Supplemental Materials for "Amount of brain edema correlates with neurologic recovery in pediatric cerebral malaria"

1 Model, MCMC, and Diagnostics

We consider a joint model of three outcomes, NSQ status (1 for developing NSQ; 0 otherwise), CSF measurements in log scale, and time in coma. We connect the three outcomes via a latent variable U_i for each patient i and consider the following model:

$$\begin{aligned}
\text{logit}(P(NSQ = 1)) &= \alpha_0 + \alpha_1 \times U_i \\
\log(CSF) &\sim \text{Normal}(U_i + \beta_0 + (\beta_1 + \kappa \times U_i) \times \text{time}, \sigma_e^2) \\
\log(\text{Time in coma}) &= \text{Normal}(\lambda_0 + \lambda_1 \times U_i, \sigma_c^2) \\
U_i &\sim \text{Normal}(0, \sigma_u^2)
\end{aligned}$$

Note in our model, the positive latent variable U_i corresponds to a tendency to have higher CSF measurement at baseline $t = 0$. It is difficult to directly obtain the maximum-likelihood estimator for model parameters; instead, we consider using a Bayesian approach. For each of the model parameters, we put a very weak, almost non- informative prior. Specifically, we have the following:

- $\sigma_u^2 \sim$ Inverse Gamma(0.01, 0.01);
- $\sigma_e^2 \sim$ Inverse Gamma(0.01, 0.01);
- $\sigma_c^2 \sim$ Inverse Gamma(0.01, 0.01);
- $\alpha_0 \sim$ Normal(0, 100);
- $\alpha_1 \sim$ Normal(0, 100);
- $\lambda_0 \sim$ Normal(0, 100);
- $\lambda_1 \sim$ Normal(0, 100);
- $\kappa \sim$ Normal(0, 100);
- $\beta_0 \sim$ Normal(0, 20);
- $\beta_1 \sim$ Normal(0, 20);

To take into account of the case-control study design, we weight each treated and control subject by (1/probability of being included in the study). We run four chains for a long time (20000 iterations for warmup and 22000 in total) and look at the mixing of these four chains. The particular diagnostic statistic of interest is the potential scale reduction factor on split chains $\hat{R}$. Note this test statistic is 1 at convergence. According to the recommendation in Gelman et al.(2014), it is generally satisfying with setting $\hat{R} = 1.1$ as a threshold. Figure 1 shows a histogram of the $\hat{R}$ statistics for all parameters (including all the random effects). Note the majority of $\hat{R}$ values is very close to 1, indicating the four chains mix very well. We can also look at the mixing of the chains more directly via the traceplots. See Figure 2. From these convergence

diagnostics, we are confident the chains have reached a stationary distribution and we can conduct inference based on the samples from the posterior.

Figure 1: Histogram of the potential scale reduction factor on split chains $\hat{R}$

