Bayesian Modeling of Neonatal Mortality: Interpretable Risk Factors in Small Clinical Datasets

Introduction

In this project, we built an interpretable Bayesian logistic regression model to identify key risk factors associated with neonatal mortality in very low birth-weight infants. Using clinical data from a neonatal intensive care unit, we compared pooled and sex-stratified hierarchical models to assess whether mortality risk differs by sex while accounting for small sample sizes and uncertainty. The analysis highlights how Bayesian methods can provide stable, interpretable insights in healthcare settings where data is limited, noisy, and high-stakes.

Problem

Neonatal mortality remains a major public health concern, particularly among very low birth-weight infants. Clinicians often rely on multiple clinical indicators (e.g., birth weight, platelet count, respiratory complications) to assess mortality risk.
However, real clinical datasets are often small, incomplete, and heterogeneous, making traditional modeling approaches unstable or difficult to interpret. This project explores how Bayesian modeling can be used to quantify uncertainty and extract meaningful risk signals from limited neonatal health data.

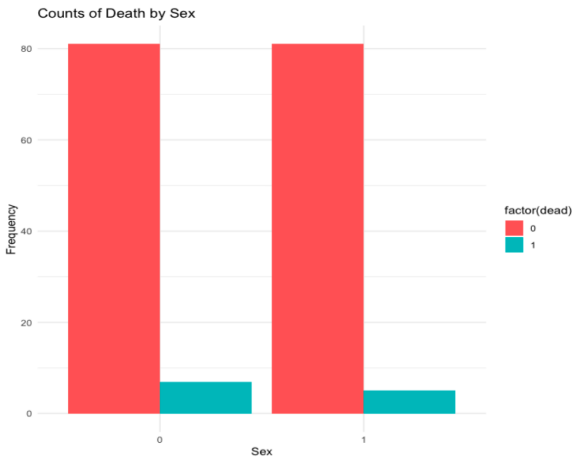

Figure 1. Observed neonatal mortality counts grouped by sex, motivating comparison of pooled vs hierarchical models

Skills demonstrated

- Bayesian logistic regression
- Hierarchical (multilevel) modeling
- Prior specification informed by domain literature
- Feature selection and multicollinearity analysis
- Model diagnostics and convergence assessment
- Posterior predictive checks

- Leave-One-Out Cross-Validation (PSIS-LOO)
- Interpretation under data and domain limitations

Tools & Methods

R, Tidyverse, brms (Stan backend), loo, Bayesian model diagnostics and posterior analysis

Data

- Publicly available *Very Low Birth Weight Infants* dataset (Duke University Medical Center)
- Historical clinical data (1981–1987)
- Initial dataset: 671 infants
- Final analyzed subset after cleaning: **174 infants**
- Binary outcome: neonatal mortality
- Key predictors:
    - Birth weight
    - Platelet count
    - Pneumothorax
    - Intraventricular hemorrhage (IVH)
    - Sex

Although the dataset is small and historical, it reflects common challenges in clinical and bioinformatics research, where data collection is expensive and uncertainty must be explicitly modeled rather than ignored

Approach

Before modeling, feature selection was performed using correlation analysis and variance inflation factors (VIF) to reduce multicollinearity and improve interpretability.

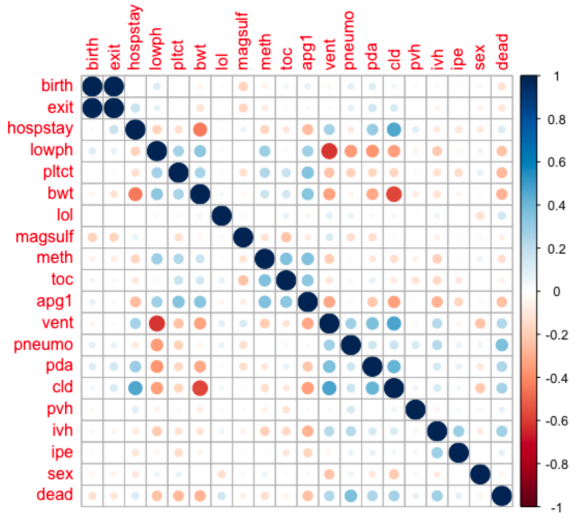

Figure . Correlation matrix used to assess multicollinearity prior to feature selection (VIF + correlation thresholds).

Rather than optimizing a single predictive model, this analysis focuses on **comparing modeling assumptions**. Two Bayesian logistic regression models were constructed:

1. Pooled Bayesian Logistic Regression

- Assumes a shared mortality risk structure across all infants
- Models mortality as a function of selected clinical predictors
- Serves as a baseline model

Goal: Establish whether core clinical variables alone explain mortality risk.

2. Hierarchical Bayesian Logistic Regression

- Includes sex as a group-level (random intercept) effect
- Allows baseline mortality risk to vary between male and female infants
- Retains the same clinical predictors as the pooled model

Goal: Assess whether stratifying by sex improves inference or predictive performance under limited data conditions.

<u>Model Validation & Evaluation</u>

To ensure reliability and interpretability, both models were evaluated using:

- Convergence diagnostics ($\hat{R} \approx 1$, high effective sample sizes)
- Posterior predictive checks to assess model fit whether simulated outcomes from the models aligned with the observed data distribution.

