How Drug Usage During Labor and Delivery Affects Infant Health: A Predictive Analysis of APGAR5 Scores*

!!! MAIN RESULTS !!!

Yunkyung Ko

November 25, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

Table of Contents

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

1 Introduction

Overview paragraph

Estimand paragraph

Results paragraph

Why it matters paragraph

Telegraphing paragraph: The remainder of this paper is structured as follows. Section 2....

2 Data

2.1 Overview

We use the '2023 Natality Data for the United States' to analyze the relationship between drug treatments administered to infants and their health outcomes. Provided by NBER and sourced from the National Center for Health Statistics (National Bureau of Economic Research (NBER) 2023), 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 the statistical programming language R (R Core Team 2023) and 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 n.d.). 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 n.d.).

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 n.d.). Each is rated on a scale of 0 to 2, with 2 being the best score (KidsHealth from Nemours n.d.). 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.

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).

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) n.d.).

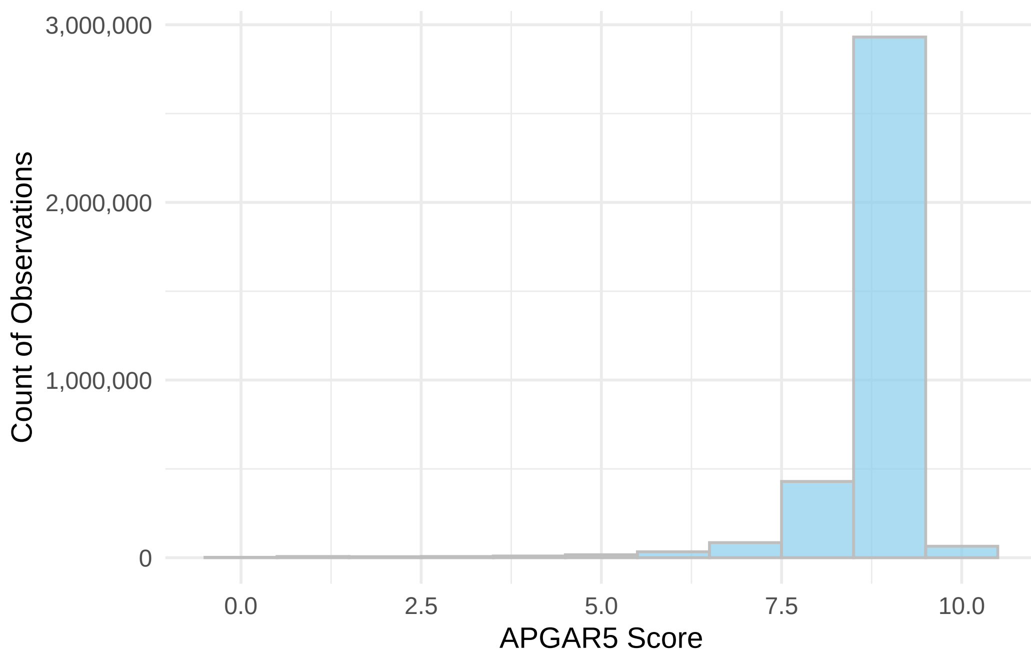


Figure 1: The distribution of APGAR5 scores across the entire observation in the original dataset **Note:** Unknown or unreported observations were excluded so that the distribution could be clearly visualized

