which are in the regime of interest to antimalarial researchers) there is no evidence of serial correlation.

Web Appendix B

Propriety of Posterior Distribution

We want to show that the posterior that results from our prior specification is proper. As we have placed proper priors on the change points and the probabilities of lag and tail phases (a mixture of a point mass with a bounded uniform, and $Beta(1, 1)$ respectively), the joint prior on $\theta_1 = [\{\delta_i^\ell, \delta_i^\tau\}, \pi^\ell, \pi^\tau]$ is proper. We also know that given values for θ_1, the posterior distribution on the rest of the parameters will also be proper, as this is the posterior from

a hierarchical model with priors on the hyperparameters chosen to avoid impropriety (i.e., to avoid the issue described in Hobert and Casella (1996)). Letting $\mathbf{y}$ denote our data and letting $\theta_2 = [\{\alpha_i, \beta_i\}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \sigma_\alpha^2, \sigma_\beta^2, \sigma_\epsilon^2]$, this means that the distribution $\theta_2|\mathbf{y}, \theta_1$ is proper for any $\mathbf{y} \in \mathcal{Y}, \theta_1 \in \Theta_1$. Finally, note that our priors on θ_1 and θ_2 are independent.

To show that the distribution $\theta_1, \theta_2|\mathbf{y}$ is almost always proper, we need to show that $p(\mathbf{y}) = \int_{\Theta_1 \times \Theta_2} f(y|\theta_1, \theta_2)p(\theta_1, \theta_2)\, d\theta < \infty$. We focus on $\int_{\mathcal{Y}} p(\mathbf{y})\, d\mathbf{y}$:

$$
\begin{aligned}
\int_{\mathcal{Y}} p(\mathbf{y})\, d\mathbf{y} &= \int_{\mathcal{Y}} \int_{\Theta_1} \int_{\Theta_2} f(\mathbf{y}|\theta_1, \theta_2)p(\theta_1, \theta_2)\, d\theta_2\, d\theta_1\, d\mathbf{y} \\
&= \int_{\mathcal{Y}} \int_{\Theta_1} \int_{\Theta_2} f(\mathbf{y}|\theta_1, \theta_2)p(\theta_1)p(\theta_2)\, d\theta_2\, d\theta_1\, d\mathbf{y} && \text{Prior Independence of } \theta_1, \theta_2 \\
&= \int_{\Theta_1} p(\theta_1) \int_{\mathcal{Y}} \int_{\Theta_2} f(\mathbf{y}|\theta_1, \theta_2)p(\theta_2)\, d\theta_2\, d\mathbf{y}\, d\theta_1 && \text{Tonelli's Theorem} \\
&= \int_{\Theta_1} p(\theta_1) \int_{\mathcal{Y}} \int_{\Theta_2} f(\theta_2|\mathbf{y}, \theta_1)p(\mathbf{y}|\theta_1)\, d\theta_2\, d\mathbf{y}\, d\theta_1 && \text{Defn of Conditional Density} \\
&= \int_{\Theta_1} p(\theta_1) \int_{\mathcal{Y}} p(\mathbf{y}|\theta_1)\, d\mathbf{y}\, d\theta_1 && \text{Propriety of } \theta_2|\mathbf{y}, \theta_1 \\
&= \int_{\Theta_1} p(\theta_1)\, d\theta_1 \\
&= 1 && \text{Propriety of Prior on } \theta_1
\end{aligned}
$$

Hence, $p(y)$ must be finite except on a set of measure 0, meaning that the posterior distribution is almost always proper

Web Appendix C

Gibbs Sampler

We use a Gibbs Sampler to simulate from our posterior distribution. As shorthand, we use **rest** to denote the set of all parameters and data excluding those being sampled. Let N denote the total number of individuals in our data set, and let n_i denote the number of time observations for individual i. We proceed as follows:

Sample Censored Observations

As discussed in Section 1 of the manuscript, there are individuals whose counts are left censored at a particular observation time due to the data collection process. At this point in the Gibbs sampler, note that we have the vector $[\alpha_i, \beta_i, \delta_i^\ell, \delta_i^\tau]$ for each individual. Hence, we know the expected value of the individual's log parasite count at a censored time t_{ij}.

