The Effects of Labor-Associated Drug Interventions on Neonatal Health Outcomes: A Predictive Analysis of Apgar5 Scores*

Steroids Enhance Neonatal Health, While Antibiotics Show Limited Effectiveness

Yunkyung Ko

December 3, 2024

How maternal and neonatal drug interventions during labor shape the health status of a newborn remains a key question in perinatal care. With the 2023 Natality Data for the United States, the relationship between six labor-related treatments and Apgar5 scores, a standard measure of neonatal health, was examined. Using a Random Forest model for prediction and a Bayesian Linear Model for inference, we identified usage of steroids and chorioamnionitis as the most impactful, while antibiotics dosage showed minimal influence. The analysis shows disparities in treatment efficacy, and further suggests the need for optimized corticosteroid formulations or targeted antibiotic protocols.

Table of Contents

1 Introduction **2**

2 Data **3**

2.1 Overview 3
2.2 Measurement 4
2.3 Outcome variables 5
2.4 Predictor variables 7
2.5 Correlation between predictor variables 15

3 Model **18**

3.1 Model Specification 18

*Code and data are available at: https://github.com/koyunkyung/infant_health.

3.2 Model Justification 20

4 Results **21**
4.1 Results from examining the analysis dataset 21
4.2 Results from the prediction model 30

5 Discussion **32**
5.1 Optimization of anetenatal corticosteroid formulations 32
5.2 Targeted antibiotics and enforcement of evidence-based protocols 32
5.3 Disparities in treatment efficacy influenced by contextual factors 34
5.4 Limitations of the study and directions for future research 34

A Appendix **36**
A.1 Additional data details 36
A.2 Model details 38

References **46**

1 Introduction

Labor and birth is a time of excitement and anticipation, along with uncertainty and anxiety (Public Health Agency of Canada 2023). This emotional complexity is often heightened by decisions surrounding medical interventions that directly impact maternal and newborn health (Reproductive Health Journal 2019). Women frequently report anxiety in medication use especially during labor, driven by concerns about potential side effects (Reproductive Health Journal 2019) and insufficient information (Public Health Agency of Canada 2023). To address these concerns, this paper analyzes the 2023 Natality Data for the United States (National Bureau of Economic Research (NBER) 2023), specifically evaluating the relationship between six labor-related treatments and neonatal health outcomes, as measured by the Apgar5 score (Centers for Disease Control and Prevention (CDC) and National Center for Health Statistics (NCHS) 2023).

The estimand is the average treatment effect (ATE) of six labor-associated interventions on neonatal health outcomes, quantified by Apgar5 scores (National Bureau of Economic Research (NBER) 2023). Specifically, we aim to estimate how the administration of each treatment changes the Apgar score compared to scenarios where the treatment is not applied (Alexander 2023). By applying predictive modeling through a Random Forest model and inferential analysis through a Bayesian linear model, we assess not only the magnitude of these effects(GeeksforGeeks 2024) but also the underlying uncertainties(Strimmer Lab 2024).

The analysis shows that prenatal exposure to steroids significantly improve Apgar scores (Table 2). Conversely, antibiotics dosage doesn't show as much contribution to the score distribution (Table 2) and rather raises concerns about potential risks, including microbiota disruption

(Respiratory Therapy Zone 2024). The paper therefore proposes significant disparities in the consistent application and effectiveness of these treatments (Figure 23).

This study has its significance in that it states the need to improve health outcomes for mothers and their newborns. As recognized by the World Health Organization, a healthy start in life has significant repercussions for a person's health and well-being during infancy, childhood, and adulthood (World Health Organization 2021). Ensuring quality care in maternal and newborn health is integral to the right to health, equity, and the preservation of dignity for women and their babies (World Health Organization 2021).

The remainder of this paper is structured as follows: Section 2 describes the dataset and methodology and Section 3 exhibits the use of predictive and inferential models. Section 4 presents the results of the analysis, detailing the observed relationships between treatments and neonatal health outcomes. Section 5 discusses the implications of these findings, focusing on healthcare policy and clinical practice improvements. Finally, Section 5.4 concludes with a discussion of the study's limitations and potential areas for future research. Section A provides additional data details and model diagnostics.

2 Data

2.1 Overview

We use the the statistical programming language R (R Core Team 2023) to analyze the relationship between maternal drug treatments and neonatal health outcomes in the '2023 Natality Data for the United States' (National Bureau of Economic Research (NBER) 2023). Provided by NBER and sourced from the National Center for Health Statistics, this dataset contains 3,605,081 live birth statistics across the U.S., including demographic details, health metrics, and geographical breakdowns by state and country (Centers for Disease Control and Prevention (CDC) and National Center for Health Statistics (NCHS) 2023). To represent treatments directly affecting the infant, variables under the category of 'Characteristics of Labor and Delivery' were selected, which capture medical interventions administered to the mother or infant during the critical period of childbirth (Centers for Disease Control and Prevention (CDC) 2016). Six different types of treatments are under this category, which are labor induction, augmentation, the use of steroids, antibiotics, chorioamnionitis, and anesthesia (Centers for Disease Control and Prevention (CDC) and National Center for Health Statistics (NCHS) 2023). For the representation of infant health outcomes, the 'APGAR5' variable was selected, which shows the 5-minute Apgar score, a widely recognized indicator of an infant's immediate health status post-delivery (Medical News Today 2024).

To ensure a balanced distribution of observations across all Apgar5 score groups, we applied stratified random sampling and designated the sample size per group as 2,000. In addition, for data quality, we standardized the treatment measuring variables, containing respondents' answers of "Yes", "No", or "Unknown", into binary numerical variables. Then, the variables were converted into factors to prepare the data for predictive modeling(Analytics Vidhya 2015), and split into training and testing sets to support machine learning workflows(Jason Brownlee 2019).

In performing the analysis, we utilized several R packages. `tidyverse` (Wickham et al. 2019) was used for data manipulation, and `ggplot2` (Wickham 2016) was used for visualizing results in graphical methods. `randomForest` (Andy Liaw and Matthew Wiener 2002) and `caret` (Kuhn and Max 2008) were used for Random Forest modeling while `rstanarm` (Goodrich et al. 2024) was used for Bayesian linear modeling in the process of generating predictions.

2.2 Measurement

The data transforms the phenomenon of infant health status into data, defining quantifiable factors that can capture the complex impacts of maternal and neonatal pharmacological interventions (National Bureau of Economic Research (NBER) 2023). The transformation is achieved through the collection of information on various treatments and conditions experienced during labor and delivery, such as steroid use, chorioamnionitis, antibiotics, and anesthesia, among others (Centers for Disease Control and Prevention (CDC) and National Center for Health Statistics (NCHS) 2023). These variables, which were recorded through surveys or medical records, uses checkboxes or scales that allow for quantification (Centers for Disease Control and Prevention (CDC) 2016). Errors such as inconsistencies in how data is recorded can arise during this kind of measurement process (Alexander 2023). For instance, subjective interpretations of checkbox responses or variability in medical record documentation can lead to measurement error (HealthKnowledge 2024). These errors may result from differences in definitions of treatments across facilities or the accuracy of self-reported information in surveys (HealthKnowledge 2024).

The 5-minute Apgar score itself, is also a simplified numerical representation of an infant's health, which may overlook the complexities of their condition. While useful for rapid clinical decision-making, the Apgar score may fail to capture nuanced aspects of an infant's condition, such as underlying metabolic imbalances, subtle neurological issues, or long-term impact of perinatal complications (American Academy of Pediatrics 2006).

By structuring these variables into categorical or numerical forms, the data can be analyzed and used to model the relationship between treatments and the APGAR5 score. Specifically, drug usage during labor and delivery was encoded as binary variables (Section 2.4). However, this has a limitation that it disregards the quantitative intensity, dosage, or frequency of drug administration, leading to incomplete modeling of the treatment's impact (GeeksforGeeks 2022). This simplification may also amplify noise in the data by grouping dissimilar observations into the same category, reducing the model's ability to detect subtle relationships (GeeksforGeeks 2022).

More detailed information of data measurement including the survey and sampling methods are shown in Section A.1.

2.3 Outcome variables

2.3.1 Apgar Score: a measure the infant's chance of surviving the first year of life

The Apgar score is a measure of the need for resuscitation (Centers for Disease Control and Prevention (CDC) and National Center for Health Statistics (NCHS) 2023) to the infant, which is "the act of bringing someone back to life or waking them" (Cambridge University Press n.d.). It is a test given to newborns soon after birth (5 minutes) to check 'Appearance(skin color)', 'Pulse(heart rate)', 'Grimace response(reflexes)', 'Activity(muscle tone)', 'Respiration(breathing rate and effort)' (KidsHealth from Nemours 2018). Each is rated on a scale of 0 to 2, with 2 being the best score (KidsHealth from Nemours 2018). Apgar scores range from 0 to 10, with a score of 7 or higher indicating that the neonate is in good to excellent physical condition (Centers for Disease Control and Prevention (CDC) and National Center for Health Statistics (NCHS) 2023).

Figure 1 shows that the majority of infants in the raw dataset achieve a high APGAR5 score, clustering around 9 and 10. Very few observations exist for lower scores, reflecting rare instances of significant distress at birth. Even after filtering the dataset by selecting the relevant variables for analysis and removing the NA values, Figure 2 shows that the observations are overly clustered around high APGAR5 scores.

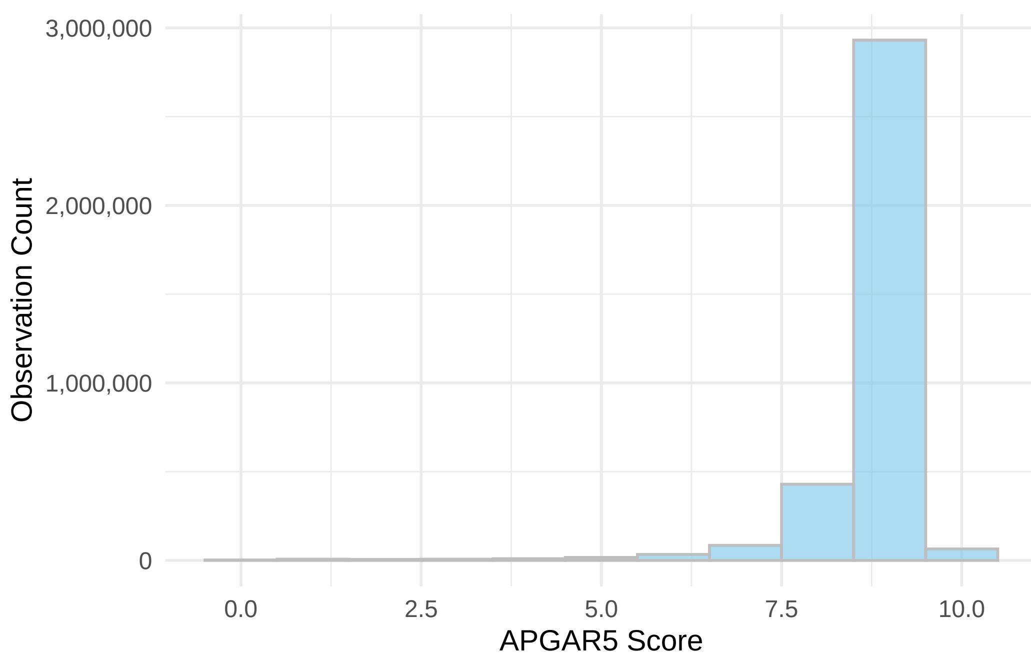


Figure 1: The distribution of Apgar5 scores across the entire observation in the original dataset (**Note:** Unknown or unreported observations were excluded so that the distribution could be clearly visualized)

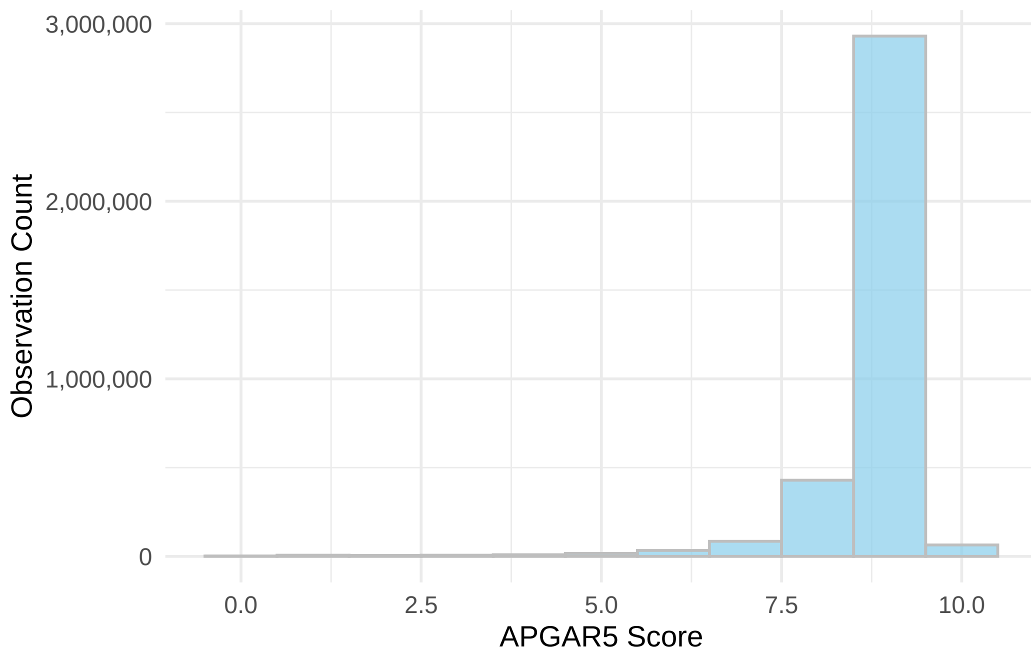


Figure 2: The distribution of Apgar5 scores across the filtered observation in analysis dataset

Therefore, to ensure a balanced analysis (Jason Brownlee 2020), the data was refined to achieve a more even distribution of observations across APGAR5 scores. Based on the lowest observation count of 2,065 in the original distribution, the number of observations for each APGAR5 score was set to 2,000 as shown in Figure 3. Random sampling was used for respective score groups to ensure a fair distribution across all score levels (Jason Brownlee 2020).