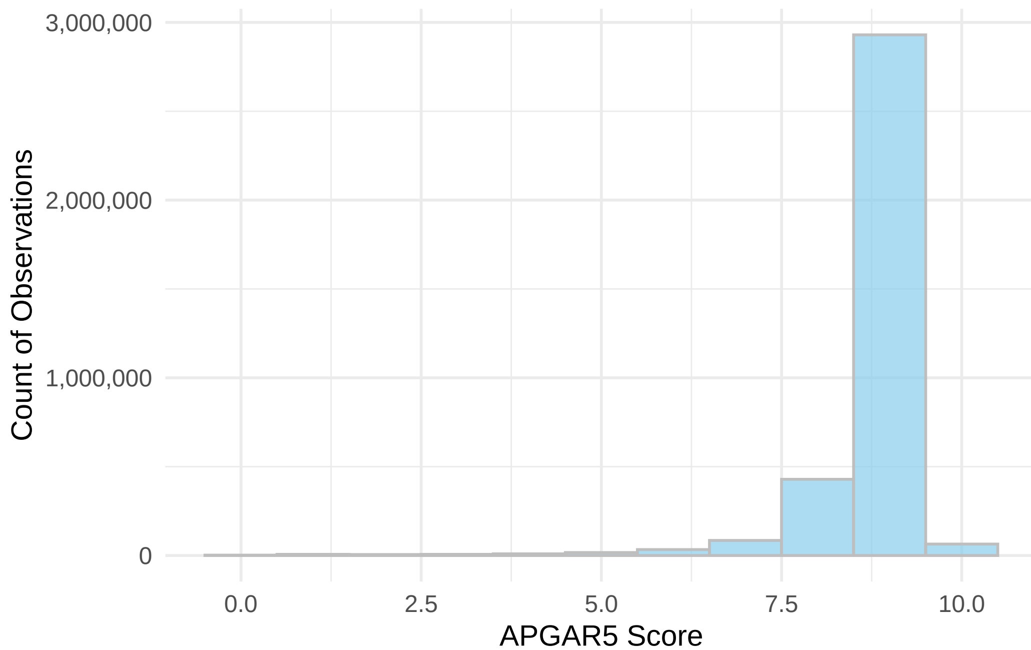


Figure 2: The distribution of APGAR5 scores across the filtered observation in analysis dataset, which selected the relevant variables and removed the NA values

	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

2.4.1 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.

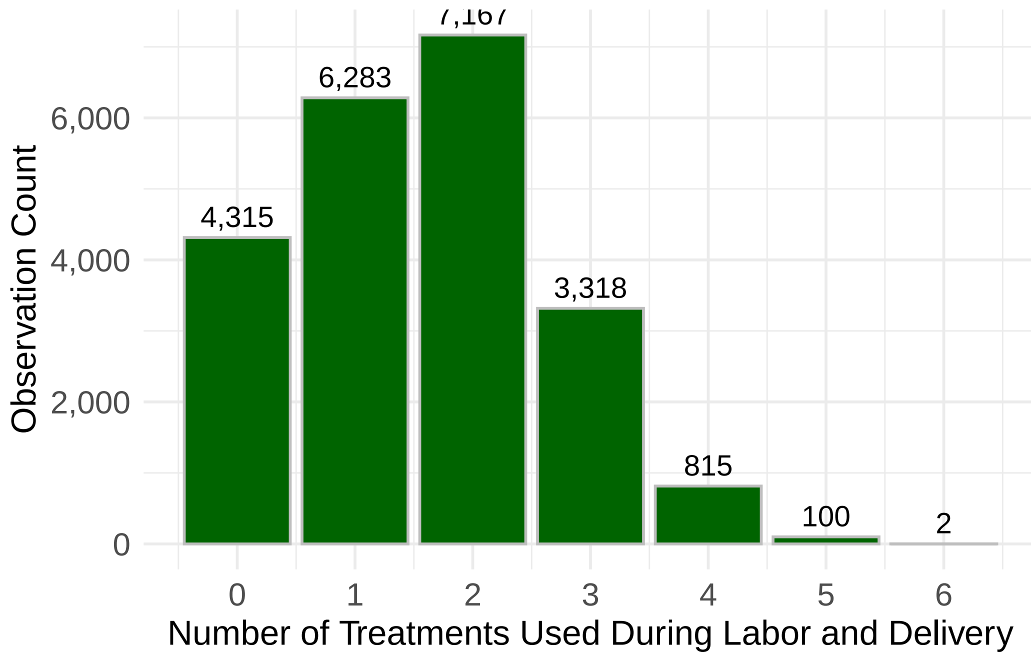


Figure 4: The distribution of observations across different number of treatments used during delivery and labor

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).