Define the modified observation time, z_{ij} as

$$z_{ij} = \begin{cases} \delta_i^\ell & t_{ij} < \delta_i^\ell \\ t_{ij} & \delta_i^\ell \leqslant t_{ij} \leqslant \delta_i^\tau \\ \delta_i^\tau & t_{ij} > \delta_i^\tau \end{cases}$$

Sampling the latent true count given the censoring threshold simply requires sampling from a truncated normal, with truncation at the censored value, call it ζ.

$$\log(y_{ij})|\mathbf{rest} \sim \mathcal{N}(\alpha_i - \beta_i z_{ij}, \sigma_\epsilon^2) \times \mathbb{1}\{y_{ij} < \zeta\}$$

Sample $\{\alpha_i, \beta_i\}|\mathbf{rest}$

We begin by sampling the individual level slopes and intercepts for the profile trajectories, given all other information. Note that at this point in the Gibbs sampler we know where the lag and tail phases, and hence can explicitly describe the modified time points for this iteration. Suppose we have n_i modified time points, $\{z_{i1}, ..., z_{in_i}\}$ for individual i.

Since we have placed priors on $\{\log(\alpha_i), \log(\beta_i)\}$ we do not have conjugacy of the posterior distribution. To sample from the joint posterior $\{\alpha_i, \beta_i\}$, we instead use an independent jump Metropolis-Hastings algorithm. As a proposal distribution, we use the posterior distribution that would have been generated if we used a normal prior with first and second moments equal to those of the lognormal. To be more explicit, let α_i^* and β_i^* denote the proposed intercepts and clearance rates. Then, we define our jumping distribution:

$$Q(\alpha_i^*, \beta_i^*|\mathbf{rest}) = f(\mathbf{y}_i|\alpha_i^*, \beta_i^*, \mathbf{rest})\pi^*(\alpha_i^*, \beta_i^*|\mathbf{rest})$$

where $f(\mathbf{y}_i | \alpha_i^*, \beta_i^*, \mathbf{rest})$ is the likelihood for observing the vector of counts $\mathbf{y}_i$:

$$f(\mathbf{y}_i | \alpha_i^*, \beta_i^*, \mathbf{rest}) \propto \exp\left\{-\frac{1}{2\sigma_\epsilon^2} \sum_{j=1}^{n_i} (\log(y_{ij}) - (\alpha_i^* - \beta_i^* z_{ij}))^2\right\}$$

and $\pi^*(\cdot)$ is our normal approximation to the lognormal prior $\pi(\cdot)$:

$$\pi^*(\alpha_i^*, \beta_i^* | \mathbf{rest}) \propto \exp\left\{-\frac{(\alpha_i^* - \exp\{\boldsymbol{\eta}\mathbf{X}_i + 0.5\sigma_\alpha^2\})^2}{2(\exp\{\sigma_\alpha^2\} - 1)\exp\{2\boldsymbol{\eta}\mathbf{X}_i + \sigma_\alpha^2\}}\right\}$$

$$\times \exp\left\{-\frac{(\beta_i^* - \exp\{\boldsymbol{\gamma}\mathbf{X}_i + 0.5\sigma_\beta^2\})^2}{2(\exp\{\sigma_\beta^2\} - 1)\exp\{2\boldsymbol{\gamma}\mathbf{X}_i + \sigma_\beta^2\}}\right\}$$

Note that we've constructed $Q(\alpha_i^*, \beta_i^*)$ such that it is a mulitvariate normal, and that it mimics the effect of the lognormal priors. Once we sample from this jumping distribution, we accept the proposal $\{\alpha_i^*, \beta_i^*\}$ with probability:

$$\begin{aligned}
\mathbb{P}(\mathbf{Accept}) &= \min\left\{1, \frac{p(\alpha_i^*, \beta_i^* | \mathbf{rest})}{p(\alpha_i, \beta_i | \mathbf{rest})} \frac{Q(\alpha_i, \beta_i | \mathbf{rest})}{Q(\alpha_i^*, \beta_i^* | \mathbf{rest})}\right\} \\
&= \min\left\{1, \frac{f(\mathbf{y}_i | \alpha_i^*, \beta_i^*, \mathbf{rest})\pi(\alpha_i^*, \beta_i^* | \mathbf{rest})}{f(\mathbf{y}_i | \alpha_i, \beta_i, \mathbf{rest})\pi(\alpha_i, \beta_i | \mathbf{rest})}\right. \\
&\quad \times \left.\frac{f(\mathbf{y}_i | \alpha_i, \beta_i, \mathbf{rest})\pi^*(\alpha_i, \beta_i | \mathbf{rest})}{f(\mathbf{y}_i | \alpha_i^*, \beta_i^*, \mathbf{rest})\pi^*(\alpha_i^*, \beta_i^* | \mathbf{rest})}\right\} \\
&= \min\left\{1, \frac{\pi(\alpha_i^*, \beta_i^* | \mathbf{rest})}{\pi(\alpha_i, \beta_i | \mathbf{rest})} \frac{\pi^*(\alpha_i, \beta_i | \mathbf{rest})}{\pi^*(\alpha_i^*, \beta_i^* | \mathbf{rest})}\right\}
\end{aligned}$$

We found that this step was highly efficient in practice, and required minimal thinning to result in nearly iid draws.