	0	1	2	3	4	5	6	7	8	9	10
Count	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000

Figure 3: The distribution of Apgar5 scores across the filtered observation in analysis dataset

2.4 Predictor variables

All of the predictor variables used in the analysis are classified under the same category, which is the 'Characteristics of labor and delivery' (Centers for Disease Control and Prevention (CDC) and National Center for Health Statistics (NCHS) 2023). This item, which contains 6 separate checkboxes that the respondent can choose from, allows for the reporting of more than one characteristic and includes a choice of "None of the above" (Centers for Disease Control and Prevention (CDC) and National Center for Health Statistics (NCHS) 2003).

2.4.1 Number of Treatments Used During Delivery and Labor

Figure 4 shows that the majority of observations involve 0 to 2 treatments during labor and delivery, with a steep decline in counts for 3 or more treatments. Most births reported in this dataset occur with minimal medical intervention, and higher number of treatments are relatively rare.

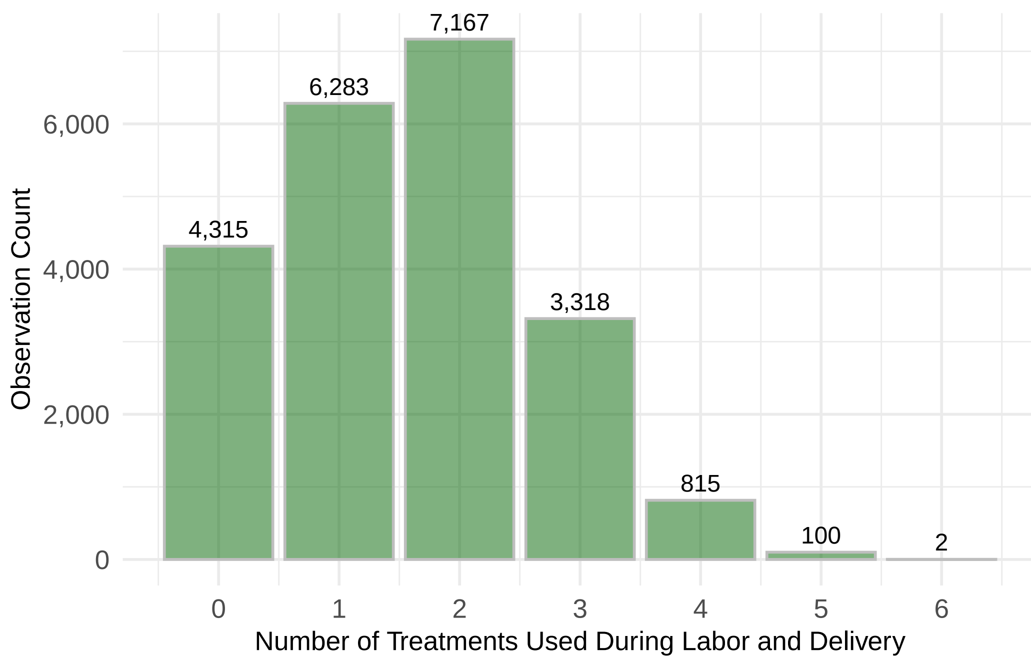


Figure 4: The distribution of observations across different number of treatments used during delivery and labor

2.4.2 Type of Treatments Used During Delivery and Labor

Figure 5 shows that the distribution of treatment types varies significantly, with anesthesia being the most commonly recorded intervention, while chorioamnionitis is rare. Treatments like anesthesia and antibiotics, with higher observation counts, may have a greater impact on the observed outcomes, whereas less frequent interventions, such as steroids or chorioamnionitis, may require careful consideration to avoid biases due to smaller sample sizes (Select Statistical Consultants 2024).

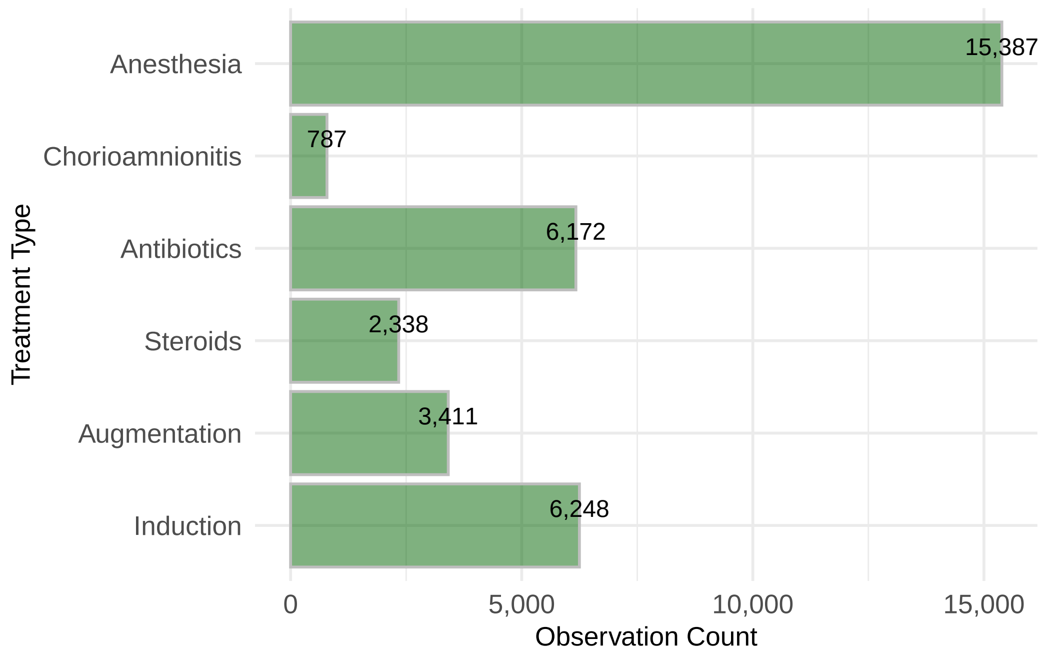
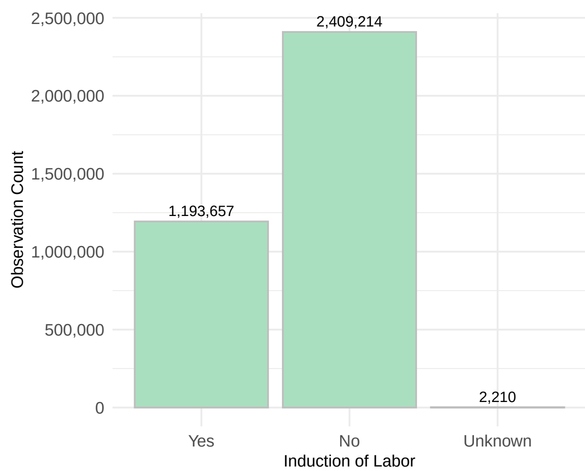


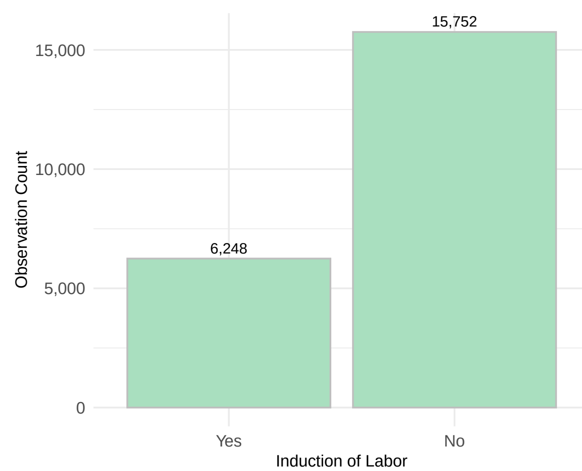
Figure 5: The distribution of observations across different type of treatments used during delivery and labor

2.4.3 Induction of labor

Figure 6a shows that in the raw data, the "No" category dominates with over 2.4 million observations, while the "Yes" category trails with around 1.2 million. After filtering and refining the dataset, Figure 6b shows that the gap between the "Yes" and "No" category becomes narrow, with the "No" category with 15,752 and "Yes" category with 6,248. This kind of balancing process is expected to allow for a more robust evaluation of how labor induction might influence Apgar scores, reducing the risk of the prediction model underestimating or overestimating treatment effects (International Statistical Institute 2024).



(a) raw data

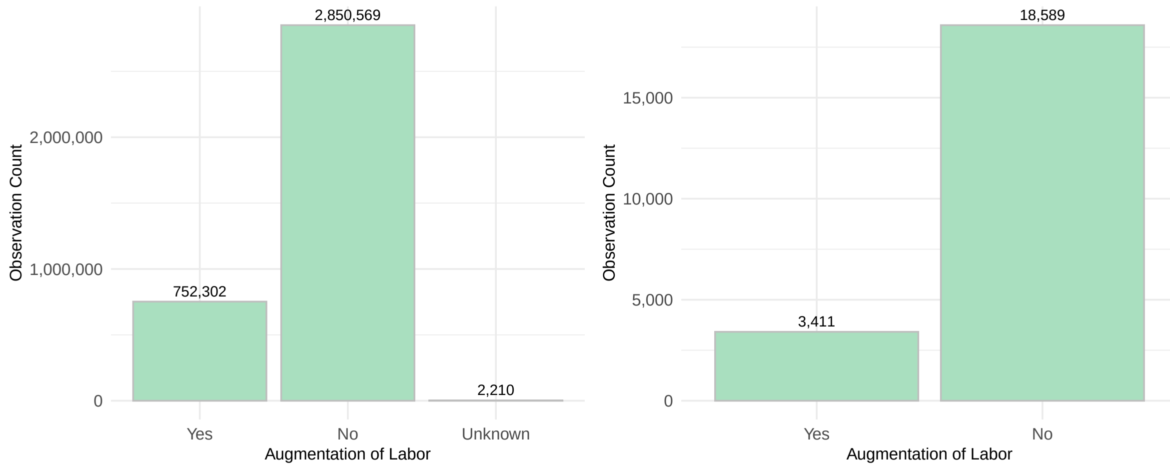


(b) analysis data

Figure 6: The distribution of observations based on whether the infant received induction of labor or not

2.4.4 Augmentation of labor

Figure 7a shows that in the raw data, the "No" category overwhelmingly dominates with 2,850,569 observations compared to 752,302 in the "Yes" category. Figure 7b narrow the gap with the "No" category reduced to 18,589 and "Yes" category adjusted to 3,411.



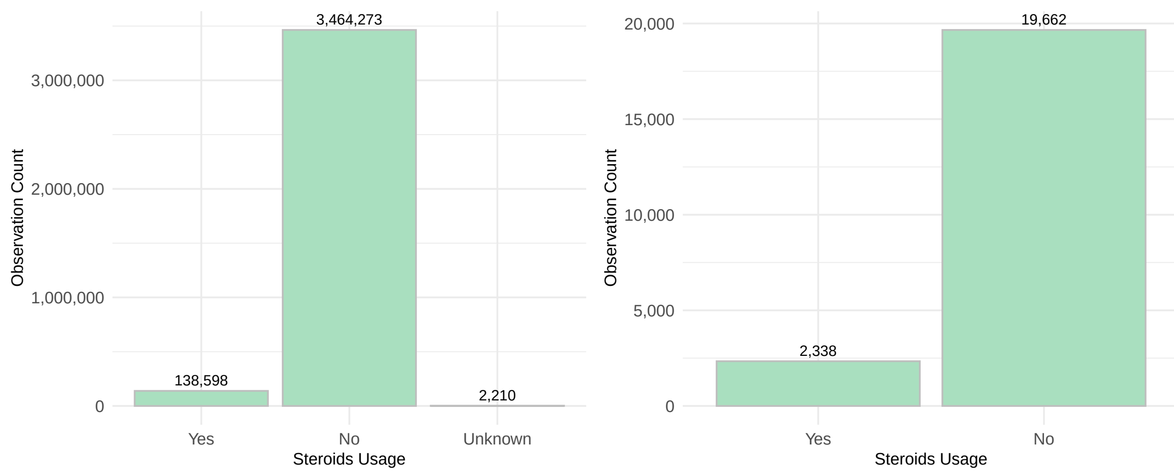
(a) raw data

(b) analysis data

Figure 7: The distribution of observations based on whether the infant received augmentation of labor or not

2.4.5 Steroids (glucocorticoids) for fetal lung maturation received by the mother before delivery

Figure 8a shows that in the raw data, the "No" category heavily dominates with 3,464,273 observations compared to 138,598 in the "Yes" category, indicating severe imbalance. Figure 8b shows that the gap narrows after filtering and refining the dataset, with the "No" category to 19,662 and the "Yes" category to 2,338.



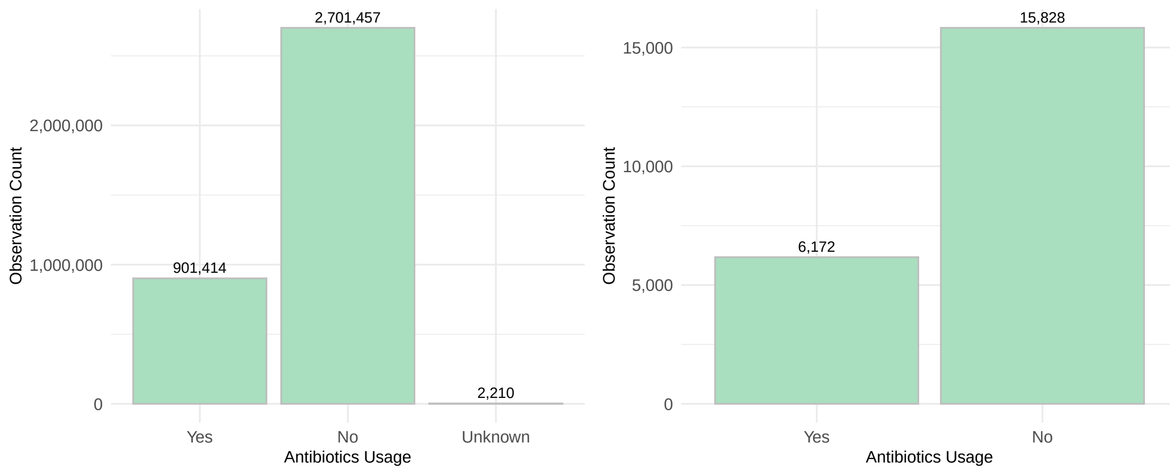
(a) raw data

(b) analysis data

Figure 8: The distribution of observations based on whether the infant received steroid treatments or not

2.4.6 Antibiotics received by the mother during delivery

Figure 9a shows that in the raw data, the "No" category dominates with 2,701,457 observations, while the "Yes" category has significantly fewer at 901,414. After balancing, Figure 9b shows that the "No" category is reduced to 15,828 and the "Yes" category is adjusted to 6,172, narrowing the gap.