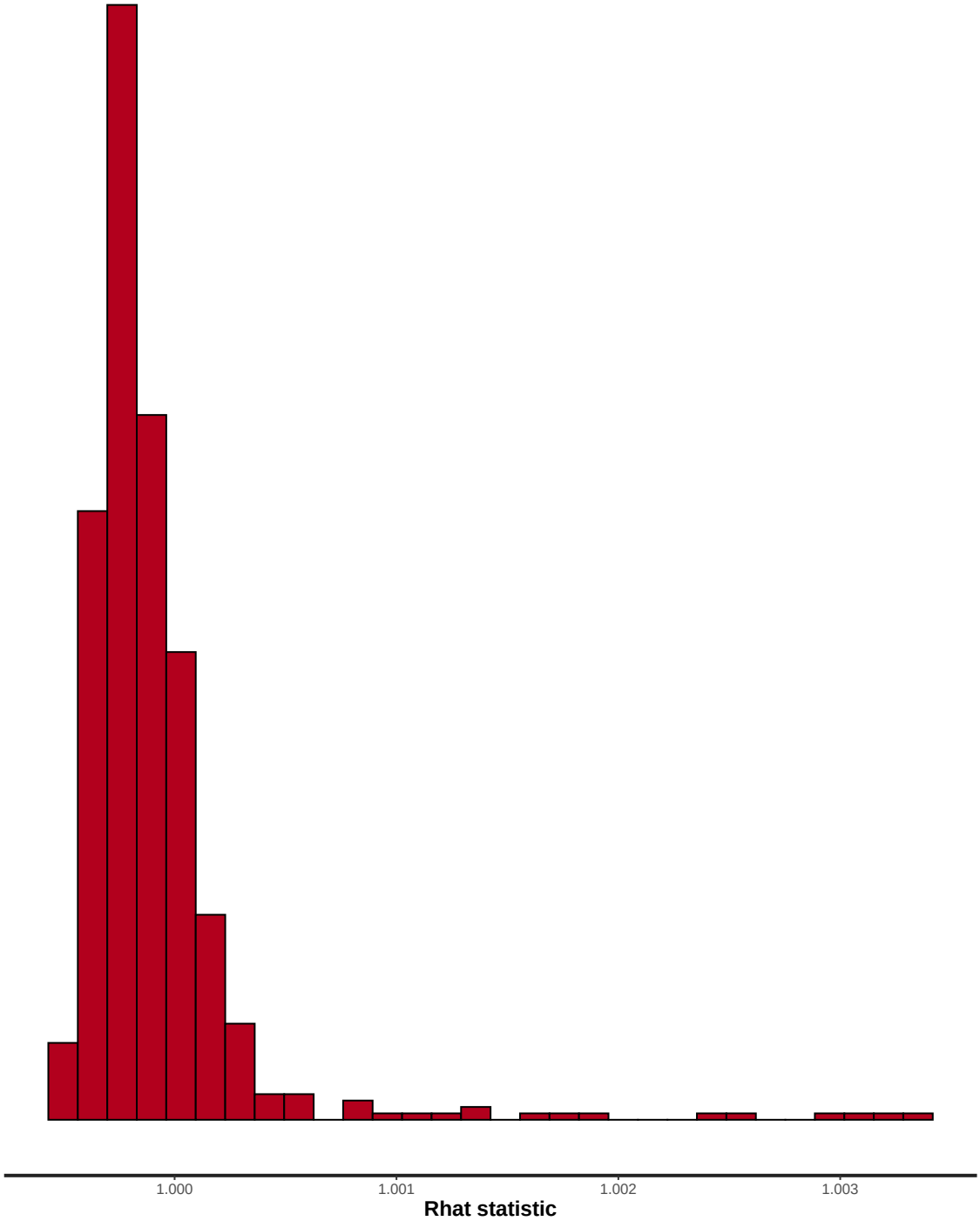