Sample $\{\delta_i^\ell, \delta_i^\tau\} | \mathbf{rest}$

Let ℓ_i denote the index of the last point in time during a lag phase for the ith individual, and let τ_i denote the index of the last time point during a decay phase (that is, the time point before a tail phase begins) for that individual. To sample the change point between lag and decay phase for each individual, we begin by considering the posterior measure of the collection of sets $\mathcal{A}_i = \{\{0\}, (0, t_{i1}], (t_{i1}, t_{i2}], ..., (t_{i,\tau_i}, \delta_i^\tau]\}$ via numerical integration for each individual. Explicitly, let $g(\mathbf{y} | x, \delta_i^\tau, \mathbf{rest})$ denote the likelihood of the observed counts for individual i given $\delta_i^\ell = x$ and given the other parameters (denoted by $\mathbf{rest}$). Then, we

calculate the following for each $A_{ik} \in \mathcal{A}_i$

$$I(A_{ik}) = \begin{cases} (1 - \pi^\ell)g(\mathbf{y}_i|0, \delta_i^\tau, \mathbf{rest}) & A_{ik} = \{0\} \\ \\ \pi^\ell \int_{x \in A_{ik}} g(\mathbf{y}_i|x, \delta_i^\tau, \mathbf{rest})p_\ell(x|\delta_i^\tau, \mathbf{rest}) \, dx & A_{ik} \neq \{0\} \end{cases}$$

Given our priors specifications, $p_\ell(x|\delta_i^\tau)$ takes on the following form:

$$p_\ell(x|\delta_i^\tau, \mathbf{rest}) = \frac{1}{x} \frac{\phi((\log(x) - a)/c)}{\Phi((\log(\delta_i^\tau) - a)/c)} \mathbb{1}\{x < \delta_i^\tau\},$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the PDF and CDF respectively of a standard normal. $I(\cdot)$ thus represents the posterior measure of each set (up to a constant that is the same for each set) given the other parameters. After normalizing the integrals to get probabilities, we then select one of these intervals with probability equal to the respective measure. If the singleton zero is selected, we are done; otherwise, we use rejection sampling within the sampled interval to arrive at our change point with a uniform proposal distribution.

For this rejection sampling, let $[a, b]$ be the interval selected above. The sampling proceeds as follows:

(1) Find $M = \max_{x \in [a,b]} g(\mathbf{y}_i|x, \delta_i^\tau, \mathbf{rest})p_\ell(x|\delta_i^\tau, \mathbf{rest})$

(2) Draw $u \sim \mathcal{U}[0, 1]$

(3) Draw $v \sim \mathcal{U}[a, b]$

(4) Set $\delta_i^\ell = v$ if $u \leqslant g(\mathbf{y}_i|v, \delta_i^\tau, \mathbf{rest})p_\ell(v|\delta_i^\tau, \mathbf{rest})/M$; else, return to step (2)

Sampling the change point between the decay and tail phase is very similar. We first consider the posterior measure of the sets $\{(\delta_i^\ell, t_{i,\ell_i+1}], (t_{i,\ell_i+1}, t_{i,\ell_i+2}], ..., [t_{i,n_i-1}, t_{i,n_i}), \{t_{i,n_i}\}\}$. If we sample the right endpoint, we are done; otherwise, we again use rejection sampling.

*Sample $a, b, c^2, d^2|$***rest**

Let $\mathcal{L} = \{i : \delta_i^\ell \neq 0\}$, and let $\mathcal{T} = \{i : \delta_i^\tau \neq t_{in_i}\}$. We then use an independent Metropolis hastings algorithm to sample a, b, c^2, d^2. We demonstrate it here with a and c^2. To begin,

given a value of c^2 we propose a value of a as follows:

$$a^*|c^2, \mathbf{rest} \sim \mathcal{N}\left((|\mathcal{L}|/c^2 + 1/0.25)^{-1}(\sum_{i \in \mathcal{L}} \log(\delta_i^\ell)/c^2 + \log(6)/0.25), (|\mathcal{L}|/c^2 + 1/0.25)^{-1}\right)$$

If we did not have the constraint that $\delta_i^\ell < \delta_i^\tau \; \forall i$, the sampling would simply be a Gibbs sampler and hence the above proposal would be accepted with probability 1. To account for the truncation, we accept the above with the following probability:

$$\mathbb{P}(\mathbf{Accept}) = \min\left\{1, \prod_{i \in \mathcal{L}} \frac{\Phi((\delta_i^\tau - a)/c)}{\Phi((\delta_i^\tau - a^*)/c)}\right\}$$

This probability ends up being extremely close to one, as lag and tail phases tend to occur far from each other.