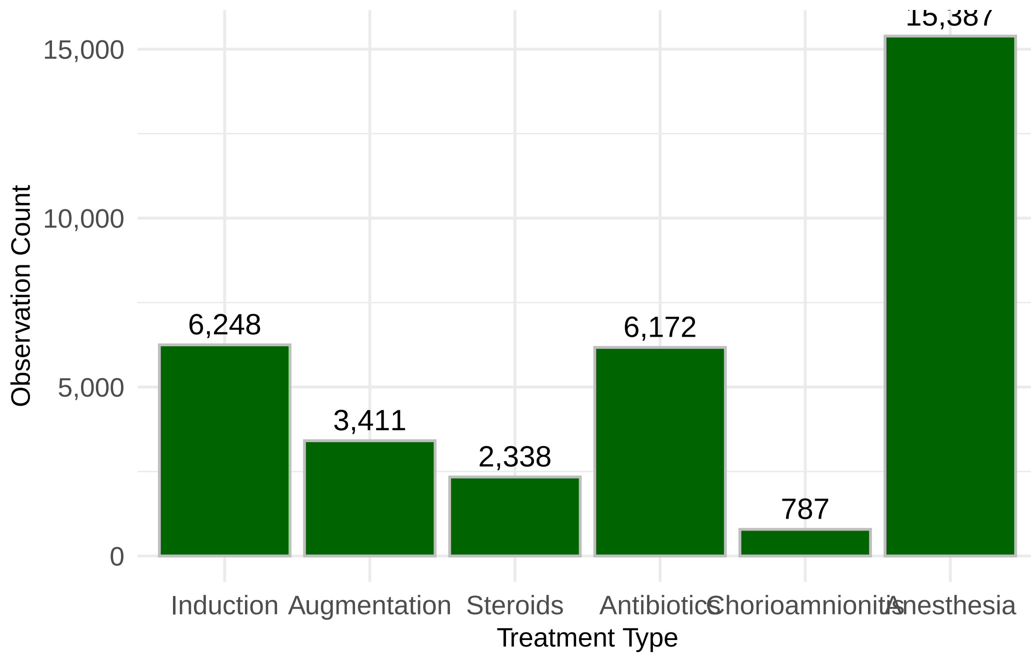


Figure 5: The distribution of observations across different type of treatments used during delivery and labor

2.4.3 Induction of labor

Figure 6 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 7 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).

2.4.4 Augmentation of labor

Figure 8 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 9 narrow the gap with the "No" category reduced to 18,589 and "Yes" category adjusted to 3,411.

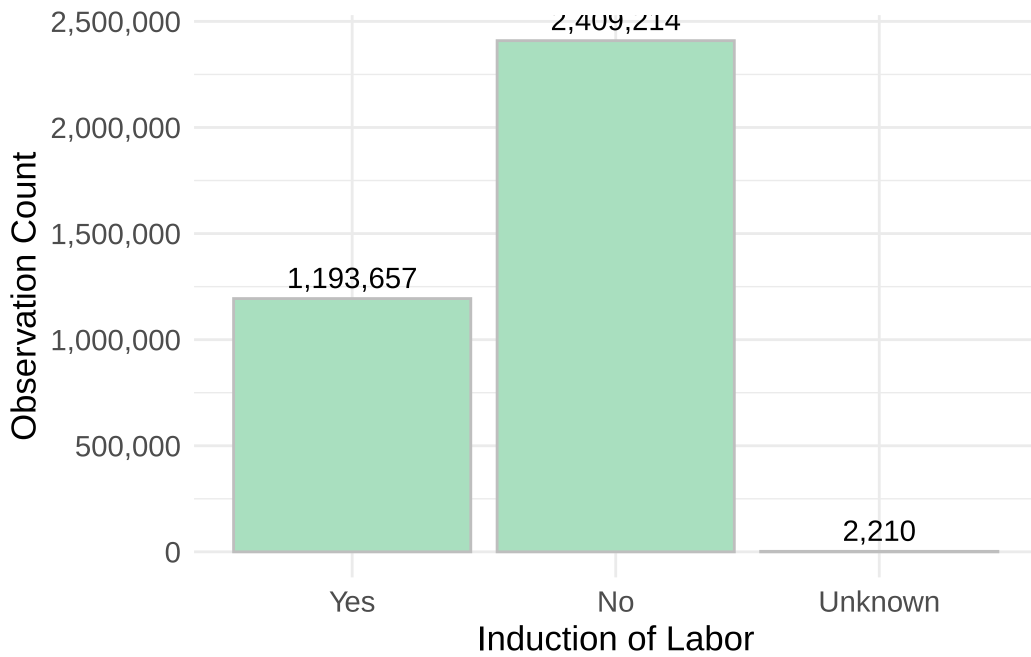


Figure 6: The distribution of observations in the raw dataset based on whether the infant received induction of labor or not

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

Figure 10 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 11 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.

2.4.6 Antibiotics received by the mother during delivery

Figure 12 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 13 shows that the “No” category is reduced to 15,828 and the “Yes” category is adjusted to 6,172, narrowing the gap.