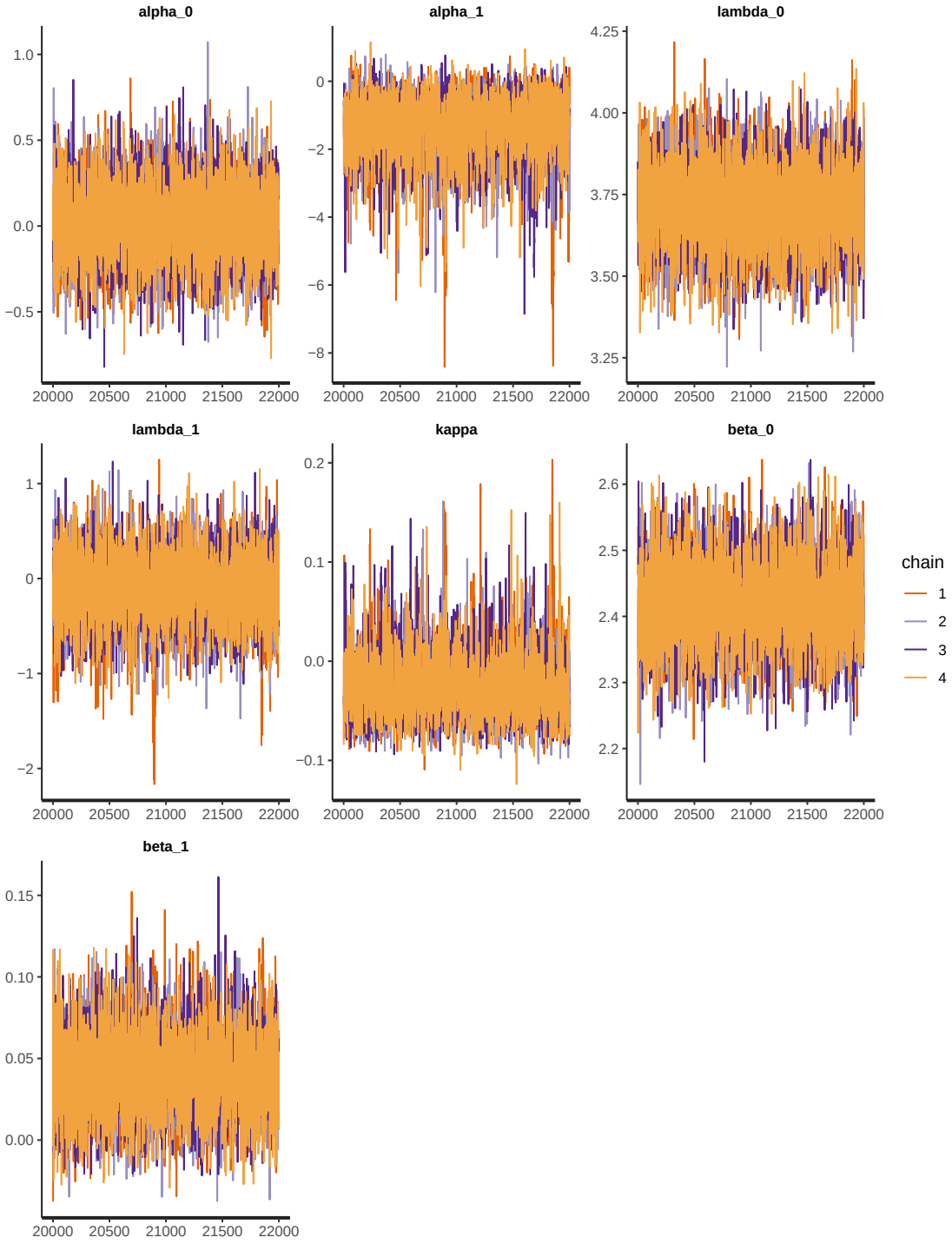