Now, given a value of a, we propose c^2 as follows:

$$(c^2)^*|a, \mathbf{rest} \sim IG\left(\frac{|\mathcal{L}|}{2}, \frac{\sum_{i \in \mathcal{L}}(\log(\delta_i^\ell) - a)^2}{2} + 1\right)$$

This proposal is accepted with the following probability:

$$\mathbb{P}(\mathbf{Accept}) = \min\left\{1, \prod_{i \in \mathcal{L}} \frac{\Phi((\delta_i^\tau - a)/c)}{\Phi((\delta_i^\tau - a)/c^*)}\right\}$$

The sampling of b and d^2 follows in an analogous fashion.

Sample $\{\boldsymbol{\gamma}, \boldsymbol{\eta}\}|\mathbf{rest}$

To sample $\boldsymbol{\gamma}$, note that we have conjugacy and that our hyperprior on $\boldsymbol{\gamma}$ is flat. Let $\tilde{\beta}$ denote the vector of clearance rates $[\beta_1, ..., \beta_N]$. Our Gibbs step is as follows:

$$\boldsymbol{\gamma}|\mathbf{rest} \sim \mathcal{N}\left((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \log(\tilde{\beta}), \sigma_\beta^2(\mathbf{X}^T\mathbf{X})^{-1}\right)$$

Similar argument leads to the following sampler for $\boldsymbol{\eta}$:

$$\boldsymbol{\eta}|\mathbf{rest} \sim \mathcal{N}\left((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \log(\tilde{\alpha}), \sigma_\alpha^2(\mathbf{X}^T\mathbf{X})^{-1}\right)$$

Sample $\{\sigma_\alpha^2, \sigma_\beta^2, \sigma_\epsilon^2\}|\mathbf{rest}$

Sampling the variances σ_α^2 and σ_β^2 is simple, as our priors on these parameters result in conjugate posteriors:

$$\sigma_\alpha^2|\mathbf{rest} \sim IG\left(\frac{N}{2}-1, \frac{1}{2}||\log(\tilde{\alpha}) - \boldsymbol{\eta}\mathbf{X}||_2^2\right)$$

$$\sigma_\beta^2|\mathbf{rest} \sim IG\left(\frac{N}{2}-1, \frac{1}{2}||\log(\tilde{\beta}) - \boldsymbol{\gamma}\mathbf{X}||_2^2\right)$$

Sampling σ_ϵ^2 is also straightforward for the same reason:

$$\sigma_\epsilon^2|\mathbf{rest} \sim IG\left(\frac{\sum_{i=1}^N n_i}{2}, \frac{1}{2}\sum_{i=1}^N\sum_{j=1}^{n_i}(\log(y_{ij}) - (\alpha_i - \beta_i z_{ij}))^2\right)$$

Sample $\{\pi^\ell, \pi^\tau\}|\mathbf{rest}$

To sample the probabilities of having a lag phase and a tail phase, let $L = \sum_{i=1}^N \mathbb{1}\{\delta_i^\ell \neq 0\}$ and $T = \sum_{i=1}^N \mathbb{1}\{\delta_i^\tau \neq t_{i,n_i}\}$. These simply count the number of profiles that had lag and tail phases. Then, our posterior distributions are as follows:

$$\pi^\ell|\mathbf{rest} \sim Beta(1 + L, 1 + N - L)$$

$$\pi^\tau|\mathbf{rest} \sim Beta(1 + T, 1 + N - T)$$

Web Appendix D

Bias Due to Sparse Number of Measurement Times

We first recreate the simulation study conducted in Section 3.1 of the manuscript with $n = 20$ individuals instead of $n = 60$ individuals, but with all other aspects of the simulation study unchanged. Recall that the true slope coefficient for the covariate x_i should be 0, as x_i was simply random noise.

When using the standard two stage analysis, the average estimated slope coefficient for x_i was 0.759 with a standard deviation of 0.502, demonstrating a concentration around a spuriously positive value. The estimated mean squared error for the two-stage method in

this simulation is 0.83. For each simulation, we also computed 95% confidence intervals for the true slope using a t interval based on the two-stage method. We found that of the 1000 confidence intervals constructed, 69.6% actually captured the true slope of 0. Hence, these intervals based on the two-stage procedure fail to match their coverage guarantee.