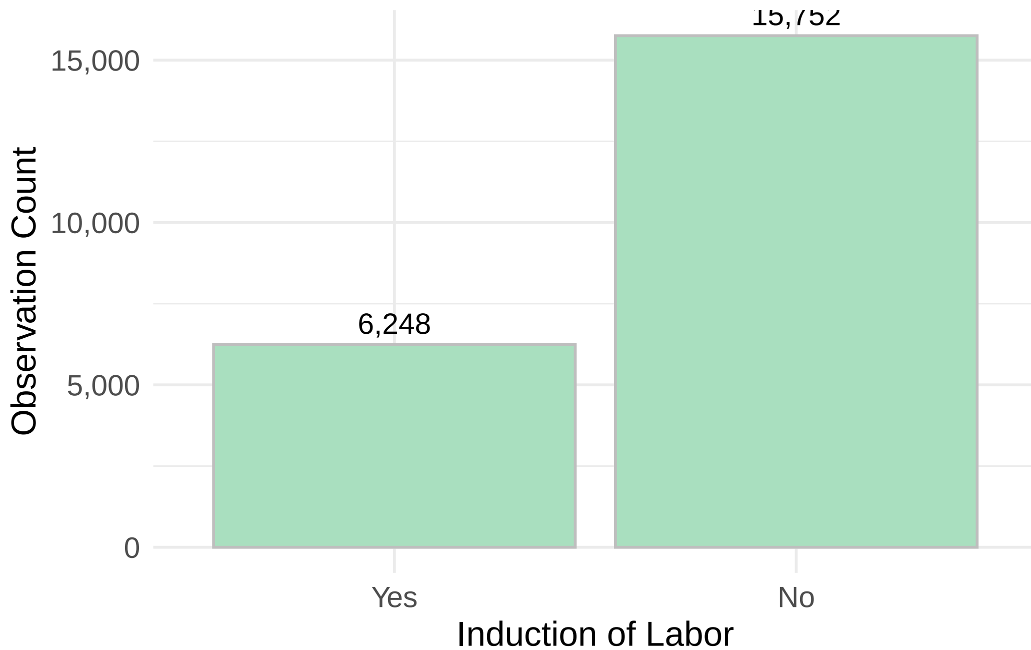


Figure 7: The distribution of observations in the analysis dataset based on whether the infant received induction of labor or not

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

Figure 14 shows that the "No" category dominates with 3,539,254 observations, compared to 63,617 in the "Yes" category. Figure 15 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.

2.4.8 Epidural or spinal anesthesia during labor

Figure 16 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 17 showing the "Yes" category of 15,387 and the "No" category of 6,613.

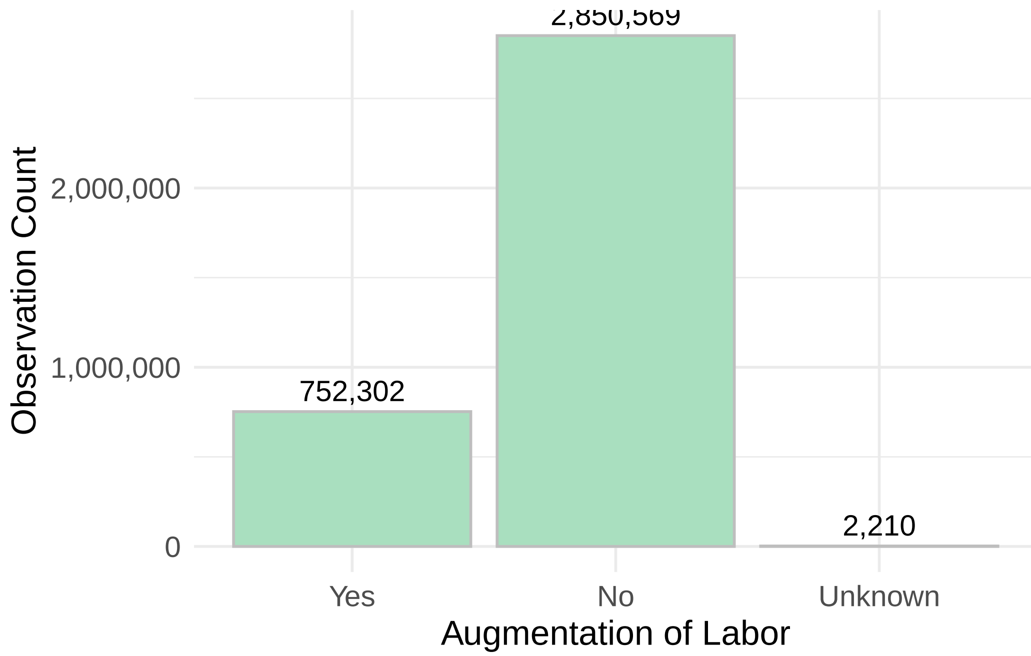


Figure 8: The distribution observations in the raw dataset based on whether the infant received augmentation of labor or not

2.5 Correlation between predictor variables

2.5.1 Induction of Labor and Augmentation of Labor

Figure 18 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).

2.5.2 Usage of Steroids and Antibiotics

In Figure 19, 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.