Figure 2: Traceplot for parameters of interest



2 Inference

To conduct inference, we plot the posterior distributions of parameters of interest. See Figure 3 below. Table 2 tabulates the 95% equal-tailed posterior credibility interval for each parameter. Of particular interest is α_1. Note its 95% posterior credibility interval is $[-3.59, 0.01]$.

The implication is that a larger latent variable U_i corresponds to a smaller probability that a patient develops NSQ. On the other hand, a larger U_i corresponds to a larger CSF measurement in log scale at baseline $t = 0$.

Figure 3: Posterior distributions of parameters of primary interest

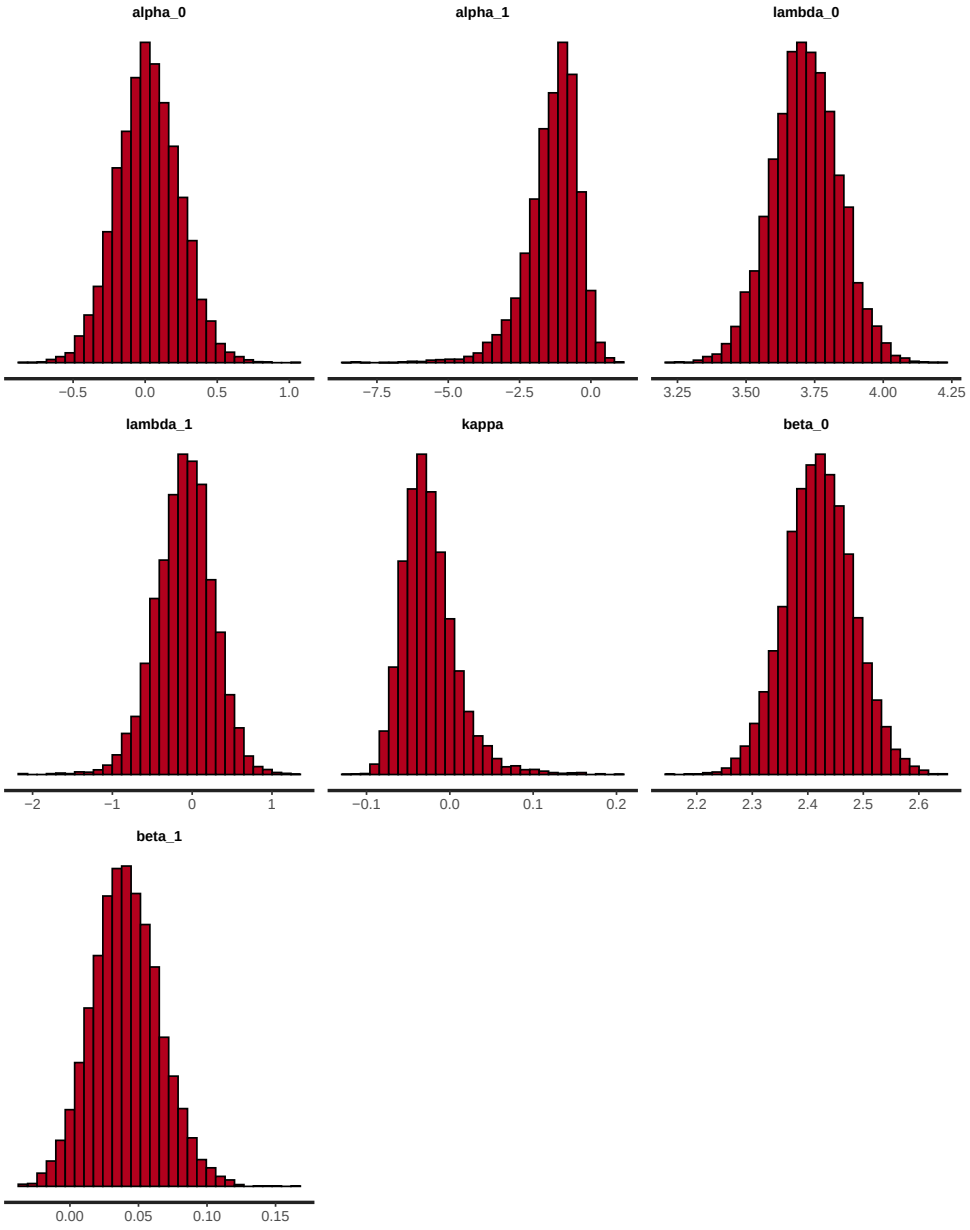


Table 1: Summary statistics for the posterior distribution

	mean	sd	2.5%	97.5%
α_0	0.02	0.22	-0.40	0.45
α_1	-1.35	0.93	-3.59	0.01
λ_0	3.72	0.12	3.48	3.96
λ_1	-0.10	0.36	-0.83	0.57
κ	-0.02	0.03	-0.07	0.05
β_0	2.42	0.06	2.30	2.54
β_1	0.04	0.02	-0.01	0.09

3 Prediction

To help prognosis, we want to leverage our model in the following way: given a patient still in coma at time t^* (log scale), and his/her log(CSF) measurements at certain timepoints, e.g. $t = 0$ and $t = 1$, we would like to leverage our model and make predictions of the probability that the patient develops NSQ. To compute the conditional distribution $P(NSQ = 1 \mid t_{\text{coma}} > t^*, \log(\text{CSF})_{t=0} = C_0, ..., \log(\text{CSF})_{t=k} = C_k)$, it is essential to first draw samples from the posterior distribution $P(U \mid t_{\text{coma}} > t^*, \log(\text{CSF})_{t=0} = C_0, ..., \log(\text{CSF})_{t=k} = C_k))$.

Note we can compute the conditional probability

$$\begin{aligned} &P(t_{\text{coma}} > t^*,\ \log(\text{CSF})_{t=0} = C_0, ...,\ \log(\text{CSF})_{t=1} = C_1 \mid U) \\ =&P(t_{\text{coma}} > t^* \mid U) \times P(\log(\text{CSF})_{t=0} = C_0 \mid U) \times \cdots \times P(\log(\text{CSF})_{t=k} = C_k \mid U) \end{aligned}$$

as follows:

- Draw samples from the posterior of λ_0, λ_1, and σ_c^2. Simulate t_{coma} according to the log-normal model and approximate $P(t_{\text{coma}} > t^* | U)$;
- Draw samples from the posterior of β_0 and σ_e^2. Calculate $P(\log(\text{CSF})_{t=i} = C_i | U = u)$ by averaging over the randomness in β_0 and σ_e^2;
- Obtain the conditional probability $P(t_{\text{coma}} > t^* \mid U) \cdot \prod_{i=0}^{k} P(\log(\text{CSF})_{t=i} = C_i \mid U)$

On the other hand, we can compute $P(U = u)$ by averaging over the randomness in σ_u^2. Finally, we can compute the unnormalized posterior probability by

$$p(u) = P(t_{\text{coma}} > t^*,\ \log(\text{CSF})_{t=0} = C_0, ...,\ \log(\text{CSF})_{t=k} = C_k \mid U = u) \cdot P(U = u).$$

Since U is only one-dimensional, we can first compute the unnormalzied posterior probability on a fine grid and then normalize. Once we have the posterior distribution for U, we can compute the posterior distribution of $P(NSQ = 1 \mid t_{\text{coma}} > t^*,\ \log(\text{CSF})_{t=0} = C_0, ...,\ \log(\text{CSF})_{t=1} = C_1 \mid U)$ by averaging over the randomness in U, α_0, and α_1.