Within each iteration, we also used the Bayesian method to calculate clearance rates and estimate the slope coefficient for a regression of x_i on the clearance rates. In the 1000 simulations, the average posterior mean for the slope coefficient was -0.125 with a standard deviation of 0.72, resulting in a mean squared error of 0.59. This represents a reduction in mean squared error by 29% relative to the two-stage procedure. We also computed 95% credible intervals for this slope coefficient in order to assess their frequentist coverage properties. We found that of the 1000 intervals constructed, 96% captured the true mean of 0. Hence, these credible intervals have much better coverage properties than the confidence intervals resulting from the two-stage method. Hence, not only does the Bayesian procedure produce shorter intervals, but it also produces intervals with valid coverage due to the lack of bias in slope estimation.

Behavior of Estimators with a Sufficient Number of Measurement Times

We now reproduce the simulation studies of Section 3.2 of the manuscript with $n = 20$ instead of $n = 60$, but with all other aspects left unchanged. Table 1 presents the results of our simulation studies with reduced sample size. We found that in all three scenarios, the Bayesian estimator outperformed the standard two-stage estimator in terms of bias, mean squared error, coverage of intervals, and length of intervals. The mean squared errors for the slope estimate produced by the two-stage procedure were 0.00527, 0.00632, and 0.0244 for Simulation scenarios 1, 2, and 3. For the Bayesian procedure, the mean squared errors in these three scenarios were 0.00429, 0.00469, and 0.00813. Once again, the performances are

very similar when lag and tail phases are absent; however, when the they are present, the Bayesian procedure provides an improvement.

[Table 1 about here.]

Web Appendix E

Comparing Uniform and Log-Normal Priors on Changepoints

As was described in Section 2.2, the general form of our prior distribution on the changepoints is a mixture of a point mass (at time 0 for lag phases and time t_{in_i} for tail phases) and a continuous distribution on the feasible points. Although uniform priors are commonly employed in changepoint problems, we decided to use the log-normal distribution for the continuous portion. Hierarchical log-normal priors allow for different hazard rates at different time values and sharing of information between the observed clearance profiles while restricting the changepoints to take on positive values. The log-normal distribution also has its mode away from the boundary and can accommodate a long right tail, which is often found in studying time to event quantities.

To assess the impact of this choice, we conducted the simulation studies of Section 3.2 using both the uniform and log-normal priors. In addition, we repeated the data analysis of Section 4 under both distributions.

Simulation Studies

Table 2 presents the results of the simulation studies of Section 3.2 under both the log-normal and uniform distributions. We found that when lag and tail phases were truly absent, the uniform prior performed better. This is in part due to the fact that the uniform prior appears to induce "sparser" fits than does the log-normal. Many trajectories that the uniform identifies as having no lag or tail phase are identified by the log-normal as having extremely short (but nonzero) lag and tail phases. On the other hand, the log-normal specification

outperforms the uniform when lag and tail phases are present. Some trajectories which the log-normal correctly identified as having a short lag or tail phase were instead identified as having no lag phase. The log-normal specification also allows for shrinkage estimation, which can be beneficial in reducing loss in estimation.

[Table 2 about here.]

Data Analysis

We analyzed the data set described in Section 4 using both the uniform and log-normal priors. For this data set the estimated slopes of the covariates of interest on the log half-lives were virtually identical, as is shown in Table 3. By looking at the individual fits, we are able to shed further light onto the relative merits of the log-normal prior versus the uniform prior on the changepoints.

[Table 3 about here.]

Figure 1 shows the fits for an individual that appears to have a small lag phase. The log-normal prior identifies this apparent lag-phase in the vast majority of posterior samples, whereas the uniform identifies that no lag phase is present in over 50% of posterior samples. The log-normal prior resulted in a median clearance rate for this individual of 0.162, with a 95% credible interval of [0.141, 0.187]. The uniform prior resulted in a median clearance rate for this individual of 0.152, with a 95% credible interval of [0.135, 0.180].

Figure 2 shows the fit for an individual that appears to only have a decay phase. The uniform prior identifies the lack of lag phase in the vast majority of samples, whereas the log-normal prior estimates a median changepoint between lag and decay phases of 0.61 hours. We found that the identification of a short lag phase by the log-normal resulted in virtually no difference in the posterior distribution of this individual's clearance rate. The log-normal prior resulted in a median clearance rate for this individual of 0.107, with a 95% credible

interval of [0.0961, 0.120]. The uniform prior resulted in a median clearance rate for this individual of 0.106, with a 95% credible interval of [0.0946, 0.119].