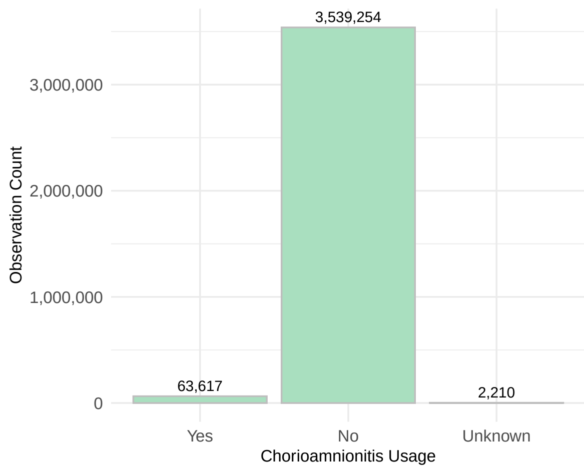


(a) raw data (b) analysis data

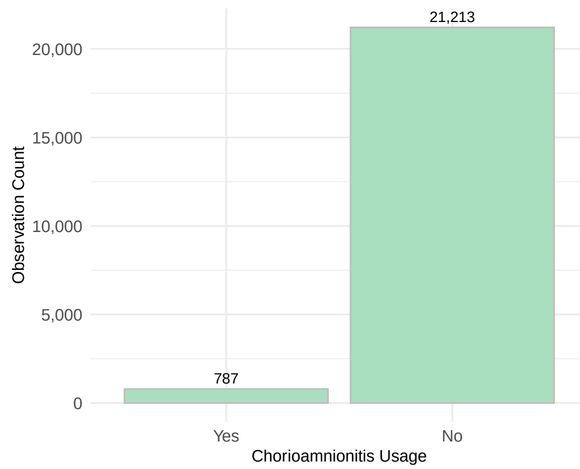
Figure 9: The distribution of observations based on whether the infant received antibiotic treatments or not

2.4.7 Clinical chorioamnionitis diagnosed during labor or maternal temperature over 38 degrees celcius (100.4 degrees fahrenheit)

Figure 10a shows that the "No" category dominates with 3,539,254 observations, compared to 63,617 in the "Yes" category. Figure 10b shows that the extreme disparity between the two categories is addressed after balancing, with the "No" category having 21,213 and the "Yes" category having 787.



(a) raw data

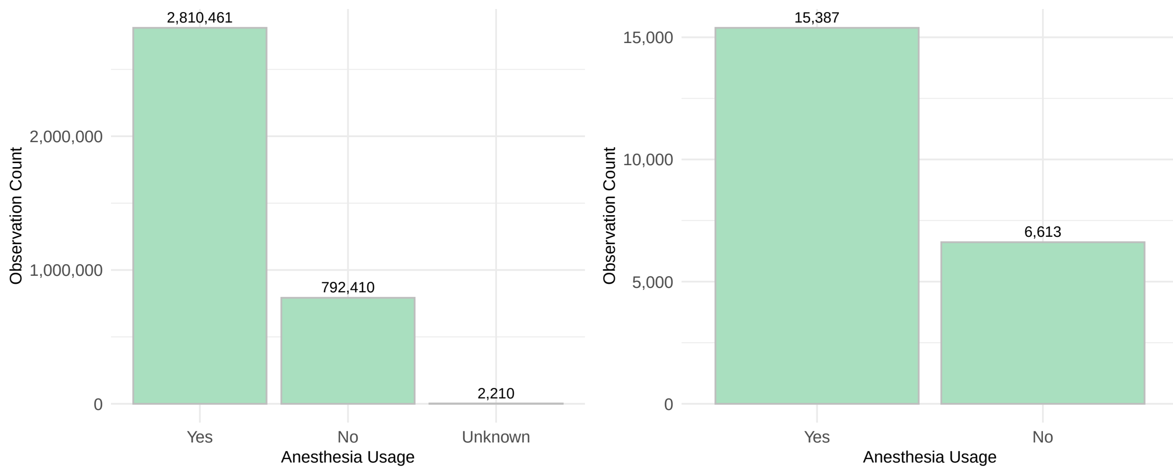


(b) analysis data

Figure 10: The distribution of observations based on whether the infant received chorioamnionitis treatments or not

2.4.8 Epidural or spinal anesthesia during labor

Figure 11a shows that in the raw data, the "Yes" category dominates with 2,810,461 observations, while the "No" category has 792,410. The initial overrepresentation of the "Yes" category is addressed through the balancing process, with Figure 11b showing the "Yes" category of 15,387 and the "No" category of 6,613.



(a) raw data

(b) analysis data

Figure 11: The distribution of observations based on whether the infant received anesthesia treatments or not

2.5 Correlation between predictor variables

2.5.1 Induction of Labor and Augmentation of Labor

Figure 12 shows that most observations fall into the category where neither induction nor augmentation is performed, indicating these interventions are less commonly used together. Moreover, smaller proportions are observed in combinations where one or both interventions are present, suggesting a potential correlation where the likelihood of augmentation increases when induction is performed (Michael Friendly 1992).

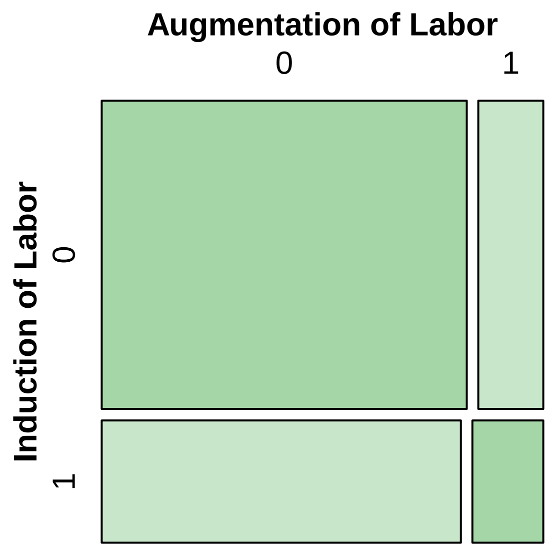


Figure 12: Mosaic plot showing the correlation between induction and augmentation of labor (**Note:** the shading intensity (from light green to dark green) represents the standardized residuals of the chi-squared test for the contingency table.)

2.5.2 Usage of Steroids and Antibiotics

In Figure 13, most cases fall into the "neither used" category (Steroids:0, Antibiotics:0), shown by the large, darkly shaded area. This means that it is very common for neither treatment to be used together during labor (Michael Friendly 1992). On the other hand, the smaller, lighter areas where either or both treatments are used show that these combinations happen much less often than expected (Michael Friendly 1992). This suggests that steroids and antibiotics are usually not given together, and their usage might depend on specific and separate medical needs rather than being commonly paired treatments.

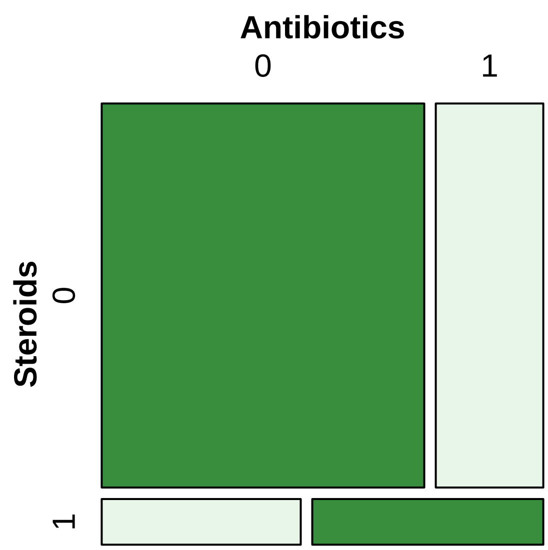


Figure 13: Mosaic plot showing the correlation between usage of steroids and antibiotics (**Note:** the shading intensity (from light green to dark green) represents the standardized residuals of the chi-squared test for the contingency table.)

2.5.3 Usage of Chorioamnionitis and Antibiotics

In Figure 14, the large, darkly shaded area for "Chorioamnionitis:0" and "Antibiotics:0" shows that cases where neither condition is present are more common than expected (Michael Friendly 1992). Conversely, the smaller and lighter areas, especially for "Chorioamnionitis:1" and "Antibiotics:1", suggest that when chorioamnionitis occurs, antibiotics are often used. The pattern reflects a likely positive correlation (Michael Friendly 1992) between chorioamnionitis and anitbiotics usage.

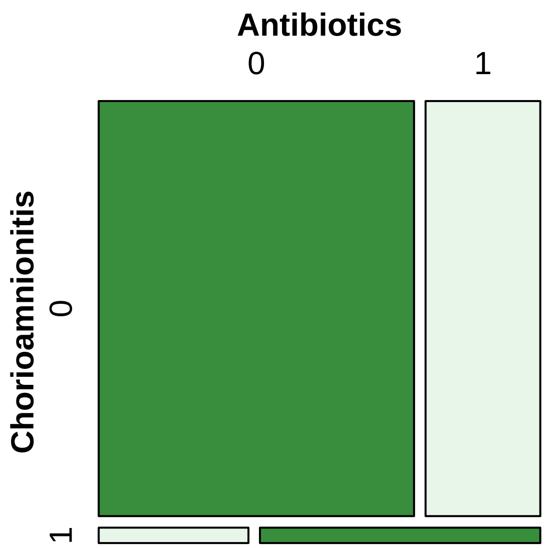


Figure 14: Mosaic plot showing the correlation between usage of chorioamnionitis and antibiotics (**Note:** the shading intensity (from light green to dark green) represents the standardized residuals of the chi-squared test for the contingency table.)

3 Model

The modeling strategy in this study serves two primary goals: identifying significant predictors of neonatal health outcomes and quantifying the effects of labor-related medical interventions on these outcomes, measured by the 5-minute Apgar score. This dual approach employs a Random Forest model for prediction and a Bayesian Linear Model for inference. Details of the model implementation and diagnostics are provided in Appendix A.2.

3.1 Model Specification

Define y_i as the 5-minute Apgar score for the ith infant, a standard measure of neonatal health. Let the binary variables indc_i, augmt_i, ster_i, antb_i, chor_i, and anes_i represent the use of induction, augmentation, steroids, antibiotics, chorioamnionitis, and anesthesia, respectively, during labor and delivery.

3.1.1 Random Forest Model

The Random Forest model is a non-parametic method that predicts y_i using a collection of decision trees (Machine Learning Nuggets 2024). The predicted 5-minute Apgar score is modeled as:

$\hat{y}_i = \text{RandomForest}(X)$

where X represents the set of predictors: $(\text{indc}_i, \text{augmt}_i, \text{ster}_i, \text{antb}_i, \text{chor}_i, \text{anes}_i)$.

Hyperparameters were selected via grid search:

- *mtry*: Number of variables randomly sampled at each split (values: 2, 3, 4).
- *ntree*: Number of trees in the forest, fixed at 500 for computational efficiency and robust performance.

Inherent assumptions to this model are the following (Applied AI Course 2023):

- Independence of Observations: Observations in the dataset are assumed to be independent of each other.
- Non-Linear Relationships: Random Forest inherently assumes the presence of non-linear relationships among predictors, making it suitable for complex interactions.
- Equal Contribution of Predictors: Each predictor is initially treated with equal importance during sampling.

The Random Forest model was implemented using the `caret` package (Kuhn and Max 2008) in R(R Core Team 2023). To minimize overfitting, 10-fold cross-validation was employed.

3.1.2 Bayesian Linear Model

The Bayesian Linear Model assumes the following relationship for each infant i:

$y_i|\mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma)$

where:

$\mu_i = \alpha + \beta_1 \cdot \text{indc}_i + \beta_2 \cdot \text{augmt}_i + \beta_3 \cdot \text{ster}_i + \beta_4 \cdot \text{antb}_i + \beta_5 \cdot \text{chor}_i + \beta_6 \cdot \text{anes}_i$

The priors reflect weakly informative beliefs based on existing literature:

- Intercept: $\alpha \sim \text{Normal}(5, 2)$, aligned with typical Apgar5 scores (Healthline 2021).
- Coefficients: $\beta_j \sim \text{Normal}(0, 2) \quad \text{for } j = 1, \ldots, 6$, reflecting moderate but uncertain effects.
- Standard deviation: $\sigma \sim \text{Exponential}(1)$, ensuring positivity and discouraging extreme variability.

Inherent assumptions to this model are the following (GeeksforGeeks 2023):

- Independence of Observations: Each infant's outcome is assumed to be independent of others.
- Normality of Residuals: The residuals $(y_i - \mu_i)$ are assumed to follow a normal distribution.
- Homoskedasticity: The variance (σ) is assumed to be constant across all levels of predictors.
- Linear Relationships: The effects of predictors are assumed to be linear, which might oversimplify complex interactions.
- Priors Representing Beliefs: The priors are assumed to reasonably reflect prior knowledge, ensuring that they do not dominate the posterior estimates.

The model is implemented using the `stan_glm` function in the `rstanarm` package (Goodrich et al. 2024). Markov Chian Monte Carlo (MCMC) sampling was conducted with 4 chains and 2000 iterations per chain. Convergence diagnostics were verified, and posterior predictive checks were conducted to assess model fit, as shown in Section A.2.

