

Bayesian Analysis of Hydroxyurea Treatment Effects on Cats with Erythrocytosis

Jason Withrow

2024-12-02

Introduction

The purpose of this study is to explore the efficacy and effects of various doses of the drug hydroxyurea and see the effect it has on the packed cell volume (PCV) in cats with erythrocytosis. Erythrocytosis is defined by the increase in total red blood cells and is diagnosed through a complete blood cell count (CBC) test. Hydroxyurea was proposed as a potential treatment to reduce packed cell volume.

The data provided was of 40 cats all with erythrocytosis with half having primary erythrocytosis and the other half having secondary erythrocytosis. This was expressed in a Type variable with Type = 1 corresponding to primary erythrocytosis and Type = 0 corresponding to secondary erythrocytosis. The cats belonged to one of five breeds which was captured by the Breed variable and that takes values A, B, C, D, or E with breeds A, B, C being classified as domestic cats and breeds D and E being classified as non-domestic cats. This domestic variable was not originally in the data set but was added with Domestic = 1 meaning the cat is domestic and Domestic = 0 meaning the cat is non-domestic. Each cat was given a specific dose of hydroxyurea for a period of 3 months and that is captured by the variable Dose. Finally recorded was the change in packed cell volume (ΔPCV) following the treatment and that is captured in the variable DeltaPCV where the bigger the decrease the more effective the treatment.

The 2 main goals of this study are to determine if a higher dose of hydroxyurea is associated with a greater reduction in PCV. Additionally we want to see when controlling for the dose of hydroxyurea administered and the type of erythrocytosis, do domestic breeds of cats have a different expected change in pack cell volume than non-domestic cats

Methodology

Model Description

To analyze the efficacy of hydroxyurea in reducing packed cell volume in cats a Bayesian linear regression model was developed.

Let Y_{ij} be the values of the response variable (ΔPCV) for cat i within breed j and we are assuming Y_{ij} are independent of each other. Let $\beta = (\beta_1, \beta_2, \beta_3)$

The sampling model is given by:

$$Y_{ij} | \alpha, \beta, \sigma^2, x_{ij} \sim N(\mu_{ij}, \sigma^2) \quad \text{for } i = 1, \dots, n_j \quad \text{and } j = 1, \dots, J$$

and the linear model is given by:

$$\mu_{ij} = \alpha_j + \beta_1 + \beta_2(Dose)_{ij} + \beta_3(Type)_{ij}$$

Where

- α_j is the breed-specific random effect (differs for breeds A,B,C,D,E)
- β_1 is the baseline effect or intercept
- β_2 is the coefficient for dose
- β_3 is the coefficient for type of erythrocytosis

Priors:

$$\beta_1, \beta_2, \beta_3 \sim N(0, 10^{10})$$

$$\alpha_1, \dots, \alpha_j | \sigma_\alpha^2 \sim N(0, \sigma_\alpha^2)$$

$$\sigma_\alpha^2 \sim \text{Inverse} - \text{Gamma}(0.01, 0.01)$$

$$\sigma^2 \sim \text{Inverse} - \text{Gamma}(0.01, 0.01)$$

This model was chosen because we wanted to see the relationship between various predictors on the response variable ΔPCV so a regression model was an obvious choice. More specifically this model is a hierarchical model that accounts for both fixed-effects and breed-specific random effects. This was chosen because its logical to want to account for some kind of breed specific effect on the response variable. This also allows for seeing what differences there are between domestic and non-domestic cats. A fixed effect for dose and type were added to see what the relationships between those and the response variable are. There is little to none prior information about any of the parameters so very weakly informative priors were chosen as to not have a very strong influence on the posterior distributions that will be generated.

Implementation

The hallmark of Bayesian analysis is computing the posterior distribution but we cannot write it down in closed form

$$P(\beta, \sigma^2, \alpha | y, x) \propto P(y | \alpha, \beta, \sigma^2, x) P(\beta, \alpha, \sigma^2)$$

Because of this other software was need and the main tool used was R-studio and JAGS (Just Another Gibbs Sampler) to estimate the posterior distributions for the coefficients.

The data were structured into a list format that included:

- n: the total number of observations,
- p: the number of predictors (hydroxyurea dose and erythrocytosis type),
- The observed values of ΔPCV (Y)
- Dose: the hydroxyurea dose for each cat,
- Type: the type of erythrocytosis (primary or secondary),
- Breed: the breed of each cat (used as a random effect).