[Figure 1 about here.]

[Figure 2 about here.]

Web Appendix F

To assure ourselves that we have reached stationarity and have thinned sufficiently, we examined trace plots and autocorrrelation plots both before and after burn in and thinning.

[Figure 3 about here.]

References

Hobert, J. P. and Casella, G. (1996). The effect of improper priors on gibbs sampling in hierarchical linear mixed models. *Journal of the American Statistical Association* **91,** 1461–1473.

Smith, A. F. and Gelfand, A. E. (1992). Bayesian statistics without tears: a sampling–resampling perspective. *The American Statistician* **46,** 84–88.

Stephens, D. (1994). Bayesian Retrospective Multiple-Changepoint Identification. *Journal of the Royal Statistical Society, Series C* **43,** 159–178.
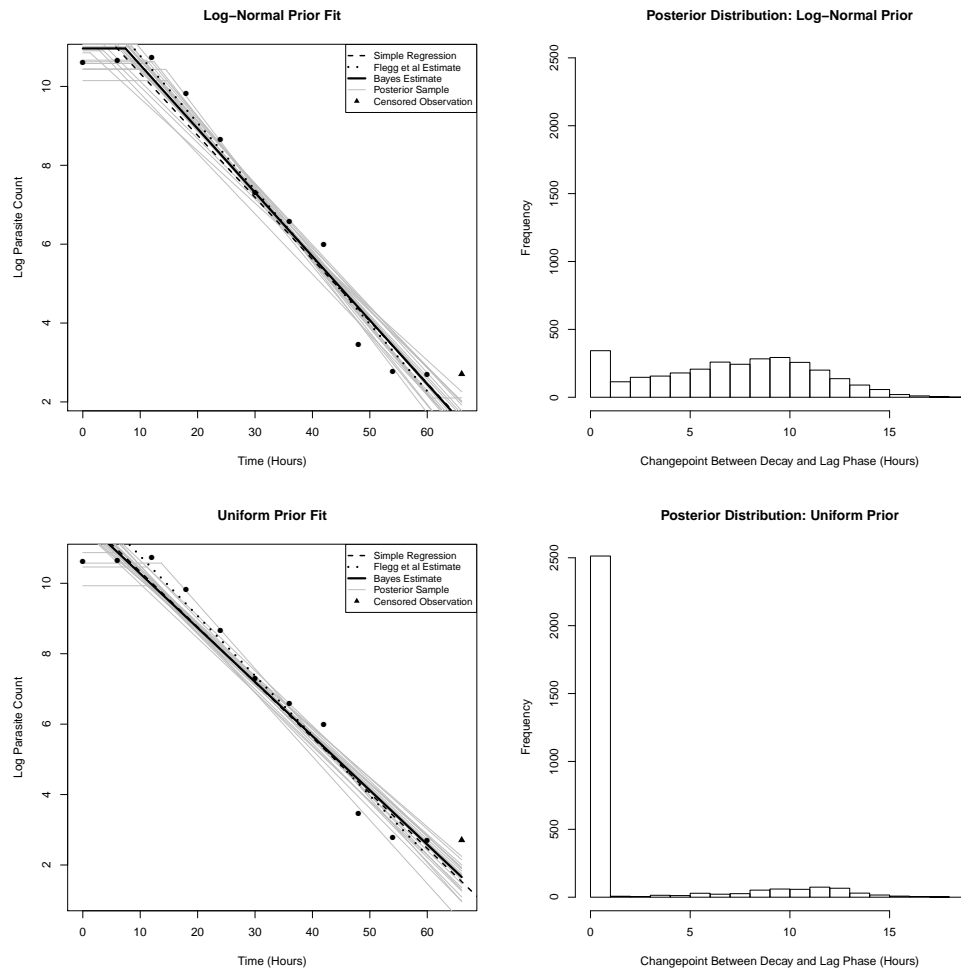
Figure 1. Posterior log parasitemia profile and posterior distribution of changepoint between lag and decay phase for an individual using a log-normal (top) and uniform (bottom) prior for the changepoints between phases.