For each patient who is still in coma at time t^* (number of hours in the log scale) and a history of log(CSF) measurements: $(t_0, c_0), (t_1, c_1), ..., (t_k, c_k)$ (t_i is in the unit of day and c_i is in log scale), we can produce the posterior distribution of the probability that this patient develops NSQ. This allows a doctor to monitor the patient's status in real time and help best relocate the precious resources in cerebral malaria-endemic regions.

4 Illustration

To illustrate, suppose a patient is still in coma at time $t^* = 55$ hours, and we have made two measurements, one at baseline (t = 0) with log(CSF) = 2, the other measurement at $t = 24$ hours with log(CSF) = 3. For such a patient, we run our algorithm and predict in real time his probability of developing NSQ. See Figure 4a for the posterior distribution of this probability. The posterior mean is 0.494 with a 95% equal-tailed

credibility interval being $[0.264, 0.714]$. To draw a contrast, suppose another patient, who is also in coma at time $t^* = 33$ hours with two log(CSF) measurements equal to 2.5 and 1.5 at $t = 0$ and $t = 1$ respectively. For such a patient, Figure 4b shows the posterior distribution of his/her probability of developing NSQ. The posterior mean is 0.596 for this patient with a 95% equal-tailed credibility interval being $[0.395, 0.853]$. In practice, a doctor may look at these two graphics and pay more attention and possibly allocate more resources towards the first patient.

5 Sensitivity Analysis

In this section, we perform a sensitivity analysis by further incorporating age and gender in the model:

$$\begin{gathered}
\text{logit}(P(NSQ = 1)) = \alpha_0 + \alpha_a \times \text{age}_i + \alpha_g \times \text{gender}_i + \alpha_1 \times U_i \\
\log(CSF) \sim \text{Normal}(U_i + \beta_0 + \beta_a \times \text{age}_i + \beta_g \times \text{gender}_i + (\beta_1 + \kappa \times U_i) * \text{time}, \sigma_e^2) \\
\log(\text{Time in coma}) = \text{Normal}(\lambda_0 + \lambda_a \times \text{age}_i + \lambda_g \times \text{gender}_i + \lambda_1 \times U_i, \sigma_c^2) \\
U_i \sim \text{Normal}(0, \sigma_u^2)
\end{gathered}$$

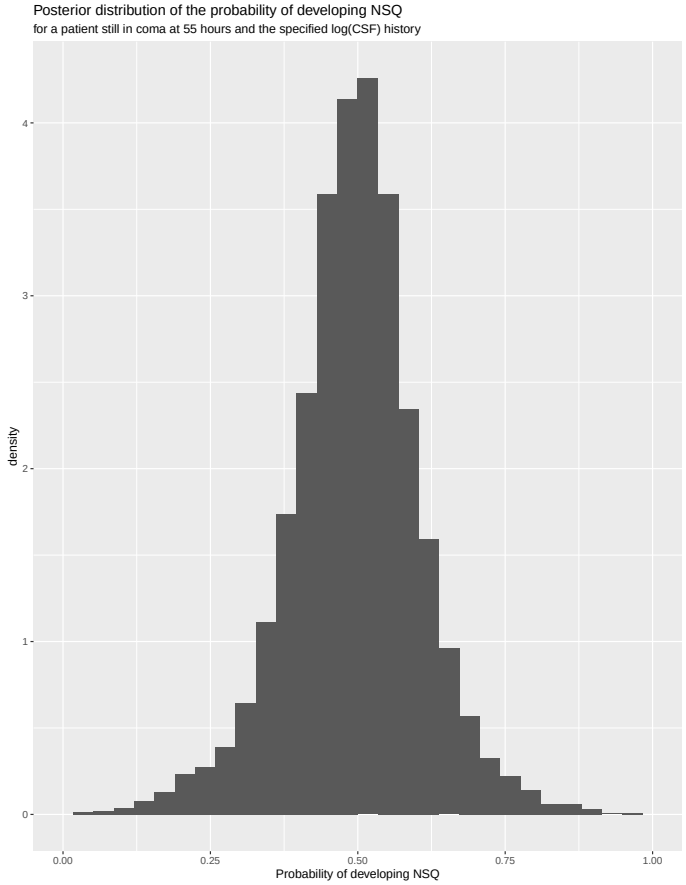
Table 2: Summary statistics for the posterior distribution of some key parameters of interest: before and after age and gender are adjusted for.

