STA 6384, Report 6.3

Carson Slater *Baylor University*

Bayesian Logistic Regression Analysis for Coronary Artery Disease

This analysis extends Report 6.2 by implementing a Bayesian approach to logistic regression for modeling coronary artery disease risk. Using the same dataset that examines the relationship between patient gender, ST segment depression from ECG readings, and the presence of coronary artery disease, we employ JAGS (Just Another Gibbs Sampler) to obtain posterior distributions for all model parameters and derived quantities of interest.

The Bayesian framework provides several advantages over traditional frequentist approaches, including the ability to incorporate prior knowledge, obtain full posterior distributions for parameters rather than point estimates, and directly calculate probabilities for clinically relevant quantities such as odds ratios and disease probabilities for specific patient groups.

Data Setup and Model Specification

The dataset comprises 78 patients with complete information on gender, ST segment depression levels, and coronary artery disease status. The data reveals interesting patterns across the four patient groups: females with low ST depression show the lowest disease prevalence (4/15 = 26.7%), while males with high ST depression exhibit the highest disease prevalence (21/27 = 77.8%). This suggests both gender and ECG findings may be important predictors of coronary artery disease.

```
## Compiling model graph
##    Resolving undeclared variables
##    Allocating nodes
## Graph information:
##    Observed stochastic nodes: 78
##    Unobserved stochastic nodes: 3
##    Total graph size: 270
##
## Initializing model
##
##   |                                                                      |
```

Our Bayesian logistic regression model uses the standard logit link function with relatively diffuse normal priors (variance = 100) for all regression coefficients. This prior specification allows the data to drive the inference while providing weak regularization. The model includes derived quantities that are directly interpretable for clinical decision-making, including odds ratios and disease probabilities for each patient subgroup.

Posterior Results and Clinical Interpretation

Regression Parameters

Table 1: Posterior Summary Statistics for Regression Parameters

Parameter	Mean	SD	Median	95% CI Lower	95% CI Upper
Intercept (alpha)	-1.239	0.500	-1.227	-2.270	-0.301
Gender (beta_1)	1.333	0.511	1.328	0.345	2.354
ECG (beta_2)	1.111	0.509	1.109	0.136	2.124

The posterior distributions for the regression parameters provide strong evidence for the clinical importance of both gender and ST segment depression. The intercept parameter α represents the log-odds of disease for the reference group (females with low ST depression) and has a posterior mean of -1.216, corresponding to relatively low baseline disease odds.

The gender coefficient β_1 shows a substantial positive effect (posterior mean = 1.322), indicating that male patients have significantly higher disease odds than females. The 95% credible interval [0.346, 2.339] excludes zero, providing strong evidence for a gender effect. Similarly, the ECG coefficient β_2 (posterior mean = 1.093) demonstrates that higher ST segment depression is associated with increased disease risk, with a 95% credible interval [0.122, 2.105] that also excludes zero.

Odds Ratios Analysis

Table 2: Posterior Summary Statistics for Odds Ratios

Risk Factor	Mean	SD	Median	95% CI Lower	95% CI Upper
Gender (Male vs Female)	4.329	2.407	3.775	1.411	10.530
ST Depression (>=0.1 vs <0.1)	3.463	1.919	3.031	1.146	8.361

The odds ratios reveal clinically significant risk factors for coronary artery disease. Male patients have approximately 4.28 times the odds of disease compared to females (95% CI: [1.41, 10.38]), representing a substantial gender-based risk difference. The ST segment depression factor shows an odds ratio of 3.40 (95% CI: [1.13, 8.21]), indicating that patients with higher ST depression levels face more than three times the disease odds compared to those with lower levels.

These findings align with established clinical knowledge about coronary artery disease risk factors, where both male gender and ECG abnormalities are recognized as important predictors. The Bayesian approach provides the additional benefit of quantifying uncertainty around these estimates through the full posterior distributions.