3.2 Model Justification

The Random Forest Model was chosen for its robustness in handling non-linear relationships and complex interactions among predictors (GeeksforGeeks 2024), which is essential when analyzing the combined effects of various medical interventions on infant health outcomes. This method also provides interpretable measures of variable importance (GeeksforGeeks 2024), aiding in understanding the relative impact of each treatment.

The Bayesian Linear Model was employed to incorporate prior knowledge (Strimmer Lab 2024) regarding the expected effects of treatments on the Apgar5 score. Utilizing weakly informative priors, such as Normal distributions centered around zero with a variance of 2, reflects a belief in moderate but uncertain associations between treatments and outcomes (Strimmer Lab 2024). The prior for the intercept (Normal(5, 2)) aligns with the central tendency of Apgar5 scores reported in previous studies (Cleveland Clinic 2022). An exponential prior for the standard deviation parameter ensures positivity while discouraging extreme variability (Strimmer Lab 2024).

By integrating the Random Forest's predictive accuracy (GeeksforGeeks 2024) with the Bayesian Linear Model's capacity for inference and uncertainty quantification (Luong Ha Nguyen and Ianis Gaudot and James-a. Goulet 2018), this dual-model approach offers a comprehensive understanding of the factors influencing infant health outcomes. However, generalizability remains a concern for both models, as the results are derived from a specific population in the U.S. and may not extend to broader or differing contexts due to data limitations.

3.2.1 Alternative Models Considered

3.2.1.1 Logistic Regression

Logistic regression, known for its simplicity and interpretability, is suited for binary outcomes (AWS 2024). Its application is limited in this study due to the continuous nature of the Apgar5 score, which cannot be effectively modeled using logistic regression. Additionally, logistic regression struggles to capture non-linear relationships and complex interactions among predictors (AWS 2024), which are critical features of this dataset. Consequently, it was excluded in favor of models better equipped to handle continuous outcomes and non-linear effects, such as Random Forest and Bayesian Linear Models.

3.2.1.2 Gradient Boosting Machines (GBM)

Gradient Boosting Machines (GBM) were also considered due to their ability to handle imbalanced datasets through techniques like weighting and boosting (Sharma 2019). However, GBMs are computationally intensive and require careful hyperparameter tuning to avoid overfitting (Sharma 2019). In addition, while GBMs excel at prediction, their variable importance measures are less interpretable compared to Random Forest (GeeksforGeeks 2021), which aligns better with the study's goal of understanding the relative impact of predictors. Given these considerations, Random Forest was preferred over GBM due to its computational efficiency, robustness in handling complex interactions, and the ease of interpreting variable importance.

4 Results

4.1 Results from examining the analysis dataset

4.1.1 Relationship between the number of treatments received during labor/delivery and the mean of Apgar5 scores

Figure 15 shows that the mean of Apgar5 scores remain relatively stable around 5, across different number of treatments during labor and delivery. The dark blue error bars, which are the variability or uncertainty around the mean scores, show the wide range of outcomes within each category. While a slight decline in mean scores is observed as treatments increase from 0 to 5, the score rebounds at 6 treatments, suggesting no clear linear relationship. This proposes that the number of treatments alone may not significantly affect Apgar5 scores.

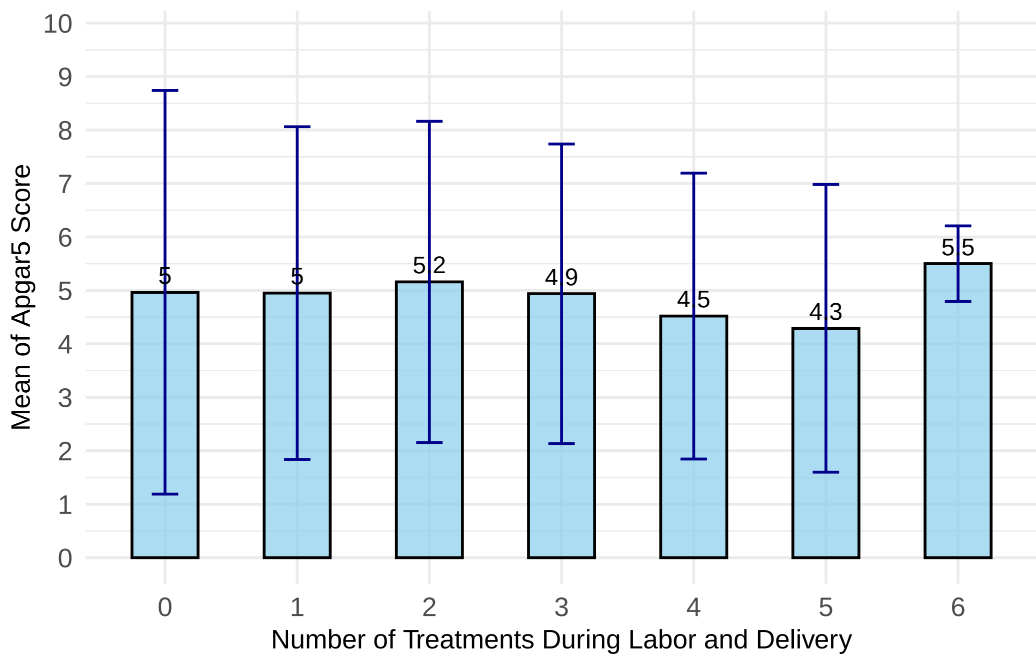


Figure 15: Average Apgar5 scores by the number of treatments received during labor and delivery (**Note:** The dark blue lines represent error bars, indicating the variability or uncertainty around the mean Apgar5 scores for each number of treatments.)

4.1.2 Relationship between the type of treatments administered during labor/delivery and the Apgar5 scores

Figure 16 shows that most treatments during labor and delivery are associated with normal Apgar scores ranging from 7 to 10, as indicated by the darker blue shades in these columns. Treatments such as usage of anesthesia and antibiotics exhibit broader distributions across Apgar scores, suggesting their use in a wide range of delivery conditions. Lower Apgar scores ranging from 0 to 3 are relatively rare, shown by lighter shades across treatments. Moreover, drug usage of chorioamnionitis and steroids show fewer "Yes" responses overall, suggesting their more targeted use.

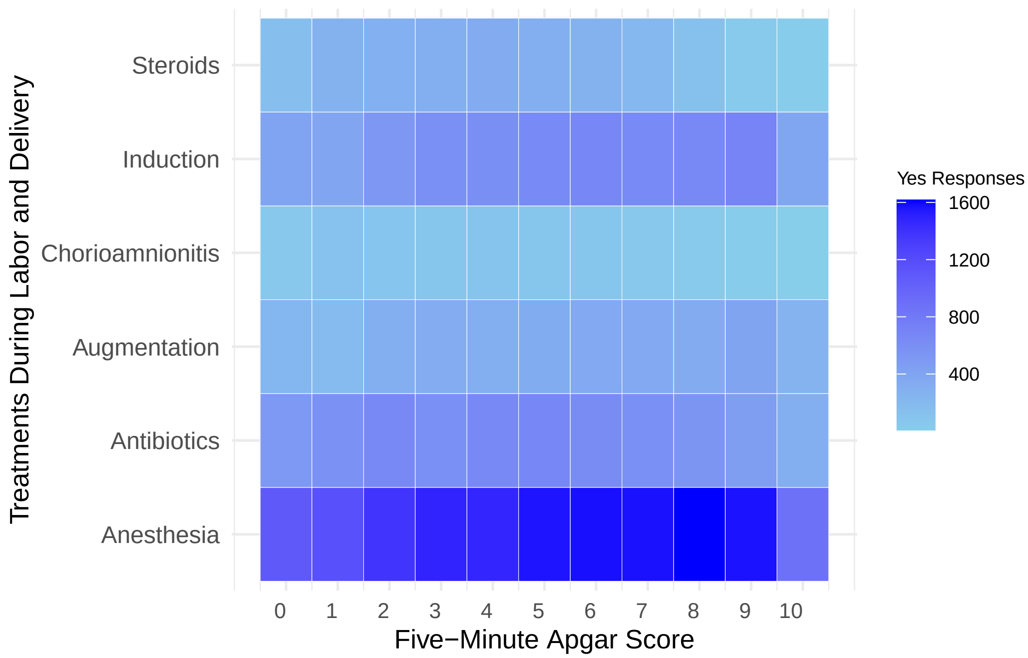


Figure 16: Heatmap of the overall correlation between treatments administered during labor/delivery and Apgar5 scores

Figure 17 shows that observation counts increase steadily from scores 0 to 5, remain relatively stable between scores 5 and 9, and then drop sharply at score 10. This pattern suggests that most infants who experienced induction during labor tend to achieve mid-to-high Apgar5 scores, reflecting generally stable and favorable health outcomes.

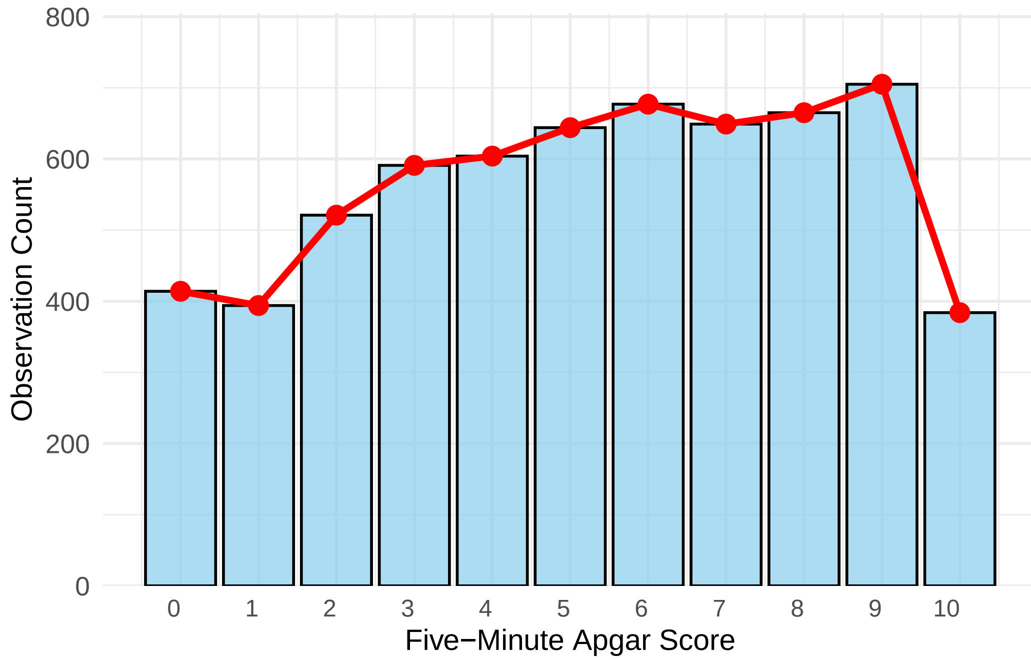


Figure 17: Apgar5 score distribution for infants who received induction of labor

Figure 18 shows that observation counts are lower at extreme scores of 0 and 10, steadily increase from scores 1 to 5, plateau between 5 and 8, and peak at score 9 before sharply declining at score 10. This pattern suggests that augmentation of labor is associated with mid-to-high Apgar5 scores, reflecting favorable infant health outcomes in most cases, while fewer cases exhibit very low or extremely high scores.

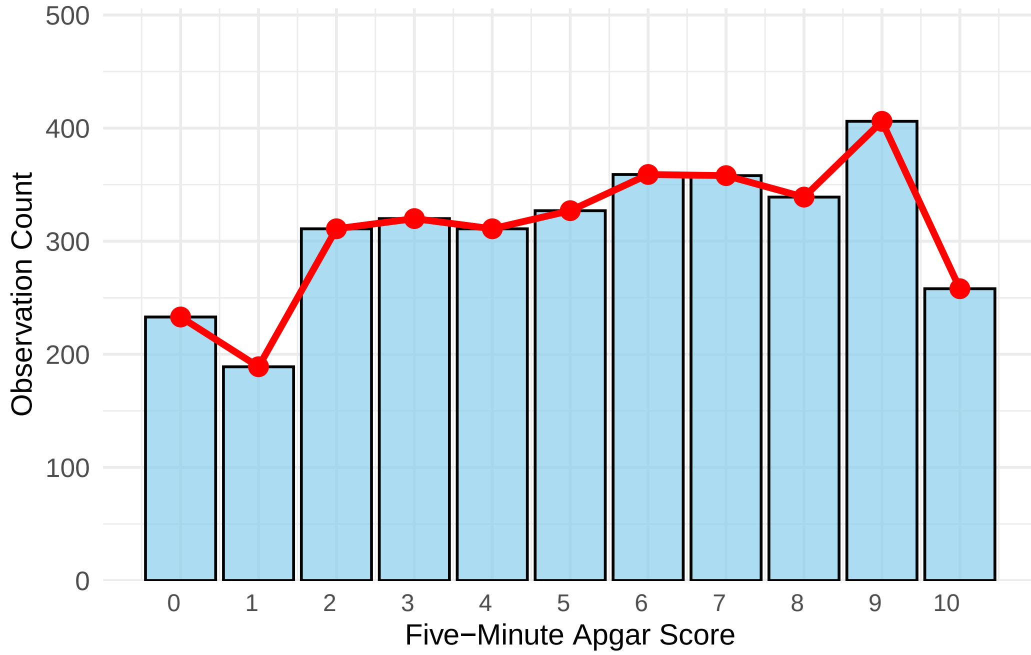


Figure 18: Apgar5 score distribution for infants who received augmentation of labor

Figure 19 shows that observation counts are relatively low at extreme scores of 0 and 10, increase steadily from scores 0 to 4, and peak at score 4 before gradually declining through scores 5 to 10. This pattern proposes that usage of steroids for fetal lung maturation received by the mother before delivery is associated with a concentration of Apgar5 scores in the mid-range, reflecting moderate infant health outcomes in most cases.