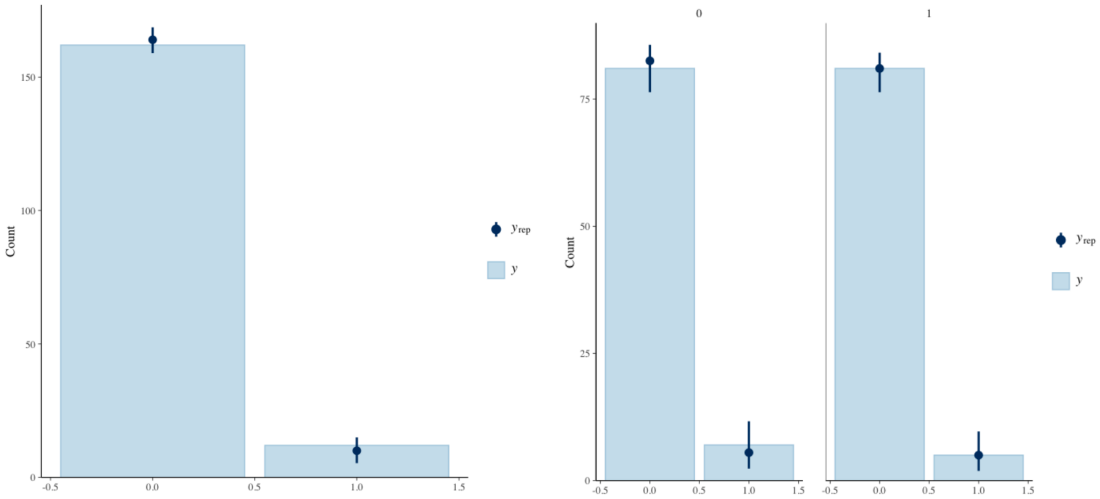

Figure 3. Posterior predictive check for pooled Bayesian logistic regression
Figure 4. Posterior predictive check of hierarchical model grouped by sex.

- Leave-One-Out Cross-Validation (PSIS-LOO) for model comparison
- Inspection of Pareto k values to identify influential observations

Both models converged without divergences and showed stable posterior behavior.

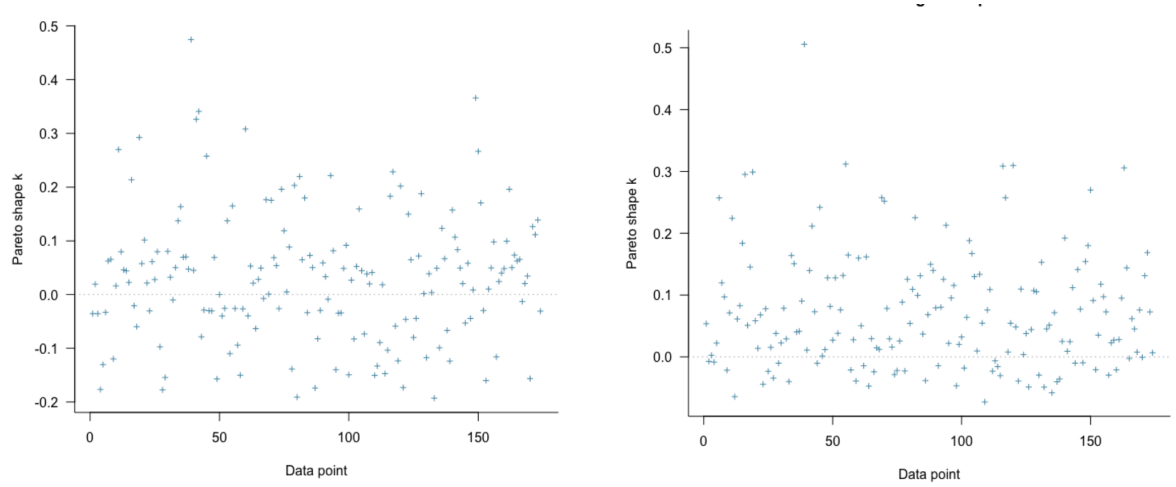

Figure 4. Non-hierarchical model vs hierarchical model of PSIS-LOO diagnostics showing absence of influential observations (Pareto k < 0.7)

Results & Insights

- **Birth weight** and **platelet count** were strongly associated with neonatal mortality
- **Pneumothorax** and **IVH** substantially increased mortality risk
- Introducing sex as a hierarchical grouping factor **did not meaningfully improve predictive performance**
- Both models achieved approximately **93% classification accuracy**, though results should be interpreted cautiously due to class imbalance and small sample size
- LOO cross-validation showed negligible differences between pooled and hierarchical models

These results suggest that, in early neonatal life, mortality risk is dominated by clinical factors rather than sex-specific baseline differences in this dataset.

Interpretation Under Data Limitations

The hierarchical model did not outperform the pooled model, highlighting an important insight: Increasing model complexity does not necessarily yield better inference when data is limited.

Bayesian modeling provided a principled framework for:

- Quantifying uncertainty
- Incorporating prior medical knowledge
- Avoiding overconfident conclusions from small samples

Potential Extensions

With larger or more recent datasets, several extensions could be explored:

- Time-to-event (survival) modeling instead of binary classification
- External validation on modern NICU cohorts
- Incorporation of additional clinical covariates

- Exploration of non-linear or interaction effects

These extensions would further clarify how data availability and modeling choices jointly affect clinical inference.

Key Takeaway - This project demonstrates how Bayesian modeling, uncertainty quantification, and careful model comparison can support interpretable analysis in small, high-stakes healthcare datasets—emphasizing critical interpretation over raw predictive performance.