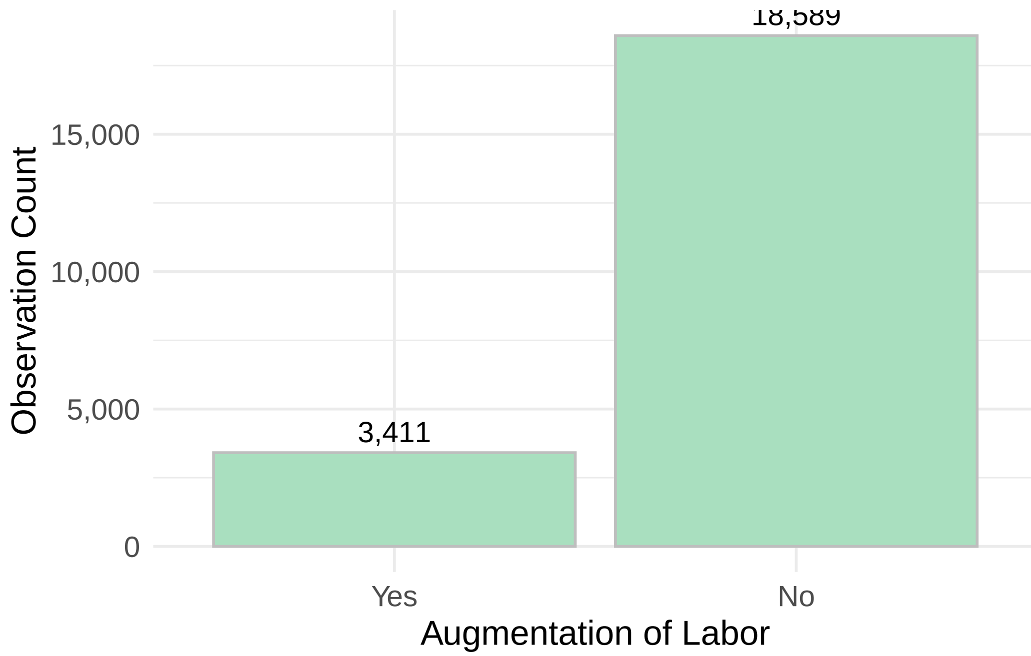


Figure 9: The distribution observations in the analysis datset based on whether the infant received augmentation of labor or not

2.5.3 Usage of Chorioamnionitis and Antibiotics

In Figure 20, 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.

3 Model

The goal of our modelling strategy is twofold. Firstly, a Random Forest model was utilized to identify significant predictors of infant health status, as measured by the 5-minute Apgar score. Secondly, a Bayesian Linear Model was applied to quantify the effects of medical interventions during labor and delivery on infant health outcomes, incorporating prior knowledge into the analysis framework.

Here we briefly describe the Random Forest and Bayesian analysis models used to investigate these relationships. Background details and diagnostics are included in Appendix C.

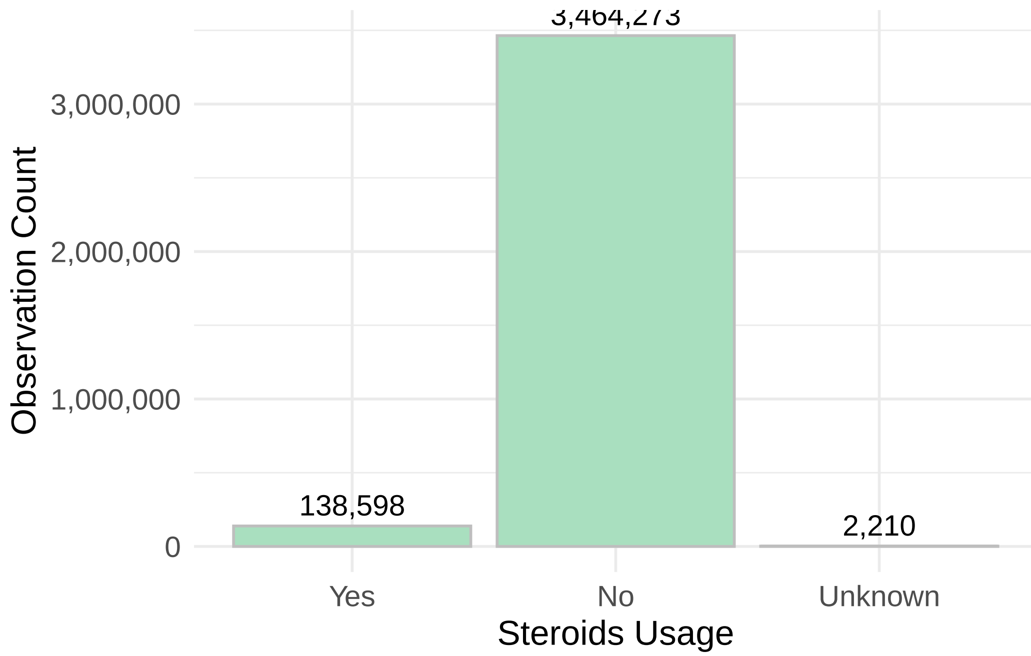


Figure 10: The distribution observations in the raw dataset based on whether the infant received steroid treatments or not

3.1 Model set-up

Define y_i as the APGAR5 score for the ith infant. Let the binary treatment 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 an ensemble of decision trees (Machine Learning Nuggets 2024):

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

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

We employ a grid search for hyperparameter tuning: - *mtry*: Number of variables randomly sampled at each split (values: 2, 3, 4). - *ntree*: Number of trees in the forest, fixed at 500.