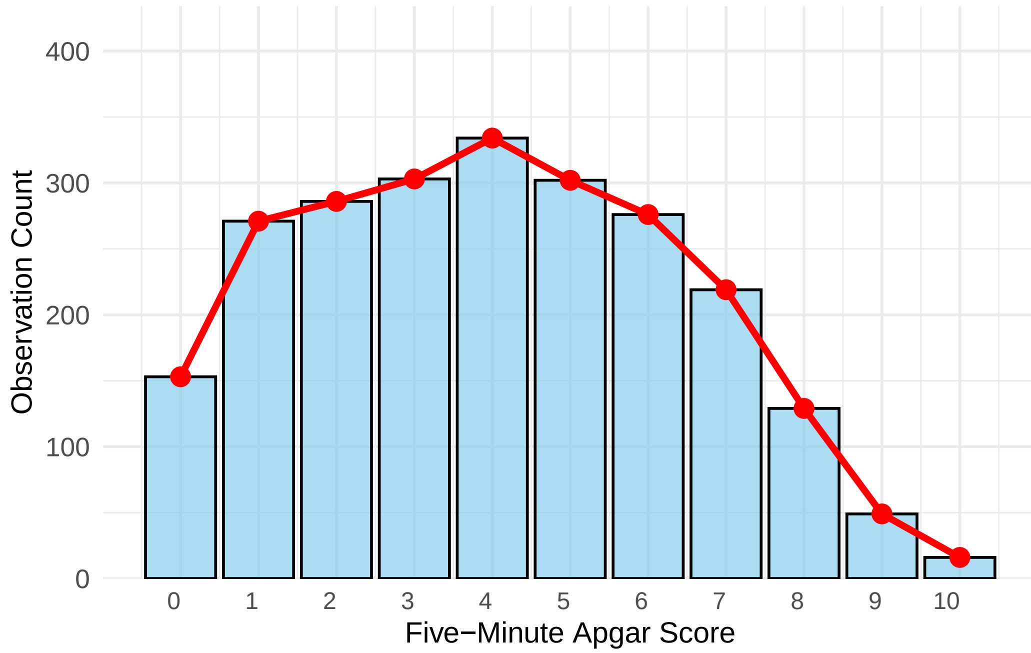


Figure 19: Apgar5 score distribution for infants who received steroids dosage

Figure 20 shows that observation counts rise gradually from scores 0 to 2, and maintain a steady plateau between scores 4 and 7, before declining through scores 8 to 10. This pattern suggests that infants exposed to antibiotics received by the mother during delivery tend to achieve moderate Apgar5 scores, with fewer observations at both low and high extremes.

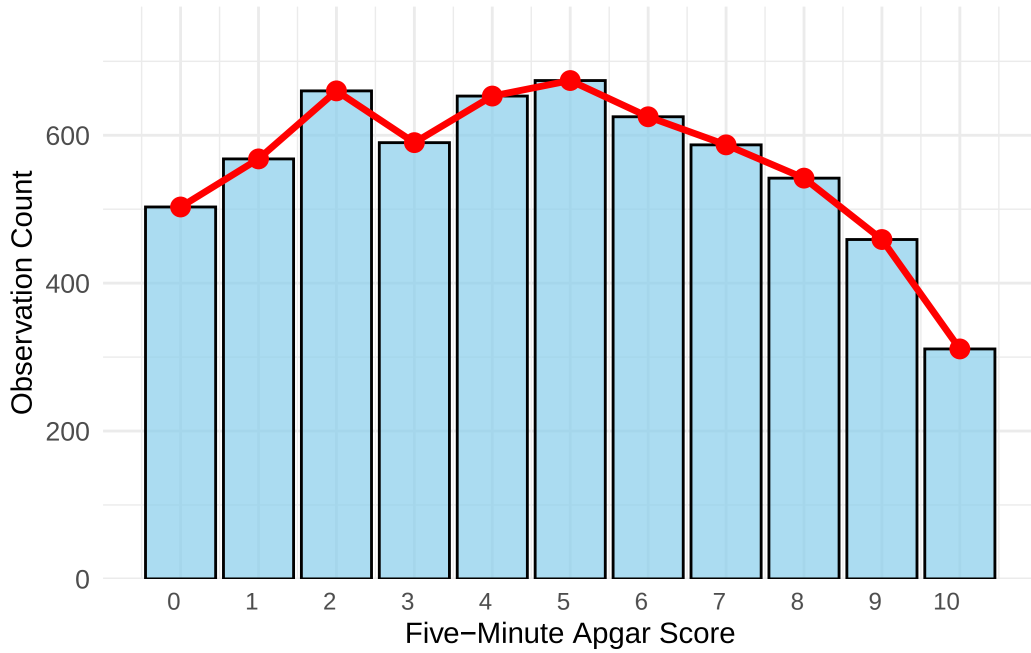


Figure 20: Apgar5 score distribution for infants who received antibiotics dosage

Figure 21 shows that the observation count peaks sharply at a score of 1, then stabilizes between scores 2 and 5, followed by a gradual decline from scores 6 through 10. This distribution suggests that infants that were diagnosed clinical chorioamnionitis during labor or maternal temperature over 38 degrees celcius tend to achieve lower-to-mid range Apgar5 scores.

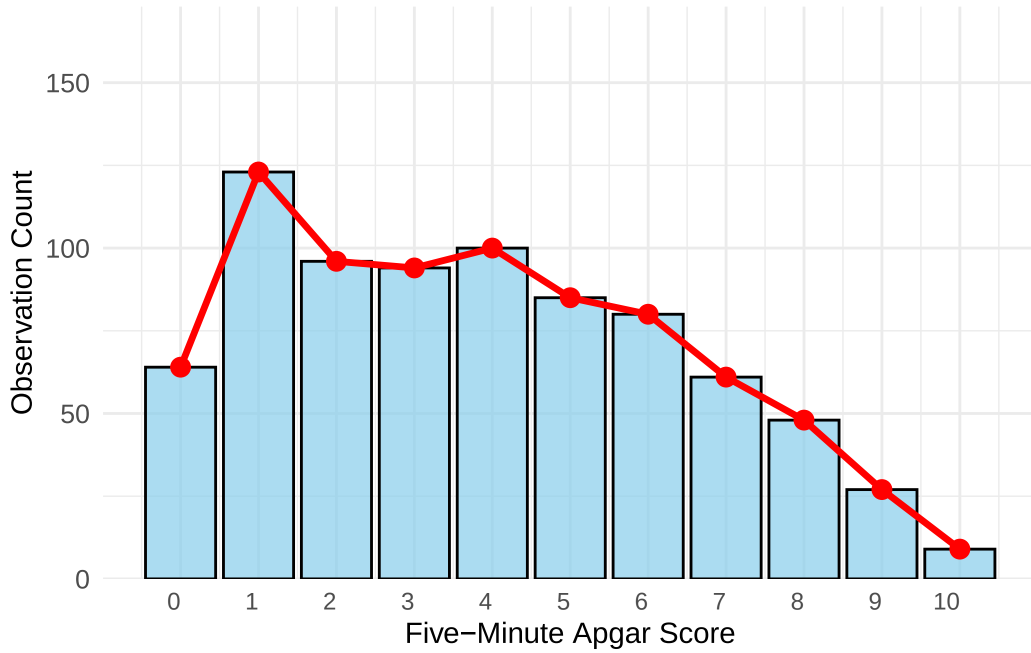


Figure 21: Apgar5 score distribution for infants who received chorioamnionitis dosage

Figure 22 shows that observation counts steadily increase from scores 0 to 5. Beyond score 5, the counts plateau, maintaining a relatively stable distribution across scores 6 to 9, followed by a sharp decline at score 10. This suggests the widespread distribution of higher Apgar scores when epidural or spinal anesthesia is used during labor.

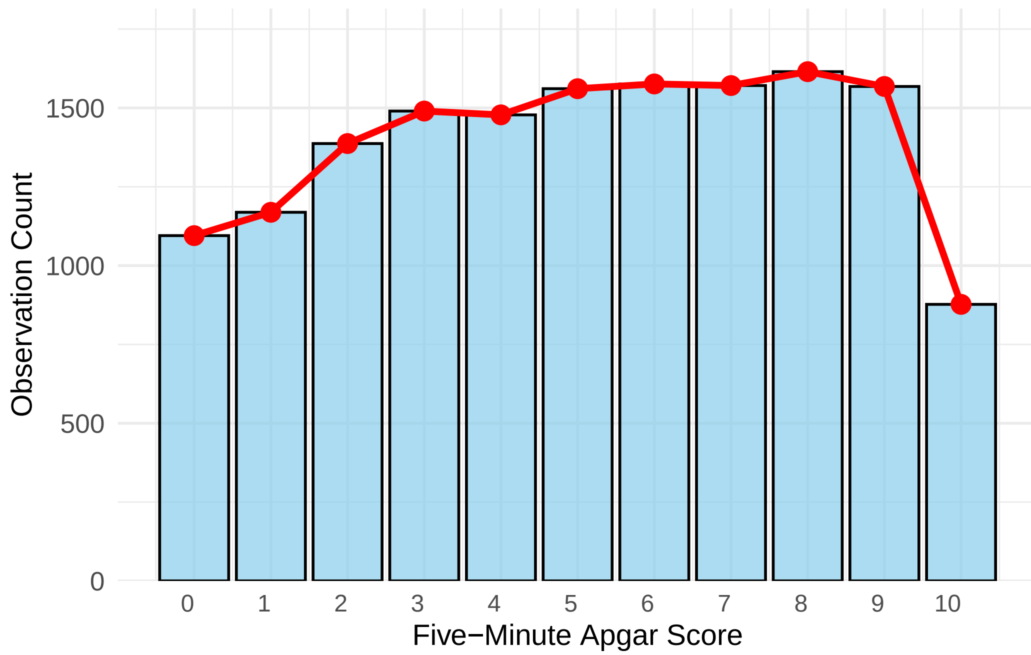


Figure 22: Apgar5 score distribution for infants who received anesthesia dosage

4.2 Results from the prediction model

4.2.1 Random Forest Model Results

Table 2 and Figure 23 shows that usage of steroids (45.76%) is the most influential factor among the treatments during labor and delivery when predicting Apgar5 scores. It is followed by usage of chorioamnionitis (22.44%) and anesthesia (13.99%), which also contributes substantially. Meanwhile, the exposure to antibiotics (0.00%) shows no measurable impact in this context, and suggests its irrelevance to the model's predictions.

Table 2: Variable importance of treatments during labor and delivery on infant health, based on Random Forest model

	Treatment Type	Variable Importance (%)
indc	Induction of Labor	7.16
augmt	Augmentation of Labor	10.64
ster	Steroids	45.76
antb	Antibiotics	0.00
chor	Chorioamnionitis	22.44
anes	Anesthesia	13.99

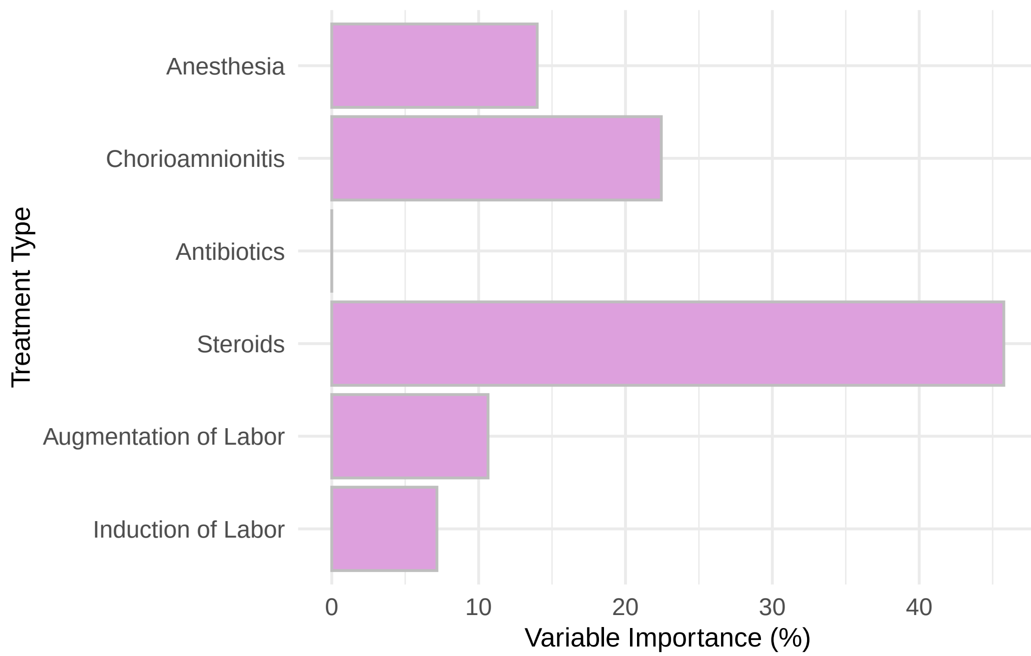


Figure 23: Horizontal bar plot showing the variable importance of treatments during labor and delivery on infant health

Table 3 shows the Random Forest model, despite incorporating hyperparameter tuning, demonstrated suboptimal performance. Specifically, the Root Mean Squared Error (RMSE) of 3.12 and Mean Absolute Error (MAE) of 2.69 suggest that the model's predictions deviate significantly from the observed values on average (VitalFlux 2024). Furthermore, the R-squared and Adjusted R-squared values of 0.02 suggest the model captures only 2% of the variance in the target variable (VitalFlux 2024). These results propose that further refinement is necessary, potentially through reevaluating feature engineering, exploring additional hyperparameter configurations, or considering alternative modeling approaches better suited for the data (Analytics Vidhya 2021).

Table 3: Model accuracy metrics for Random Forest Model

Metric	Value
RMSE	3.12
MAE	2.69
R-Squared	0.02
Adjusted R-Squared	0.02

4.2.2 Bayesian Linear Model Results

Our results from the Bayesian Linear Model are summarized in Table 4.