Figure 2. Posterior log parasitemia profile and posterior distribution of changepoint between lag and decay phase for an individual using a log-normal (top) and uniform (bottom) prior for the changepoints between phases.
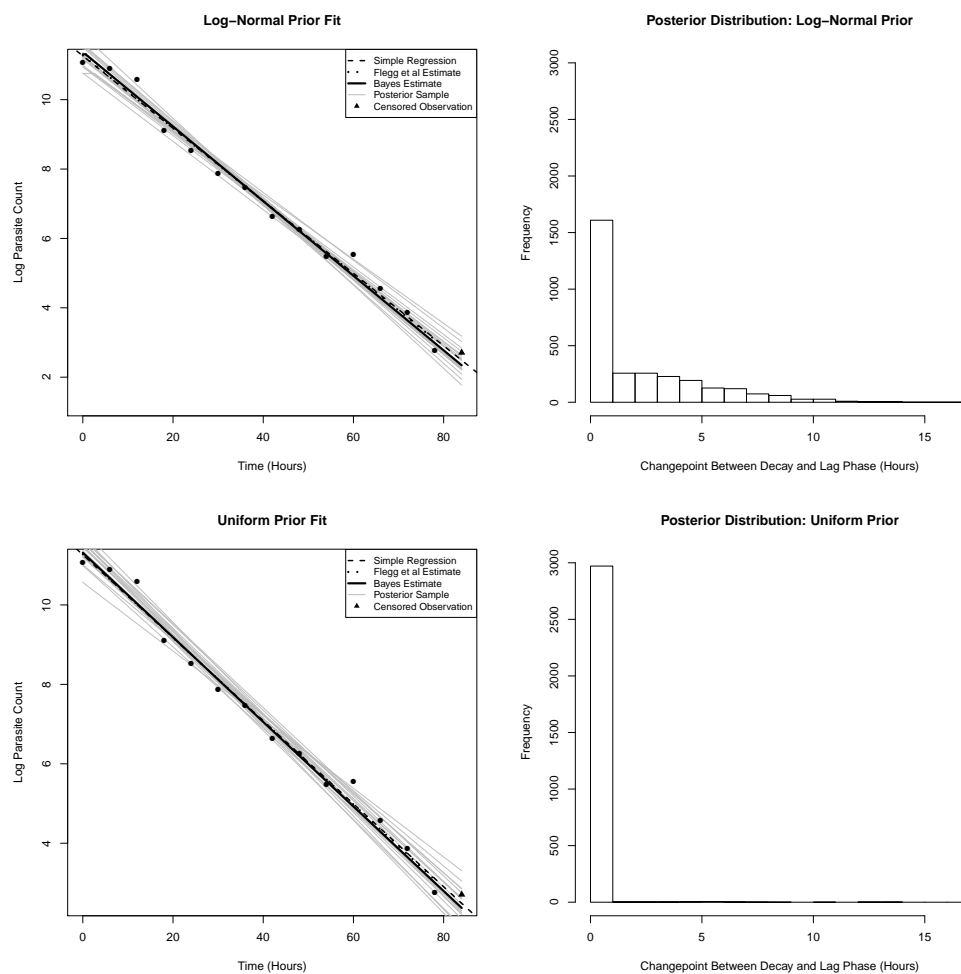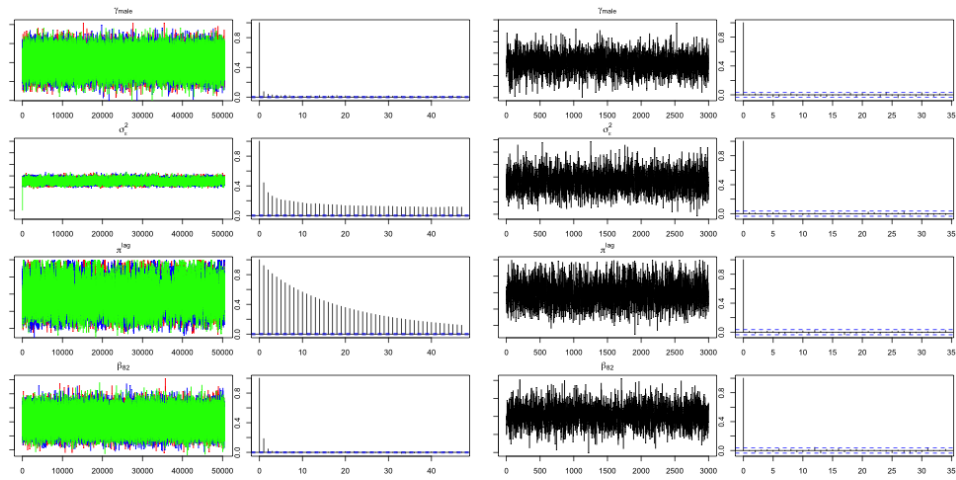
Figure 3. Diagnostic plots of values over time and ACF plots before burn in and thinning (**left**) and after burn in and thinning (**right**) for select domain points. The traceplots before burn in and thinning display three different chains with different starting values overlaid on top of one another.