The `caret` package(Kuhn and Max 2008) in R(R Core Team 2023) is used to implement the Random Forest model, with cross-validation to minimize overfitting.

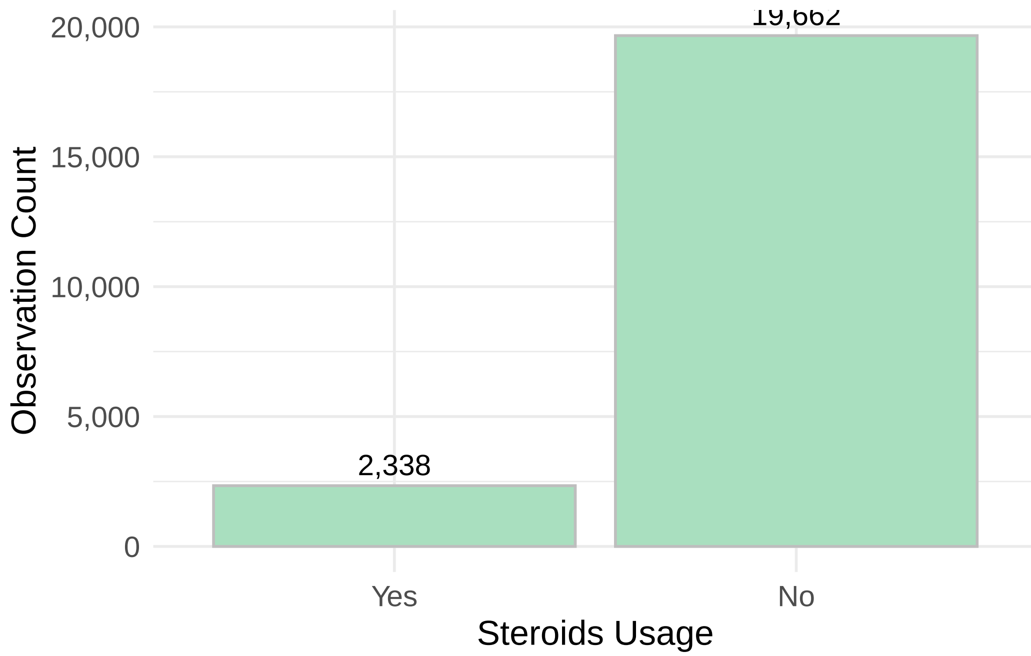


Figure 11: The distribution observations in the analysis datset based on whether the infant received steroid treatments or not

3.1.2 Bayesian Linear Model

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

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

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

Priors are specified as:

$$\alpha \sim \text{Normal}(5, 2), \quad \beta_j \sim \text{Normal}(0, 2) \quad \text{for } j = 1, \ldots, 6, \quad \sigma \sim \text{Exponential}(1)$$

We fit the Bayesian model using the `stan_glm` function in the `rstanarm`(Goodrich et al. 2024) package, which implements Markov Chian Monte Carlo (MCMC) sampling with 4 chaing and 2000 iterations per chain. We use the default priors from `rstanarm`(Goodrich et al. 2024).

3.1.3 Model justification

The Random Forest Model was chosen for its robustness in handling non-linear relationships and complex interactions among predictors (GeeksforGeeks n.d.), which is essential when analyzing the combined effects of various medical interventions on infant health outcomes.