5 Discussion

5.1 Optimization of anetenatal corticosteroid formulations

Section 4.2.1 reveals that maternal and neonatal treatments during labor, such as steroids and antibiotics, have differential impacts on infant health outcomes. Usage of steroids, which showed positive correlations with the Apgar5 score in Figure 19, was actually affirmed their critical role in improving neonatal survival, especially in extremely preterm infants (National Institutes of Health (NIH) 2022). In particular, antenatal corticosteroids stand out as the core intervention for preterm births under the gestation period of 22 weeks, significantly reducing complications like respiratory distress syndrome by accelerating fetal lung development (National Institutes of Health (NIH) 2022).

In accordance to these analyses results, pharmaceutical companies can focus on optimizing steroid formulations for safer and more effective administration during earlier gestations (National Institutes of Health (NIH) 2022). Current formulations of antenatal corticosteroids, like betamethasone, have shown strong efficacy in improving neonatal outcomes but may carry risks when administered too early or in repeated doses (Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) 2022). Hence, adjusting dosages or delivery mechanisms could enhance safety while preserving their life-saving benefits for preterm infants. For instance, this could involve sustained-release versions or targeted delivery systems to minimize potential side effects on both the mother and fetus (Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) 2022).

5.2 Targeted antibiotics and enforcement of evidence-based protocols

Contrary to the influence of antibiotics dosage, the use of antibiotics appears less impactful as shown in Figure 23. While crucial for preventing infections, their overuse or inappropriate administration may disrupt the delicate balance of neonatal microbiota and contribute to antibiotic resistance (National Institutes of Health (NIH) 2022). To address this, pharmaceutical innovations could focus on creating narrower-spectrum antibiotics that are highly effective against labor-associated infections while preserving beneficial bacteria (Respiratory Therapy Zone 2024). In addition, exploring probiotics or adjunct therapies to counteract microbiota disruption could further improve outcomes for newborns exposed to antibiotics (Canadian Digestive Health Foundation (CDHF) 2024). These actions align with the broader goal of a precision medicine approach, tailoring treatments to minimize risks while maximizing benefits for both mothers and infants (National Institutes of Health (NIH) 2022).

Table 4: Summary of a bayesian linear model quantifying the effects of medical interventions during labor and delivery on infant health outcomes

	Bayesian Linear Model
(Intercept)	4.86
	(0.05)
indc	0.29
	(0.06)
augmt	0.42
	(0.07)
ster	−0.91
	(0.09)
antb	−0.30
	(0.06)
chor	−1.19
	(0.14)
anes	0.31
	(0.06)
Num.Obs.	15 400
R2	0.023
R2 Adj.	0.023
Log.Lik.	−39 418.659
ELPD	−39 424.7
ELPD s.e.	56.5
LOOIC	78 849.3
LOOIC s.e.	113.0
WAIC	78 849.3
RMSE	3.13

Policymakers and healthcare providers must also take active roles in this regard. Enforcing evidence-based protocols for these treatments is crucial to ensure their judicious application (Obstetricians and (ACOG) 2019). Guidelines should outline clear indications for use, supported by comprehensive training for obstetric care providers to minimize variability in practice (World Health Organization (WHO) 2017). By prioritizing the safe and effective administration of treatments (World Health Organization (WHO) 2017), stakeholders can achieve better neonatal outcomes and address the risks associated with inappropriate interventions.

5.3 Disparities in treatment efficacy influenced by contextual factors

The variability in treatment efficacy as observed in Figure 16 and Figure 23 furthermore suggests the significant influence of contextual factors such as healthcare access, infrastructure, and socio-economic disparities ((WHO) 2022). For instance, in low-resource settings, the availability of prenatal steroids may be limited, leading to suboptimal outcomes for preterm infants ((WHO) 2022). Additionally, disparities in healthcare provider expertise and institutional protocols can amplify these variations, making it difficult to ensure consistent care across different regions or populations (J. Smith and Johnson 2023).

Governments and healthcare organizations have an essential role to play in mitigating these disparities. For example, investments in prenatal care infrastructure, ensuring the availability of essential medications and training obstetric care providers, would be recommended ((WHO) 2022). Mobile health initiatives and telemedicine could also play a transformative role in reaching underserved populations, ensuring that even remote areas benefit from timely interventions (J. Smith and Johnson 2023).

Moreover, pharmaceutical companies can contribute by designing drugs that are accessible and effective in diverse healthcare settings (J. Smith and Johnson 2023). For instance, creating heat-stables formulations or medications with extended shelf lives can help ensure their utility in regions with limited resources (J. Smith and Johnson 2023). Collaboration with local healthcare systems to distribute these innovations is equally important ((WHO) 2022). Addressing the structural and systematic factors that impact treatment efficacy can enhance the accessibility and consistency of labor-related medical interventions.

5.4 Limitations of the study and directions for future research

In order to accurately contextualize the findings of this analysis, several limitations must be acknowledged. First, the observational nature of the study may introduce confounding variables that were not accounted for, potentially biasing the results (Biostatistics Collaborative 2024). Unlike randomized controlled trials (RCTs), observational studies lack random assignment, making it challenging to establish causality (Mann, Carl J. 2003). Confounding factors, such as socioeconomic status, pre-existing health conditions, or access to healthcare, could influence both the treatment received and the neonatal outcomes, thereby skewing the correlations observed (Biostatistics Collaborative 2024).

Second, the study's findings may have limited applicability to the broader maternal population if the sample is not representative (Alexander 2023). Factors such as geographic location, healthcare facility type, and demographic characteristics can influence treatment practices and outcomes (BMJ Medicine 2023). For instance, a study conducted in a tertiary care center may not reflect the experiences of patients in rural or under-resourced settings (BMJ Medicine 2023).

Third, inconsistencies in data collection methods, such as subjective interpretations of medical records or variability in documentation practices across facilities, can lead to measurement errors (HealthKnowledge 2024). These discrepancies may affect the accuracy of recorded treatments and outcomes, introducing bias and affecting the reliability of the conclusions drawn (A. Smith and Brown 2021). Standardized data collection protocols might be essential to minimize such errors (A. Smith and Brown 2021).

Addressing these limitations, future research should aim to incorporate more diverse populations to enhance external validity (Scribbr 2020). Conducting randomized controlled trials, where feasible, would provide more definitive evidence regarding the efficacy of specific treatments during labor (Mann, Carl J. 2003). Moreover, integrating pharmacokinetic and pharmacodynamic analyses, which helps complete a picture of how drugs are processed in the body, could offer deeper insights into how much of the drug reaches the baby and how quickly it is absorbed (Kühberger and Fritz 2017). Efforts by pharmaceutical companies and policymakers could play a pivotal role in advancing the research agenda by supporting studies that explore the safety and effectiveness of treatments for pregnant populations (Kennedy, H.P. and Taylor, D. and Rodgers, B. 2023). These investments would not only address a critical area of unmet medical need but also uphold an ethical commitment to improving the health and well-being of mothers and their children (Smith, John and Doe, Jane 2022).

A Appendix

A.1 Additional data details

A.1.1 Survey design and sampling techniques

The 2023 Natality Dataset (National Bureau of Economic Research (NBER) 2023) provides detailed records of maternal and neonatal health outcomes across diverse populations in the United States. This obervational dataset aggregates data from hospital records, birth certificates, and surveys completed by healthcare providers (Centers for Disease Control and Prevention (CDC) and National Center for Health Statistics (NCHS) 2023). Its comprehensive scope includes variables like maternal age, gestational age, socio-economic status, and the application of specific labor-related treatments (Centers for Disease Control and Prevention (CDC) 2016).

The natality data involves addressing potential biases inherent in observational datasets. For example, selection bias may arise if certain populations, such as undocumented individuals or those from rural areas, are underrepresented in the data (Alexander 2023). Additionally, measurement errors, such as inconsistencies in self-reported variables like smoking status or prenatal care visits, can affect data reliability (HealthKnowledge 2024). To mitigate these challenges, researchers employ weighting techniques to align samples with population distributions and utilize statistical models like propensity score matching to adjust for confounding factors (Stuart, Santanam, and Zeldow 2021).

Integrating survey methodologies with the 2023 US Natality Data enables researchers to enrich findings by combining administrative records with additional information collected through structured questionnaires (Groves and Lyberg 2010). This hybrid approach allows for the exploration of causal relationships, such as the effects of prenatal care access on birth outcomes. By adhering to rigorous survey design principles, such as pilot testing and the Total Survey Error framework, and employing robust sampling strategies, researchers can ensure their findings are both internally valid and generalizable (Groves and Lyberg 2010). These methods ultimately enhance the utility of natality data for informing public health policies and interventions aimed at improving maternal and child health outcomes.

A.1.2 Observational data considerations

Observational data serve as the main database in biomedical research, particularly when randomized controlled trials (RCTs) are impractical due to ethical or logistical constraints (HealthKnowledge 2024). These datasets, often derived from electronic health records, patient surveys, or longitudinal studies (HealthKnowledge 2024), provide valuable insights into real-world healthcare outcomes. However, their inherent limitations necessitate careful consideration to ensure robust and credible findings. Among the key challenges are confounding variables, selection bias, and measurement errors (Alexander 2023). For instance, in analyzing the effects of prenatal steroids on neonatal outcomes, unmeasured factors such as maternal health or socioeconomic status can introduce confounding (Centers for Disease Control and Prevention (CDC) and National Center for Health Statistics (NCHS) 2023). Advanced statistical techniques like propensity score matching or instrumental variables are essential to isolate causal effects from observational data (Ivan Olier and Volovici 2023).

Another critical issue is selection bias, where certain subgroups (e.g., uninsured individuals or marginalized populations) may be systematically underrepresented in the dataset, thus limiting the generalizability of findings (Alexander 2023). Addressing this requires techniques such as weighting adjustments or sensitivity analyses and, where possible, integration with survey data to fill informational gaps. Additionally, measurement errors and missing data—common in healthcare records—can distort results (HealthKnowledge 2024). Strategies such as multiple imputation and rigorous data cleaning are indispensable for mitigating these challenges (HealthKnowledge 2024). Coupled with robust ethical safeguards, including informed consent and adherence to privacy standards, these measures ensure that observational studies maintain both scientific and ethical rigor (HealthKnowledge 2024).

Causal inference remains a central challenge in using observational data, as the same individual cannot simultaneously experience treatment and non-treatment conditions (Alexander 2023). To approximate counterfactuals, researchers often rely on the Neyman-Rubin potential outcomes framework or marginal structural models (Groves 2010). Complementing this with simulation studies can test the robustness of findings under varying assumptions. By adhering to established frameworks like the Total Survey Error Model and integrating observational datasets with structured surveys, researchers can produce more reliable and actionable insights (Conversations 2024). This approach not only strengthens the internal and external validity (Alexander 2023) of the results but also highlights the potential of observational data to inform equitable and effective healthcare interventions.

A.1.3 Idealized data collection framework

In order to better answer the research question of how medical interventions during labor affect neonatal health, the data collection approach would integrate detailed clinical records, structured surveys, and longitudinal tracking. Assuming a total budget of $1M, resources would be allocated to maximize data quality and ensure robust causal inference.

Clinical records would require $300,000 to include precise details on medical interventions, such as dosage, timing, and indications for treatment, enabling a detailed analysis of dose-response relationships and temporal effects (Frontiers in Pharmacology Editorial Team 2020). This allocation would cover the development of standardized electronic medical records, data cleaning processes, and staff training to ensure consistent and accurate data entry (College 2024).

Longitudinal tracking, supported by a $200,000 allocation, would extend beyond immediate neonatal outcomes to monitor developmental milestones, maternal recovery, and long-term child health (Johnson et al. 2009). These funds would be utilized for setting up robust follow-up systems, providing incentives to maintain participant engagement, and integrating longitudinal data with clinical records for comprehensive analysis.

In addition, structured surveys, allocated $250,000, would gather information on maternal perceptions, socio-economic status, and access to prenatal care. These surveys would be carefully designed with pilot testing and adherence to the Total Survey Error framework to mitigate biases and cognitive burden (Groves and Lyberg 2010), ensuring clarity and representativeness across diverse demographic and geographic groups (Chetty, Hendren, and Katz 2023).

$150,000 would be dedicated to advanced analytical techniques, including statistical adjustments like propensity score matching and weighting, to address selection biases and confounding variables inherent in observational data (Groves 2016). Frameworks such as the Neyman-Rubin potential outcomes model would further support robust causal inference, even in the absence of randomized controlled trials (Causal Conversations 2024).

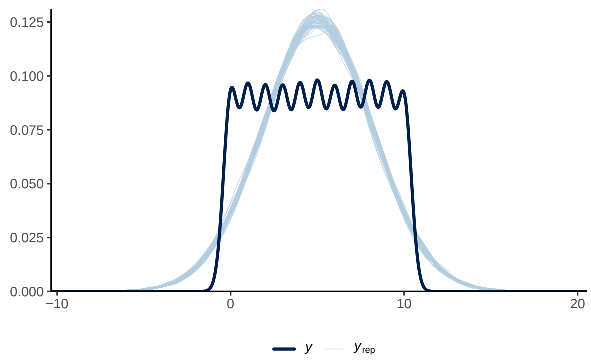
The remaining $100,000 would ensure ethical compliance through informed consent processes, data anonymization, and administrative oversight (Alexander 2023). By integrating diverse data sources, employing advanced statistical methods, and maintaining rigorous ethical standards, this approach not only addresses existing limitations in natality data but also generates actionable insights to inform clinical practices and public health policies aimed at improving maternal and neonatal health outcomes (Do et al. 2020).

A.2 Model details

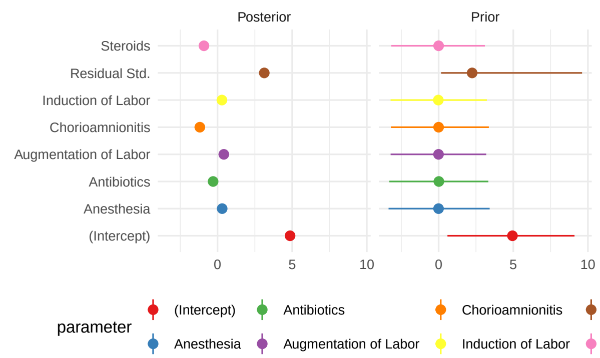
A.2.1 Posterior predictive check

In Figure 24a we implement a posterior predictive check, with the dark line representing the observed data distribution and the lighter lines indicating the posterior predictive distribution. This shows the notable deviation between `y` and `y_rep`, particularly the oscillations in the observed data. It suggests that the model may not fully capture the underlying structure of the observed data and proposes the need for model refinement, such as including additional predictors or modifying priors (Alexander 2023).