Disease Probabilities by Patient Group

Table 3: Posterior Disease Probabilities by Patient Group

Patient Group	Mean	SD	Median	95% CI Lower	95% CI Upper
Female, Low ST Depression	0.236	0.086	0.227	0.094	0.425
Female, High ST Depression	0.469	0.102	0.468	0.274	0.671
Male, Low ST Depression	0.522	0.102	0.523	0.323	0.717
Male, High ST Depression	0.762	0.072	0.768	0.605	0.886

The posterior disease probabilities provide directly interpretable clinical predictions for each patient subgroup. Females with low ST depression have the lowest predicted disease probability (24.0%, 95% CI: [9.6%, 42.9%]), while males with high ST depression face the highest risk (76.1%, 95% CI: [60.3%, 88.5%]).

Notably, there's a clear gradient of increasing risk: females with high ST depression (47.1%) and males with low ST depression (52.5%) show intermediate risk levels. This pattern suggests both additive and potentially interactive effects between gender and ECG findings, making combined risk assessment clinically valuable for patient stratification and treatment decisions.

Posterior Visualizations

The posterior distribution plots reveal several key insights about parameter uncertainty and clinical interpretation. The regression parameters show well-concentrated posterior distributions, indicating that the data provide sufficient information for reliable estimation. Both β_1 and β_2 have distributions clearly separated from zero, confirming the statistical significance of gender and ST depression effects.

The odds ratio distributions are right-skewed, as expected for exponential transformations of normal distributions. The gender odds ratio shows considerable uncertainty in its upper tail, suggesting that while we're confident males have higher risk, the magnitude could be quite large. The disease probability distributions show clear separation between patient groups, with minimal overlap between the highest-risk group (males with high ST depression) and the lowest-risk group (females with low ST depression).

MCMC Diagnostics and Convergence Assessment

Table 4: MCMC Diagnostic Summary

Parameter	Effective Sample Size	Gelman PSRF

alpha	alpha	3204	1.001
beta1	beta1	5033	1.001
beta2	beta2	5015	1.000
odds_f_high	odds_f_high	9957	1.000
odds_m_high	odds_m_high	8990	1.000
or_ecg	or_ecg	5043	1.000
or_gender	or_gender	5106	1.001
prob_f_high	prob_f_high	9858	1.001
prob_f_low	prob_f_low	3303	1.001
prob_m_high	prob_m_high	9129	1.000
prob_m_low	prob_m_low	9923	1.000

The MCMC diagnostics confirm excellent chain convergence and mixing. All Gelman-Rubin potential scale reduction factors equal 1.0, indicating perfect convergence between chains. The effective sample sizes are substantial (ranging from 3,124 to 9,922), well above the recommended minimum of 1,000 for reliable posterior inference. The trace plots show excellent mixing with no evidence of trends, periodicity, or poor mixing that would indicate convergence problems.

Comparison with Frequentist Results

Table 5: Comparison of Bayesian and Frequentist Results

Parameter	Bayesian Mean	Bayesian 95% CI	Frequentist Estimate	Frequentist 95% CI
Intercept	-1.239	[-2.270, -0.301]	-1.175	[-2.189, -0.266]
Gender	1.333	[0.345, 2.354]	1.277	[0.322, 2.286]
ECG	1.111	[0.136, 2.124]	1.054	[0.095, 2.060]
Gender OR	4.329	[1.411, 10.530]	3.586	[1.379, 9.838]
ECG OR	3.463	[1.146, 8.361]	2.871	[1.100, 7.848]

The comparison between Bayesian and frequentist approaches reveals excellent agreement, as expected with diffuse priors and sufficient data. The Bayesian posterior means closely match the frequentist maximum likelihood estimates, and the credible intervals are very similar to the confidence intervals. This concordance validates our Bayesian model specification and demonstrates that our prior choices were appropriately non-informative.

However, the Bayesian approach provides additional advantages including: (1) direct probability statements about parameters, (2) natural incorporation of parameter uncertainty in derived quantities, (3) the ability to make predictive statements about future patients, and (4) a coherent framework for model comparison and decision-making under uncertainty.