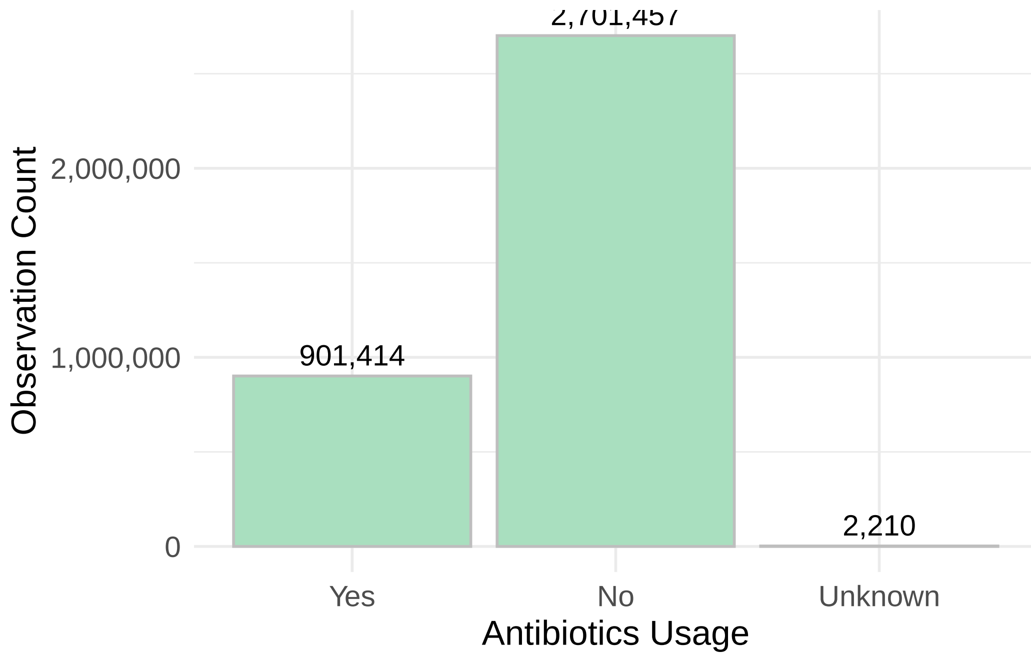


Figure 12: The distribution observations in the raw dataset based on whether the infant received antibiotic treatments or not

This method also provides interpretable measures of variable importance (GeeksforGeeks n.d.), aiding in understanding the relative impact of each treatment.

The Bayesian Linear Model was employed to incorporate prior knowledge (Strimmer Lab n.d.) 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 n.d.). The prior for the intercept (Normal(5, 2)) aligns with the central tendency of Apgar5 scores reported in previous studies (Cleveland Clinic n.d.). An exponential prior for the standard deviation parameter ensures positivity while discouraging extreme variability (Strimmer Lab n.d.).

By integrating the Random Forest's predictive accuracy (GeeksforGeeks n.d.) 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.

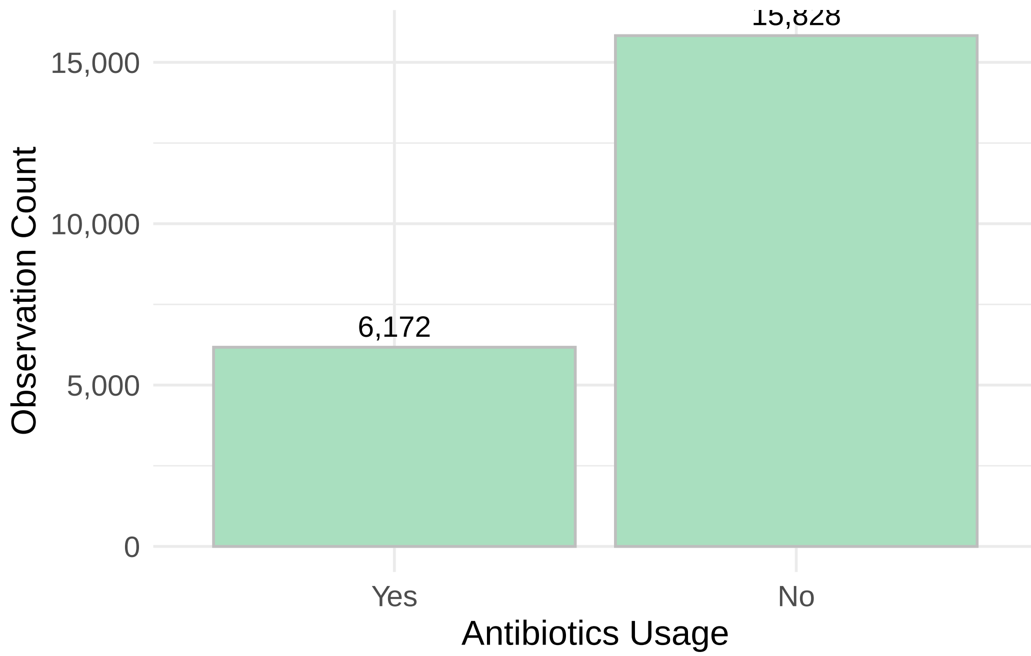


Figure 13: The distribution observations in the analysis datset based on whether the infant received antibiotic treatments or not

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 21 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.

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

Figure 22 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.