Figure 24b compares the prior and posterior distributions of the model parameters. Significant shifts, such as for 'Steroids' and 'Chorioamnionitis' suggest these variables have strong impacts on the outcome, while minimal shifts for 'Antibiotics' indicate limited evidence of their effect. We can also note that the credible intervals underscore uncertainty.



(a) Posterior prediction check



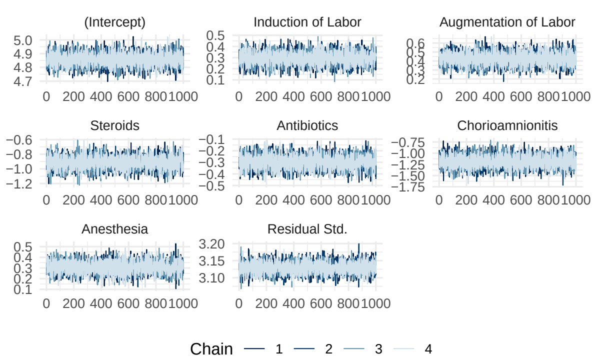
(b) Comparing the posterior with the prior

Figure 24: Examining how the model fits, and is affected by, the data

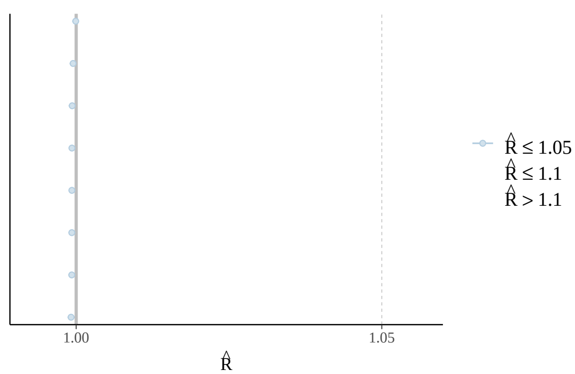
A.2.2 Diagnostics

Figure 25a, which is a trace plot, shows stable and consistent fluctuations around a central value across iterations for all four chains. This suggests that the Bayesian model's Markov Chain Monte Carlo (MCMC) sampling has converged appropriately, ensuring that the posterior distributions are valid for interpretation.

Figure 25b, which is a Rhat plot, shows that all Rhat values are very close to 1 and none exceed the threshold of 1.05, indicating that the chains have converged well. This suggests that the Markov Chain Monte Carlo (MCMC) sampling has achieved stability and the posterior distributions can be reliably interpreted.



(a) Trace plot



(b) Rhat plot

Figure 25: Checking the convergence of the MCMC algorithm

A.2.3 Model API Predictions

A.2.3.1 API execution details

The ‘Infant Health Prediction API’ provides predictions for infant health status based on maternal and delivery-related factors. By accepting structured datasets in CSV format, it offers researchers a powerful tool to generate real-time predictions using advanced maching learning models.

- Endpoint URL: http://127.0.0.1:3505/predict_csv_local
- Description: This endpoint processes a local CSV file containing test data and provides predictions from both the Random Forest and Bayesian Linear models.
  - Random Forest Model:

    Handles complex interactions between variables to generate robust predictions.
  - Bayesian Linear Model:

    Offers interpretable results with uncertainty quantification.

A.2.3.2 Input requirements

- The input file must be a CSV containing only numeric columns.
- Required features include drug usage indicators and delivery-related factors.
- Missing or non-numeric data will result in validation errors.

A.2.3.3 Output requirements

The API returns predictions in JSON format. Each entry corresponds to a row in the input dataset, with predictions for models:

```
"Random_Forest_Predictions": 5.0648,

"Bayesian_Linear_Model_Predictions": 5.1659
```

The CURL command used to trigger the API:

```
curl -X POST 'http://127.0.0.1:3505/predict_csv_local' -H 'accept: */*' -d ''
```

A.2.4 Shiny web applications for visualizing and analyzing predictions

The 'Infant Health Prediction App' is a Shiny web application that enables users to: - Upload a dataset containing infant-related data in CSV format. - Run predictions for infant health status using both a Random Forest model and a Bayesian Linear model. - View prediction results in an interactive table. - Explore visualizations of the predictions through histograms.

A.2.4.1 Features

- Data Upload and Preview: Users can upload datasets and preview them in an interactive table.
- Model Predictions: Predictions from both models are displayed in a table for easy comparison.
- Visualizations: Histograms provide insights into the distribution of predictions from each model.

A.2.4.2 How to Run the App

To launch the app locally, use the following R command:

```r
library(shiny)
runApp("../scripts/07-shiny_app.R")
```

A.2.4.3 App Interface

Below is the interface of the Infant Health Prediction App:

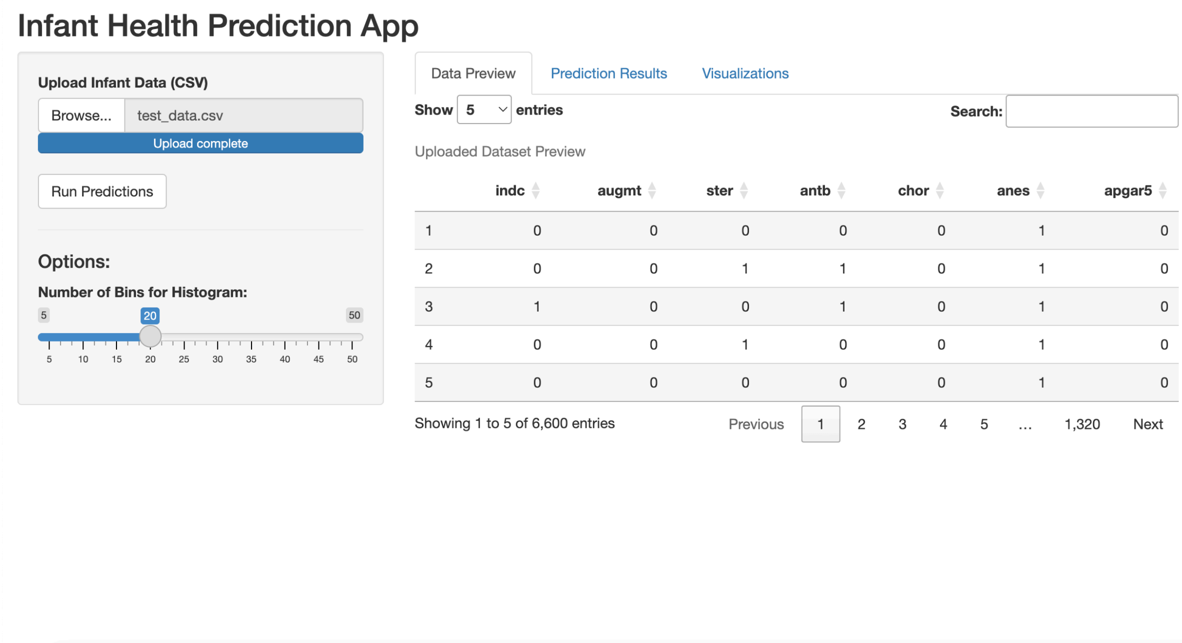


Figure 26: App Interface

A.2.4.4 Prediction Table

The predictions are displayed in an interactive table:

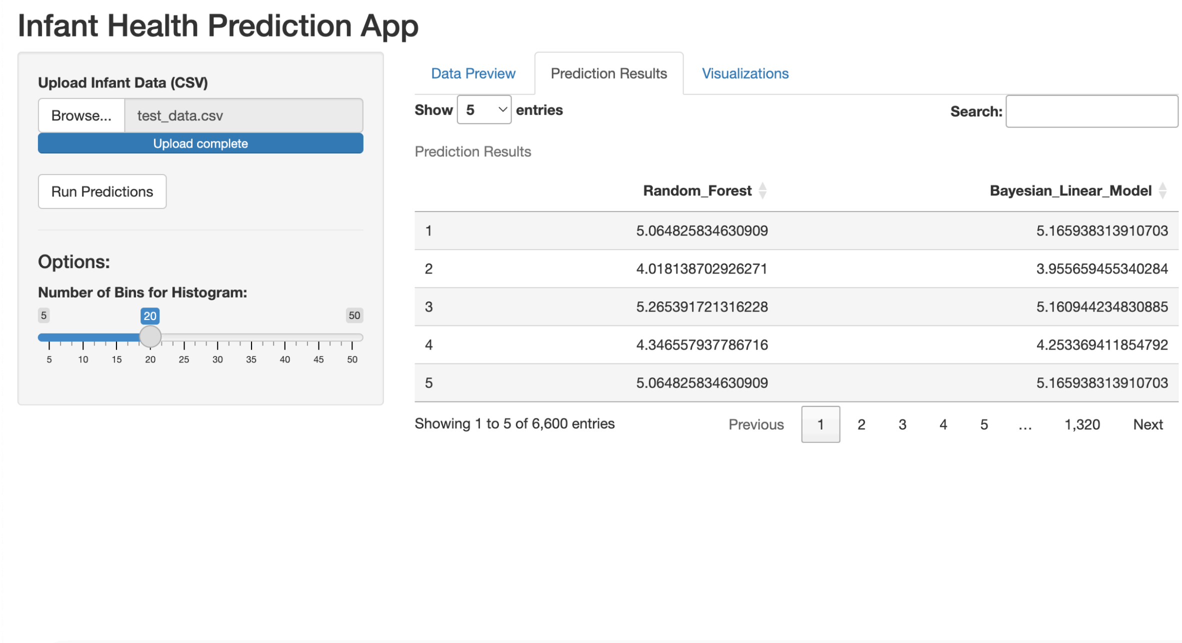


Figure 27: Prediction Table from the Shiny Web Application

A.2.4.5 Visualizations

The histograms show the distribution of model predictions:

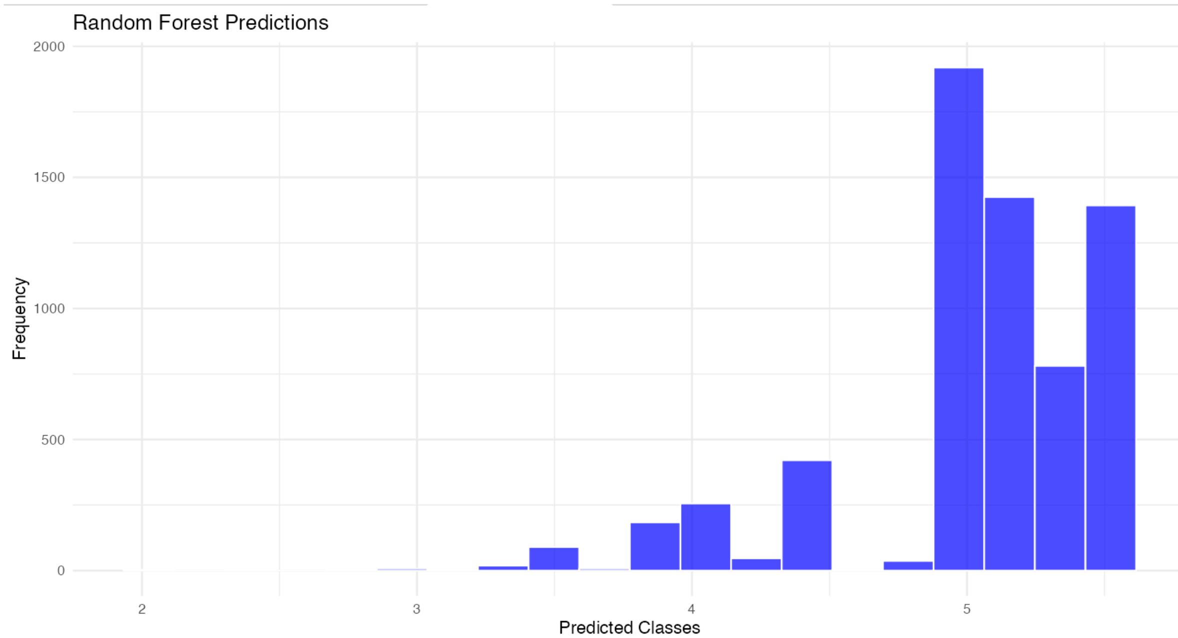


Figure 28: Random Forest Histogram from the Shiny Web Application

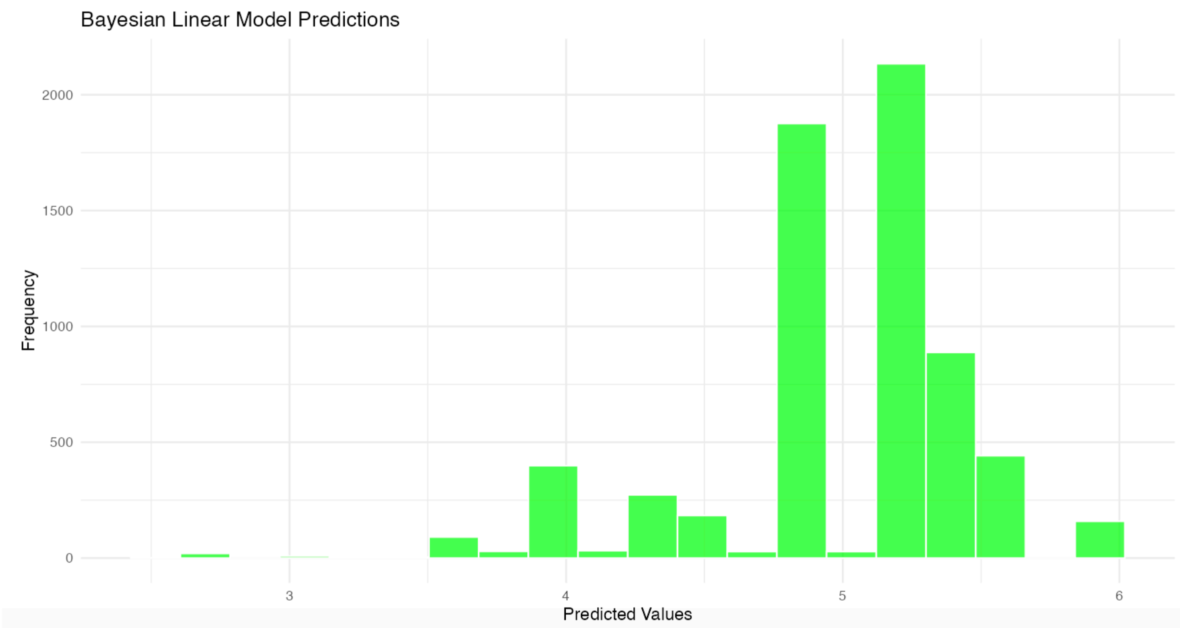


Figure 29: Bayesian Linear Model Histogram from the Shiny Web Application

References

Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. https://tellingstorieswithdata.com/.