As for initialization of parameters $\beta_1, \beta_2, \beta_3$ and $\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5$ were set to zero. The variance parameters τ_α^2 and σ^2 were set to 1.

The MCMC chains were run using JAGS, with 3 chains initialized and adapted for 10,000 iterations each. The burn-in phase was 20,000 iterations, followed by 50,000 iterations saved across the chains with thinning every 10th iteration to reduce autocorrelation in the samples. This means we took every 10th iteration and used that to generate the distribution.

Lastly convergence of the Chains was assessed using density and trace plots and can be found in the supplemental material section

Results

Research Question 1: Is a higher dose of hydroxyurea associated with a greater reduction in packed cell volume?

To assess the effect of dose of hydroxyurea on the change in packed cell volume (ΔPCV) the posterior distribution of the coefficient β_2 was examined. It was found for the coefficient of β_2 the posterior mean was -0.3968 indicating that on average for every 1 unit increase in hydroxyurea dose ΔPCV is expected to decrease by -0.3968 . The 95% credible interval for the posterior mean of β_2 was $[-0.5414, -0.2519]$.

The figure below shows the posterior distribution for β_2

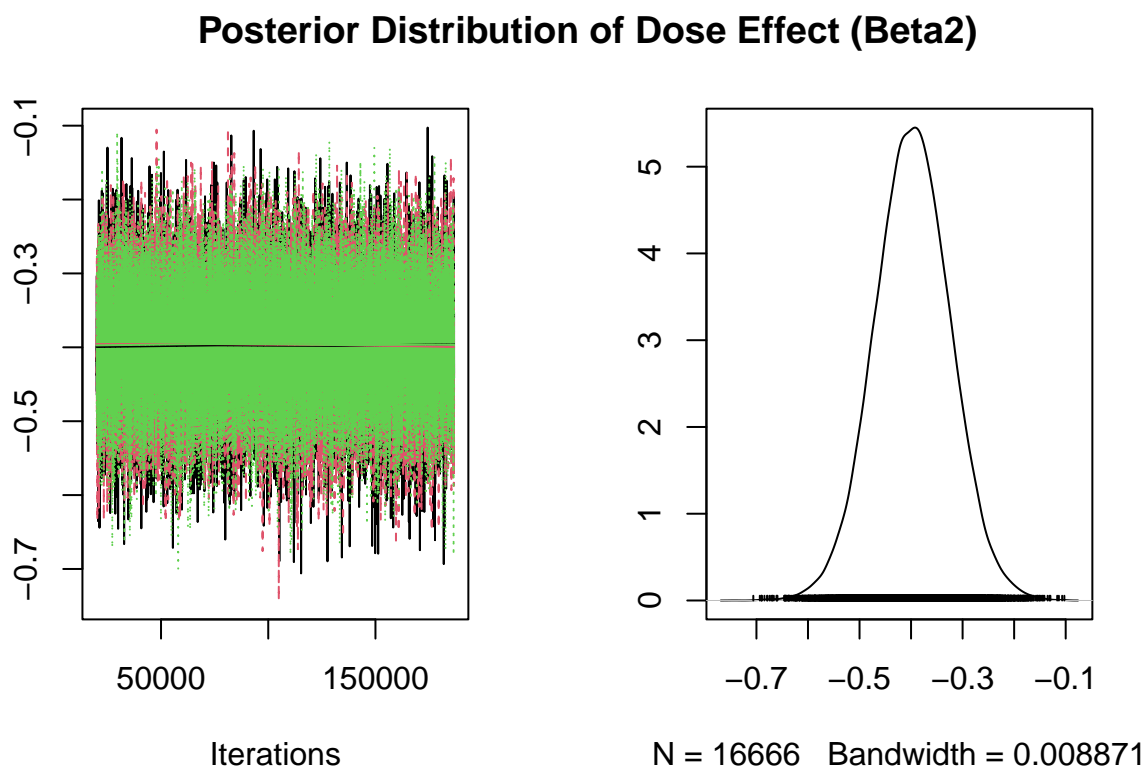


Table 1: Posterior Summary Statistics

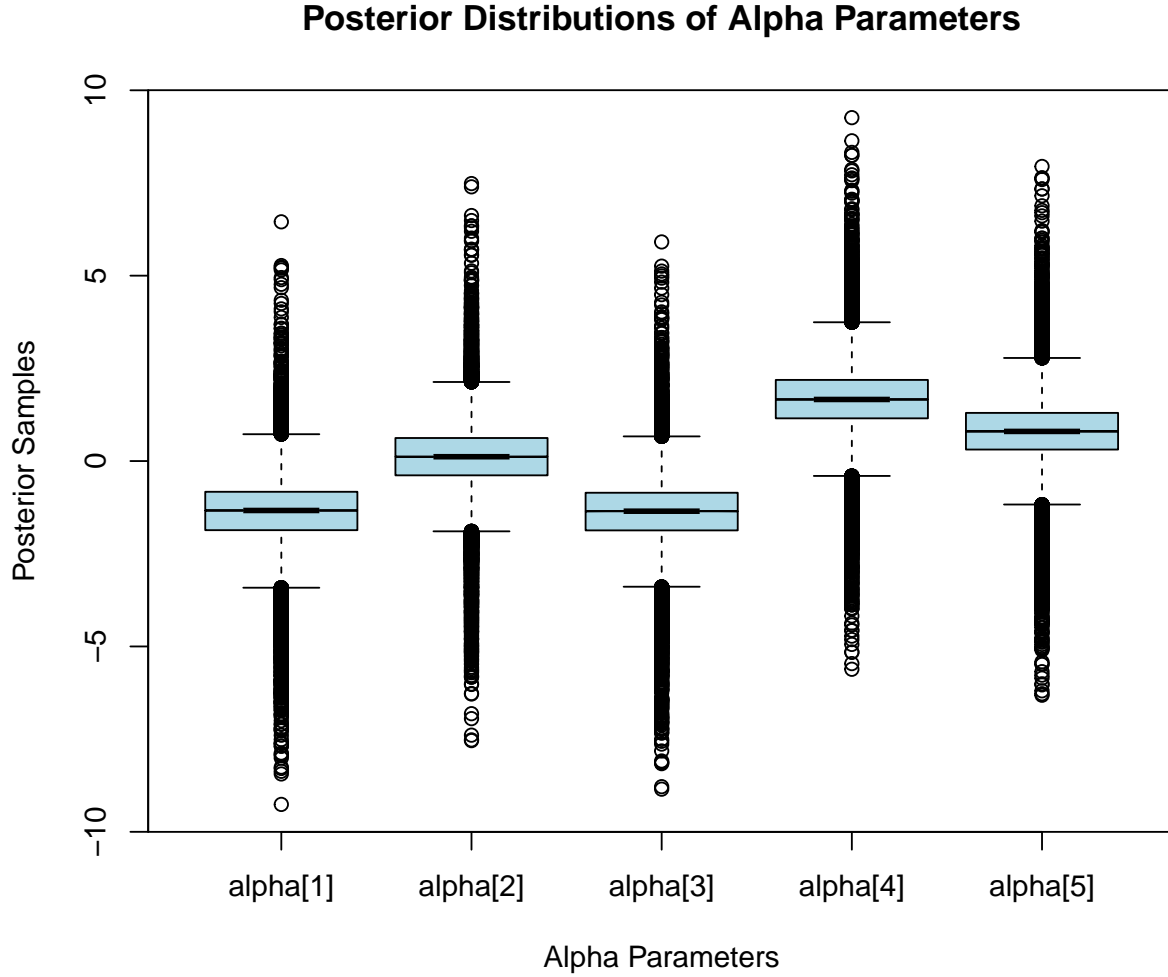
	Mean	SD	Naive SE	Time-series SE
alpha[1]	-1.3675	0.9295	0.0042	0.0121
alpha[2]	0.1064	0.9152	0.0041	0.0118
alpha[3]	-1.3824	0.9164	0.0041	0.0123
alpha[4]	1.6726	0.9255	0.0041	0.0120
alpha[5]	0.7977	0.9069	0.0041	0.0120
beta[1]	-2.7395	0.9808	0.0044	0.0130
beta[2]	-0.3968	0.0736	0.0003	0.0004
beta[3]	0.2582	0.3178	0.0014	0.0014
sig2	0.9273	0.2437	0.0011	0.0011
sig2.alpha	3.8082	7.5668	0.0338	0.0697