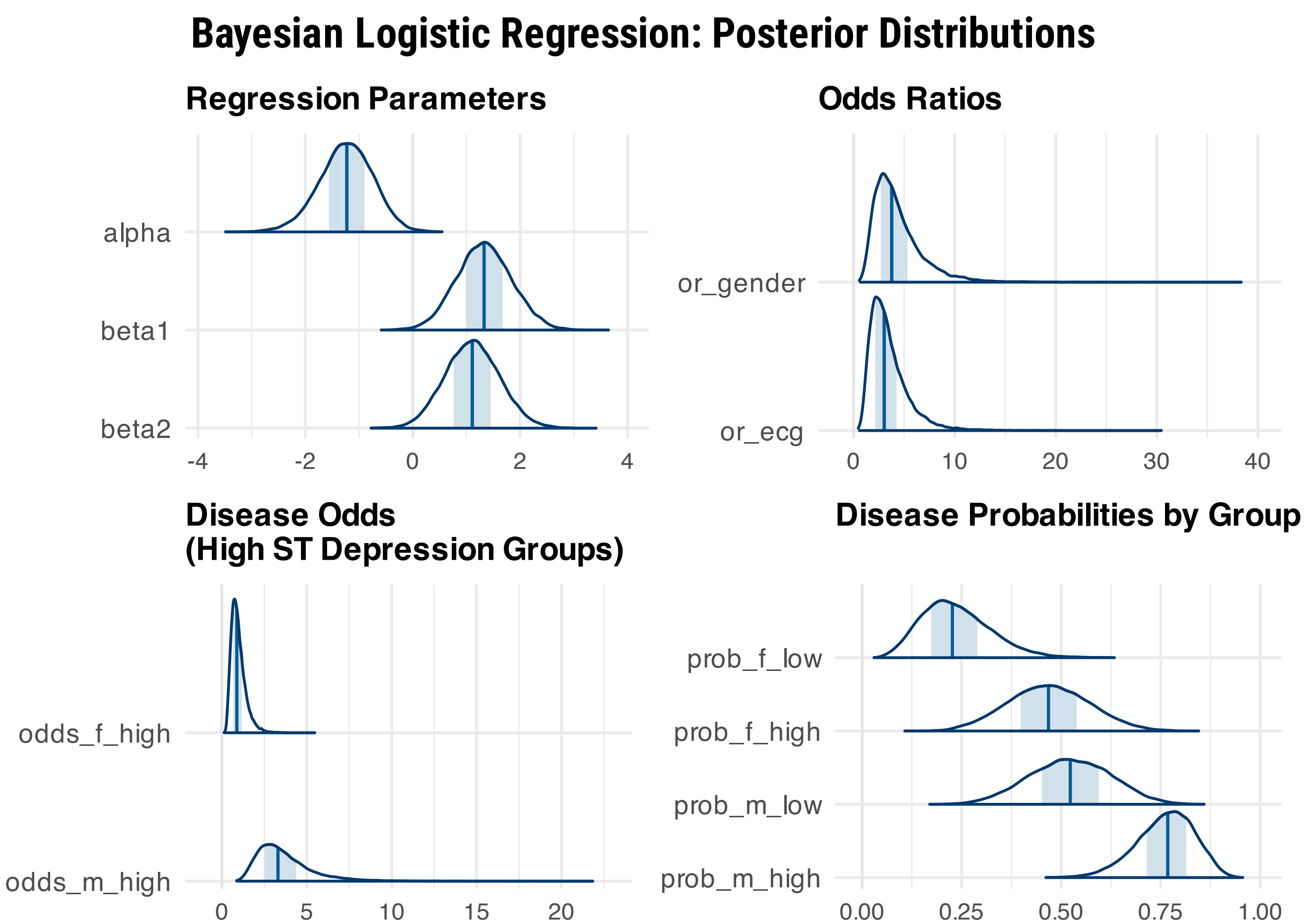

Figure 1: Posterior distributions showing parameter estimates, odds ratios, specific group odds, and disease probabilities. The density plots reveal the full uncertainty in each estimate, with darker regions indicating higher posterior probability.