	Before				After			
	Mean	SD	2.5%	97.5%	Mean	SD	2.5%	97.5%
α_0	0.02	0.22	-0.40	0.45	0.38	0.49	-0.55	1.37
α_1	-1.35	0.93	-3.59	0.01	-1.20	0.98	-3.41	0.21
λ_0	3.72	0.12	3.48	3.96	3.76	0.27	3.24	4.30
λ_1	-0.10	0.36	-0.83	0.57	-0.04	0.37	-0.81	0.67
κ	-0.02	0.03	-0.07	0.05	-0.03	0.03	-0.07	0.05
β_0	2.42	0.06	2.30	2.54	2.16	0.13	1.90	2.43
β_1	0.04	0.02	-0.01	0.09	0.04	0.02	0.00	0.09

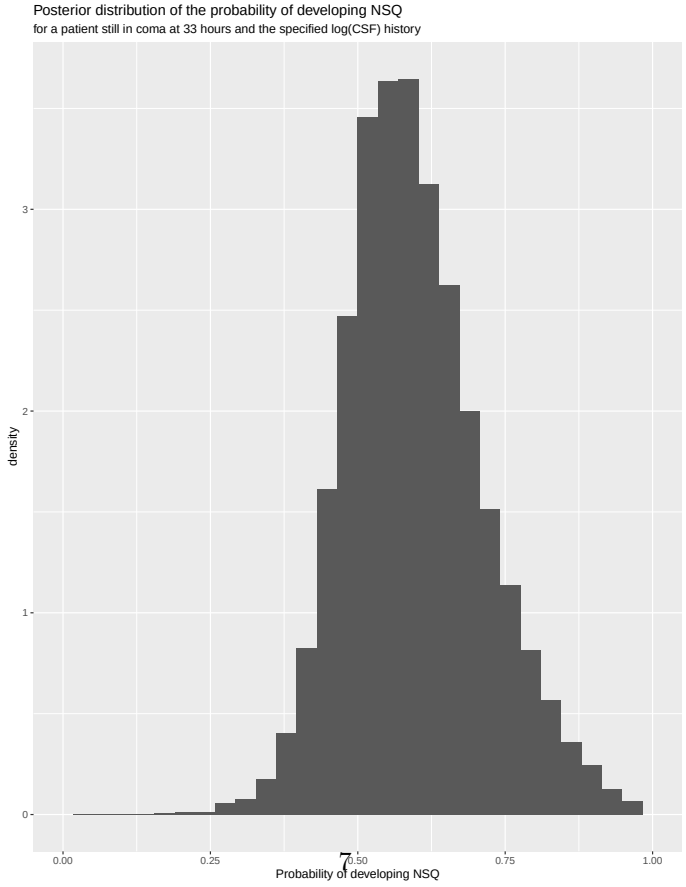
Note the posterior distributions of the intercepts α_0 and β_0 change a bit, which is as expected as age and gender are now adjusted for. The posterior distributions of coefficients on the latent factor, i.e., α_1, κ, β_1, and λ_1, remain very similar.

References

[1] Andrew Gelman, David Dunson, Donald Rubin, Hal S. Stern, and John B. Carlin *Bayesian Data Analysis*, CRC Press, Taylor & Francis Group, 3rd edition, 2014.


Posterior distribution of the probability of developing NSQ
for a patient still in coma at 55 hours and the specified log(CSF) history
density
0
1
2
3
4
0.00
0.25
0.50
0.75
1.00
Probability of developing NSQ

(a)


Posterior distribution of the probability of developing NSQ
for a patient still in coma at 33 hours and the specified log(CSF) history
density
0
1
2
3
0.00
0.25
0.50
0.75
1.00
Probability of developing NSQ

(b)

Figure 4