Table 2: Posterior Summary Credible Intervals

	2.5%	25%	50%	75%	97.5%
alpha[1]	-3.3039	-1.8638	-1.3330	-0.8286	0.3560
alpha[2]	-1.7478	-0.3853	0.1170	0.6218	1.8619
alpha[3]	-3.2889	-1.8698	-1.3525	-0.8552	0.3337
alpha[4]	-0.1095	1.1519	1.6610	2.1885	3.5238
alpha[5]	-1.0289	0.3104	0.7993	1.2987	2.5510
beta[1]	-4.6334	-3.3141	-2.7571	-2.1858	-0.7537
beta[2]	-0.5414	-0.4457	-0.3970	-0.3481	-0.2519
beta[3]	-0.3751	0.0490	0.2588	0.4691	0.8852
sig2	0.5657	0.7551	0.8886	1.0556	1.5091
sig2.alpha	0.5723	1.3377	2.2242	3.9925	16.7099

Research Question 2: Controlling for the dose of hydroxyurea administered and the type of erythrocytosis, do domestic breeds of cats have a different expected change in pack cell volume than non-domestic cats?

To answer this question we can first look at a boxplot of the posterior mean summaries for each of the α_j 's



Just looking at the boxplots we can see there appears to be a difference between domestic cats (Breeds A,B,C) and non-domestic cats (Breeds D,E). The posterior means for the domestic cats are lower than for the non-domestic cats. The Posterior means and credible intervals for all α_j 's is given by the following table

Table 3: Posterior Summary for Alpha (Breed Effects)

	Mean	2.5%	97.5%
alpha[1]	-1.3675	-3.3039	0.3560
alpha[2]	0.1064	-1.7478	1.8619
alpha[3]	-1.3824	-3.2889	0.3337
alpha[4]	1.6726	-0.1095	3.5238
alpha[5]	0.7977	-1.0289	2.5510

Due to the fact we are interested in if there is a difference in expected change in pack cell volume of domestic cats and non-domestic cats we can average the means for domestic cats (Breeds A,B,C) and for non-domestic cats (Breeds D, E) and compute the difference.

$$\bar{\alpha}_{domestic} = \frac{\alpha_1 + \alpha_2 + \alpha_3}{3}$$

$$\bar{\alpha}_{non-domestic} = \frac{\alpha_4 + \alpha_5}{2}$$

The difference between

$$\bar{\alpha}_{non-domestic}$$

and

$$\bar{\alpha}_{domestic}$$

was -2.116329 and we can also use the difference samples we generated to get a 95% credible interval of $[-2.757295, -1.448796]$ for $\bar{\alpha}_{domestic} - \bar{\alpha}_{non-domestic}$

Conclusion

In this report our goal was to assess the efficacy of the hydroxyurea treatment in reducing packed cell volume in cats with erythrocytosis. The method used to explore this was a Bayesian analysis specifically a Bayesian hierarchical linear model.

In regards to the first question that asked if a higher dose of hydroxyurea is associated with a greater reduction in packed cell volume we found that the estimate for the mean of β_2 was -0.3968 indicating that on average for every 1 unit increase in hydroxyurea dose ΔPCV is expected to decrease by -0.3968 . The 95% credible interval for the posterior mean of β_2 was $[-0.5414, -0.2519]$. This means that an increase in dose is associated with a greater reduction in packed cell volume. Due to the fact that the 95% credible interval for the mean does not include zero we can conclude that the estimate is most likely statistically significant.

In regards to the second question that asked if there was a difference in expected change in pack cell volume between domestic and non-domestic cats while controlling for dose we added a breed specific intercept term to our model α_j . The 3 α_j 's for domestic cats were averaged and the 2 α_j 's for non-domestic cats were averaged and subtracted them to get a difference of -2.116329 and which means that on average the domestic breeds (breeds A, B, C) have a lower value for the breed specific intercepts compared to non domestic breeds (breeds D,E). This indicates that the domestic breeds experience a greater baseline reduction in packed cell volume. It was also found that the 95% credible interval for $\bar{\alpha}_{domestic} - \bar{\alpha}_{non-domestic}$ was $[-2.757295, -1.448796]$ and this credible interval is only negative and doesnt include zero so we can conclude it is significant and that for domestic breeds we are 95% certain the mean is contained in this interval and it is only negative values so we can say domestic breeds have a significantly lower packed cell volume than non-domestic breeds