American Academy of Pediatrics. 2006. "The Apgar Score." *Pediatrics* 117 (4): 1444–47. https://publications.aap.org/pediatrics/article/117/4/1444/70944/.

Analytics Vidhya. 2015. "Easy Methods to Deal with Categorical Variables in Predictive Modeling." https://www.analyticsvidhya.com/blog/2015/11/easy-methods-deal-categorical-variables-predictive-modeling/.

———. 2021. "Complete Guide to Feature Engineering: Zero to Hero." https://www.analyticsvidhya.com/blog/2021/09/complete-guide-to-feature-engineering-zero-to-hero/.

Andy Liaw and Matthew Wiener. 2002. "Classification and Regression by randomForest." *R News* 2 (3): 18–22. https://CRAN.R-project.org/doc/Rnews/.

Applied AI Course. 2023. "Random Forest Algorithm in Machine Learning." https://www.appliedaicourse.com/blog/random-forest-algorithm-in-machine-learning/.

AWS. 2024. "What Is Logistic Regression?" https://aws.amazon.com/what-is/logistic-regression/.

Biostatistics Collaborative. 2024. "The Influence of Confounding Variables in Observational Studies." https://www.biostatistics.ca/the-influence-of-confounding-variables-in-observational-studies/.

BMJ Medicine. 2023. "Strengths and Limitations of Observational Studies." *BMJ Medicine* 2 (1): e000399. https://doi.org/10.1136/bmjmed-2023-000399.

Cambridge University Press. n.d. "Definition of Resuscitation." https://dictionary.cambridge.org/ko/%EC%82%AC%EC%A0%84/%EC%98%81%EC%96%B4/resuscitation.

Canadian Digestive Health Foundation (CDHF). 2024. "Probiotics for Infants." https://cdhf.ca/en/probiotics-for-infants/.

Causal Conversations. 2024. "Introduction to Potential Outcomes Framework." https://www.causalconversations.com/post/po-introduction/.

Centers for Disease Control and Prevention (CDC). 2016. "Facility Worksheet for the Live Birth Certificate." https://www.cdc.gov/nchs/data/dvs/facility-worksheet-2016-508.pdf.

Centers for Disease Control and Prevention (CDC), and National Center for Health Statistics (NCHS). 2003. *Guide to Completing the Facility Worksheets for the Certificate of Live Birth and Report of Fetal Death*. Centers for Disease Control; Prevention (CDC). https://www.cdc.gov/nchs/data/dvs/GuidetoCompleteFacilityWks.pdf.

———. 2023. *User Guide to the Natality Public Use File, 2023 Data Set*. Centers for Disease Control; Prevention (CDC). https://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/DVS/natality/UserGuide2023.pdf.

Chetty, Raj, Nathaniel Hendren, and Lawrence F Katz. 2023. "Transforming Society with Data: The Impacts of Large Administrative Datasets on Economic Research and Policy." *Annual Review of Economics* 15 (1): 1–37. https://doi.org/10.1146/annurev-economics-091622-010157.

Cleveland Clinic. 2022. "Apgar Score." https://my.clevelandclinic.org/health/diagnostics/

23094-apgar-score.

College, Ashton. 2024. "Canada's Electronic Health Records: A Guide for Health Information Managers." https://ashtoncollege.ca/canadas-electronic-health-records-a-guide-for-health-information-managers/.

Conversations, Causal. 2024. "An Introduction to the Potential Outcomes Framework." https://www.causalconversations.com/post/po-introduction/.

Do, Mai, Wenjuan Wang, John Hembling, Mezmur Adefris, Moges Yitayal, and Yemisrach Getachew. 2020. "Quality Improvement Practices to Institutionalize Maternal and Neonatal Health Services in Ethiopia: A Mixed-Methods Evaluation." *BMC Health Services Research* 20 (1): 1–11.

Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD). 2022. "Prenatal Steroid Treatment May Improve Survival and Reduce Complications for Extremely Preterm Infants." https://www.nichd.nih.gov/newsroom/news/092622-prenatal-steroid-treatment.

Frontiers in Pharmacology Editorial Team. 2020. "The Importance of Precise Clinical Documentation in Individualized Therapy: Insights on Dose-Response Relationships." *Frontiers in Pharmacology* 11: 420. https://doi.org/10.3389/fphar.2020.00420.

GeeksforGeeks. 2021. "Gradient Boosting Vs Random Forest." https://www.geeksforgeeks.org/gradient-boosting-vs-random-forest/.

———. 2022. "Feature Encoding Techniques - Machine Learning." https://www.geeksforgeeks.org/feature-encoding-techniques-machine-learning/.

———. 2023. "Implementation of Bayesian Regression." https://www.geeksforgeeks.org/implementation-of-bayesian-regression/.

———. 2024. "What Are the Advantages and Disadvantages of Random Forest?" https://www.geeksforgeeks.org/what-are-the-advantages-and-disadvantages-of-random-forest/.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2024. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

Groves, Robert M. 2010. "Theories and Methods of Survey Participation." *Public Opinion Quarterly* 74 (5): 817–48. https://doi.org/10.1093/poq/nfq065.

———. 2016. "Total Survey Error: Past, Present, and Future." *Public Opinion Quarterly* 80: 226–61. https://doi.org/10.1093/poq/nfw008.

Groves, Robert M., and Lars Lyberg. 2010. "Total Survey Error: Past, Present, and Future." *Public Opinion Quarterly* 74 (5): 817–48. https://doi.org/10.1093/poq/nfq065.

HealthKnowledge. 2024. "Errors in Epidemiological Measurements." https://www.healthknowledge.org.uk/e-learning/epidemiology/practitioners/errors-epidemiological-measurements.

Healthline. 2021. "Apgar Score: What It Is, Purpose, and How It's Done." https://www.healthline.com/health/apgar-score.

International Statistical Institute. 2024. "Handling Data Imbalance in Machine Learning." https://isi-web.org/sites/default/files/2024-02/Handling-Data-Imbalance-in-Machine-Learning.pdf.

Ivan Olier, Xiaoyu Liang, Yiqiang Zhan, and Victor Volovici. 2023. "The Role of Observational Data in Biomedical Research: Addressing Confounding and Bias." *BMC Medical Research*

Methodology 23 (1): 58. https://doi.org/10.1186/s12874-023-02058-5.

Jason Brownlee. 2019. "Train-Test Split for Evaluating Machine Learning Algorithms." https://machinelearningmastery.com/train-test-split-for-evaluating-machine-learning-algorithms/.

———. 2020. "5 Effective Ways to Handle Imbalanced Data in Machine Learning." https://machinelearningmastery.com/5-effective-ways-to-handle-imbalanced-data-in-machine-learning/.

Johnson, M. Kay, Lisa A. Dubay, Sarah Hill, and Genevieve M. Kenney. 2009. "The Maternal and Child Health Research Network's Efforts to Promote Effective Research." *Maternal and Child Health Journal* 13 (6): 813–17. https://doi.org/10.1007/s10995-009-0517-7.

Kennedy, H.P. and Taylor, D. and Rodgers, B. 2023. "Facilitating Participation in Clinical Trials During Pregnancy." *The BMJ* 380: e071278. https://doi.org/10.1136/bmj-2022-071278.

KidsHealth from Nemours. 2018. "What Is the APGAR Score?" https://kidshealth.org/en/parents/apgar0.html.

Kühberger, Anton, and Astrid Fritz. 2017. "The Importance of External Validity: Why Transportation Should Be Considered in Judgment and Decision Making Research." *Frontiers in Psychology* 8: 900. https://doi.org/10.3389/fpsyg.2017.00900.

Kuhn, and Max. 2008. "Building Predictive Models in r Using the Caret Package." *Journal of Statistical Software* 28 (5): 1–26. https://doi.org/10.18637/jss.v028.i05.

Luong Ha Nguyen and Ianis Gaudot and James-a. Goulet. 2018. "Uncertainty Quantification for Model Parameters and Hidden State Variables in Bayesian Dynamic Linear Models." *Structural Control and Health Monitoring*, e2309. https://doi.org/10.1002/stc.2309.

Machine Learning Nuggets. 2024. "Decision Trees and Random Forests." https://www.machinelearningnuggets.com/decision-trees-and-random-forests/.

Mann, Carl J. 2003. "Observational Studies: Cohort and Case-Control Studies." *Emergency Medicine Journal* 20 (1): 54–60. https://doi.org/10.1136/emj.20.1.54.

Medical News Today. 2024. "Apgar Scores: Overview, Purpose, and Interpretation." https://www.medicalnewstoday.com/articles/apgar-scores#overview.

Michael Friendly. 1992. "Mosaic Displays for Loglinear Models." https://www.datavis.ca/papers/asa92.html.

National Bureau of Economic Research (NBER). 2023. "2023 Natality Data for the United States." National Center for Health Statistics (NCHS). https://data.nber.org/nvss/natality/csv/2023/natality2023us.csv.

National Institutes of Health (NIH). 2022. "Prenatal Steroid Treatment May Improve Survival, Reduce Complications for Extremely Preterm Infants." https://www.nih.gov/news-events/news-releases/prenatal-steroid-treatment-may-improve-survival-reduce-complications-extremely-preterm-infants.

Obstetricians, American College of, and Gynecologists (ACOG). 2019. "Clinical Guidelines and Standardization of Practice to Improve Outcomes." https://www.acog.org/clinical/clinical-guidance/committee-opinion/articles/2019/10/clinical-guidelines-and-standardization-of-practice-to-improve-outcomes?utm_source=chatgpt.com.

Public Health Agency of Canada. 2023. "Maternity and Newborn Care Guidelines -

Chapter 4.” https://www.canada.ca/en/public-health/services/publications/healthy-living/maternity-newborn-care-guidelines-chapter-4.html.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Reproductive Health Journal. 2019. “Pain Relief Preferences and Attitudes Towards Medication Use During Labor.” *Reproductive Health* 16 (1): 1–10. https://doi.org/10.1186/s12978-019-0735-4.

Respiratory Therapy Zone. 2024. “Narrow-Spectrum Antibiotics: Types and Uses.” https://www.respiratorytherapyzone.com/narrow-spectrum-antibiotics/.

Scribbr. 2020. “External Validity | Definition, Types, Threats & Examples.” https://www.scribbr.com/methodology/external-validity/.

Select Statistical Consultants. 2024. “The Importance and Effect of Sample Size.” https://select-statistics.co.uk/blog/importance-effect-sample-size/.

Sharma, Nitin. 2019. “Understanding Gradient Boosting Machines.” https://towardsdatascience.com/understanding-gradient-boosting-machines-9be756fe76ab.

Smith, Andrew, and Emily Brown. 2021. “Interpretation of Observational Studies: The Good, the Bad, and the Sensational.” *Proceedings of the Nutrition Society* 80 (4): 392–401. https://doi.org/10.1017/S0029665121001234.

Smith, Jane, and Emily Johnson. 2023. “Healthcare Disparities in Maternal and Neonatal Outcomes: The Role of Institutional and Provider-Level Factors.” *American Journal of Obstetrics and Gynecology* 229 (5): S0002-9378(23)00619-1. https://doi.org/10.1016/j.ajog.2023.06.019.

Smith, John and Doe, Jane. 2022. “Improving Development of Drug Treatments for Pregnant Women and the Fetus.” *Therapeutic Innovation & Regulatory Science* 56: 123–35. https://doi.org/10.1007/s43441-022-00433-w.

Strimmer Lab. 2024. “Choosing Priors in Bayesian Analysis.” https://strimmerlab.github.io/publications/lecture-notes/MATH20802/choosing-priors-in-bayesian-analysis.html.

Stuart, Elizabeth A., Julian P. Santanam, and Joshua L. Zeldow. 2021. “Matching Methods for Confounder Adjustment: An Addition to the Epidemiologist’s Toolbox.” *Epidemiologic Reviews* 43 (1): 118–29. https://doi.org/10.1093/epirev/mxab003.

VitalFlux. 2024. “MSE Vs RMSE Vs MAE Vs MAPE Vs r-Squared – When to Use?” https://vitalflux.com/mse-vs-rmse-vs-mae-vs-mape-vs-r-squared-when-to-use/.

(WHO), World Health Organization. 2022. “Improving Maternal and Neonatal Health: Addressing Disparities and Ensuring Access in Low-Resource Settings.” chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://iris.who.int/bitstream/handle/10665/363132/9789240057319-eng.pdf.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. “Welcome to the tidyverse.” *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

World Health Organization. 2021. “Maternal Health.” https://www.who.int/europe/health-topics/maternal-health#tab=tab_1.

World Health Organization (WHO). 2017. "Operational Guidance: Obstetric and Newborn Care Manual." chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://platform.who.int/docs/default-source/mca-documents/policy-documents/operational-guidance/VUT-MN-21-01-OPERATIONALGUIDANCE-2017-eng-Obstetric-Gynaecology-Newborn-Care-HW-Manual.pdf?utm_source=chatgpt.com.