Table 1
Results from the simulation studies of manuscript Section 3.2 with $n = 20$. In all three simulation studies, the true value of γ_2, the average log difference in clearance rates between the two simulated populations, was $\log(0.25/0.15) \approx 0.511$. In each simulation study, 1000 data sets of size $n = 20$ were drawn according to that simulation's data generative model. From left the right, the columns represent the average value and standard deviation of the two-stage estimator across simulations; the average value and standard deviation of the Bayesian estimator across simulations; the percentage reduction in average loss (under squared error loss) attained by using the Bayesian estimator; the proportion of 95% intervals from the two-stage method that capture the true difference in means and length of those intervals; and the proportion of 95% intervals from the Bayesian method that capture the true difference in means and the length of those intervals

	$\hat{\mathbb{E}}[\hat{\gamma}_2^{TS}]$ (SD)	$\hat{\mathbb{E}}[\hat{\gamma}_2^{B}]$ (SD)	% Red in Avg. Loss Using Bayes	$\hat{\mathbb{P}}[\gamma_2 \in CI_{95\%}^{TS}]$ (Length)	$\hat{\mathbb{P}}[\gamma_2 \in CI_{95\%}^{B}]$ (Length)
Simulation 1 $\pi^\ell = \pi^\tau = 0$	0.526 (0.071)	0.529 (0.063)	18.4%	95.8% (0.289)	96% (0.287)
Simulation 2 $\pi^\ell = 0.5,\ \pi^\tau = 0$	0.491 (0.077)	0.525 (0.067)	25.7%	94.2% (0.323)	96.0% (0.314)
Simulation 3 $\pi^\ell = \pi^\tau = 0.25$	0.561 (0.148)	0.538 (0.086)	66.7%	94.2% (0.628)	95.6% (0.375)

Table 2

Results from the simulation studies of Section 3.2 under log-normal and uniform priors on changepoints. In all three simulation studies, the true value of γ, the average log difference in clearance rates between the two simulated populations, was $\gamma = \log(5/3) \approx 0.511$. Simulation studies were conducted with $n = 60$ individuals.

	$\hat{\mathbb{E}}[\hat{\gamma}_2^{LN}]$ (SD)	$\hat{\mathbb{E}}[\hat{\gamma}_2^{U}]$ (SD)	MSE_{LN}/MSE_U	$\hat{\mathbb{P}}[\gamma_2 \in CI_{95\%}(LN)]$ (Length)	$\hat{\mathbb{P}}[\gamma_2 \in CI_{95\%}(U)]$ (Length)
Simulation 1 $\pi^\ell = \pi^\tau = 0$	0.523 (0.037)	0.515 (0.037)	1.10	95.2 (0.145)	94.8% (0.145)
Simulation 2 $\pi^\ell = 0.5, \pi^\tau = 0$	0.515 (0.038)	0.520 (0.041)	0.83	94.0% (0.152)	94.1% (0.169)
Simulation 3 $\pi^\ell = \pi^\tau = 0.25$	0.523 (0.041)	0.525 (0.041)	0.97	94.1% (0.167)	93.5% (0.170)

Table 3
Estimates and Intervals for the Effect of Covariates on Log Half-Lives

	Log-Normal Prior			Uniform Prior		
	Estimate	2.5%	97.5%	Estimate	2.5%	97.5%
Intercept	1.58	0.823	2.33	1.58	0.815	2.38
Sex (1 if Male)	0.191	0.01	0.370	0.185	-0.001	0.369
Age Group (1 if < 21)	-0.015	-0.145	0.116	-0.011	-0.14	0.115
Kravanh or Veal Veng (1 if Yes)	-0.007	-0.134	0.127	-0.008	-0.138	0.122
Hemoglobin E (1 if Yes)	0.108	-0.004	0.219	0.104	-0.004	0.207
α-thalassaemia (1 if Yes)	-0.066	-0.200	0.063	-0.064	-0.194	0.070
G6PD Deficient (1 if Yes)	-0.001	-0.090	0.087	-0.001	-0.088	0.084
Log Initial Parasite Density	0.021	-0.054	0.091	0.021	-0.053	0.092
Year (1 if 2010)	0.044	-0.086	0.172	-0.072	0.179	0.095
Parasite Group (1 if Group 1)	0.150	0.028	0.280	0.140	0.0134	0.267