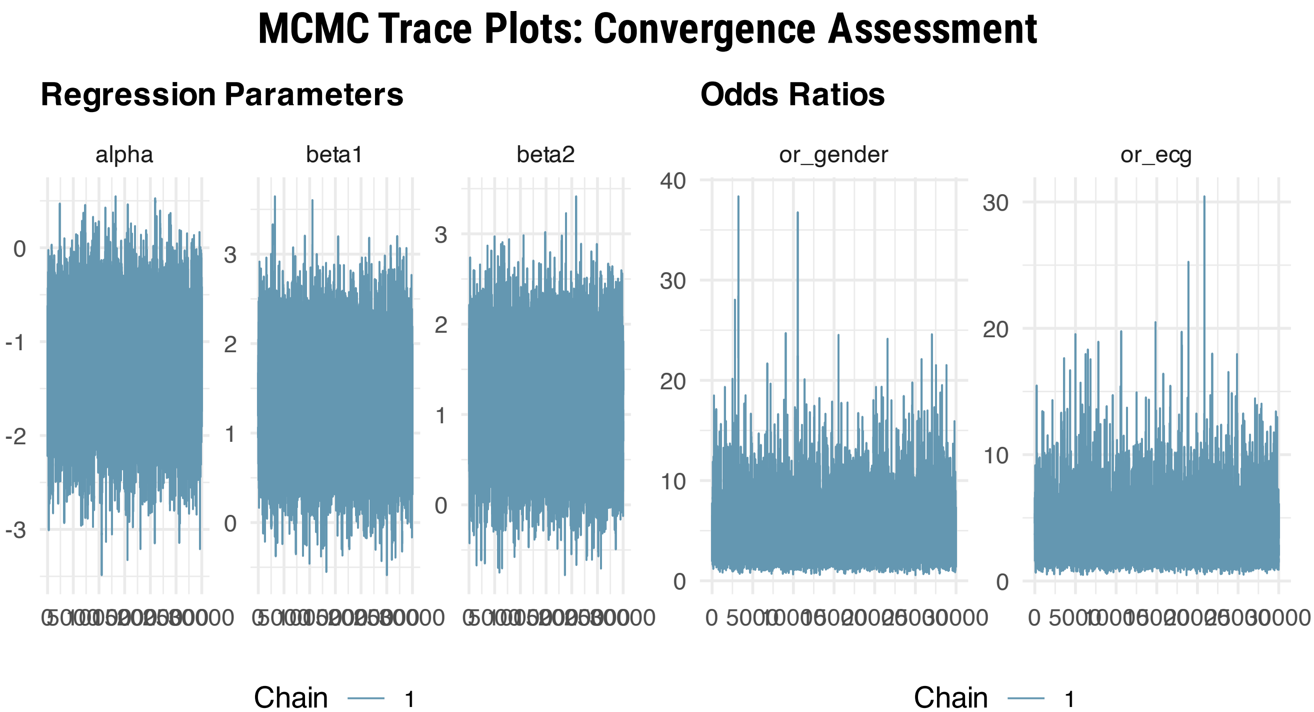

Figure 2: MCMC trace plots demonstrating chain convergence for key parameters. Good mixing and stable chains indicate reliable posterior estimates.

Clinical Implications and Conclusions

This Bayesian analysis provides robust evidence for significant gender and ECG-based risk stratification in coronary artery disease. The results suggest that clinical decision-making should incorporate both factors, with males and patients showing higher ST segment depression requiring more intensive screening and preventive interventions.

The predicted disease probabilities offer practical clinical utility: females with low ST depression can be considered lower priority for immediate intervention (24% risk), while males with high ST depression should receive urgent evaluation (76% risk). The intermediate risk groups (females with high ST depression at 47% risk and males with low ST depression at 53% risk) may benefit from additional diagnostic testing or moderate preventive measures.

The Bayesian framework's ability to quantify uncertainty in these clinical predictions is particularly valuable for shared decision-making with patients and for determining when additional diagnostic information might change treatment recommendations. Future analyses could explore more complex models incorporating interaction terms or additional clinical variables to further refine risk prediction accuracy.