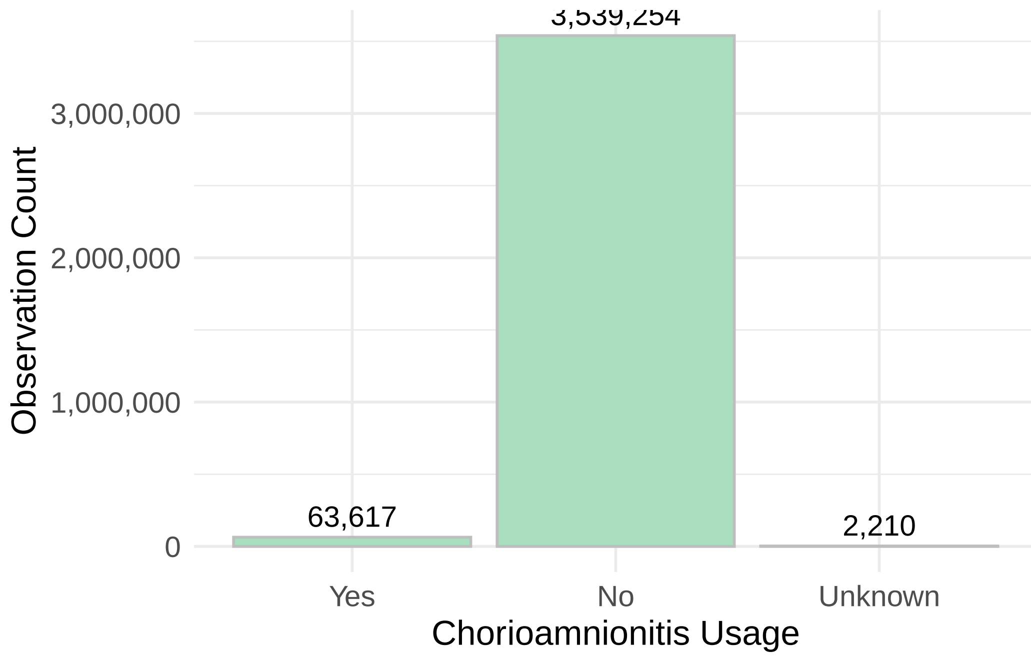


Figure 14: The distribution observations in the raw dataset based on whether the infant received chorioamnionitis treatments or not

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 23 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 24 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 25 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

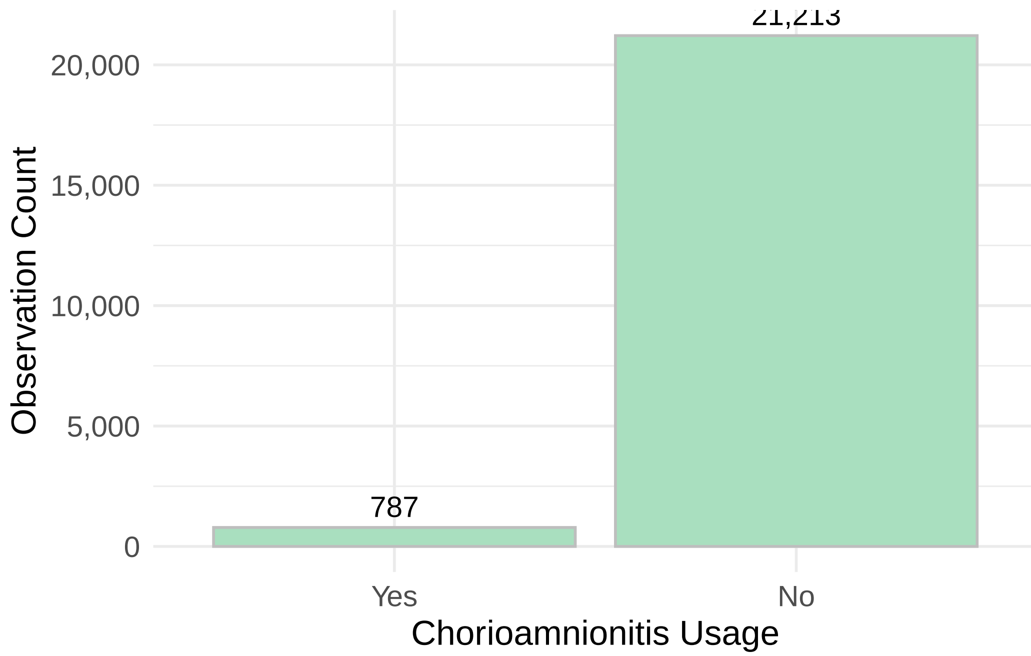


Figure 15: The distribution observations in the analysis datset based on whether the infant received chorioamnionitis treatments or not

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 26 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 27 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 28 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.

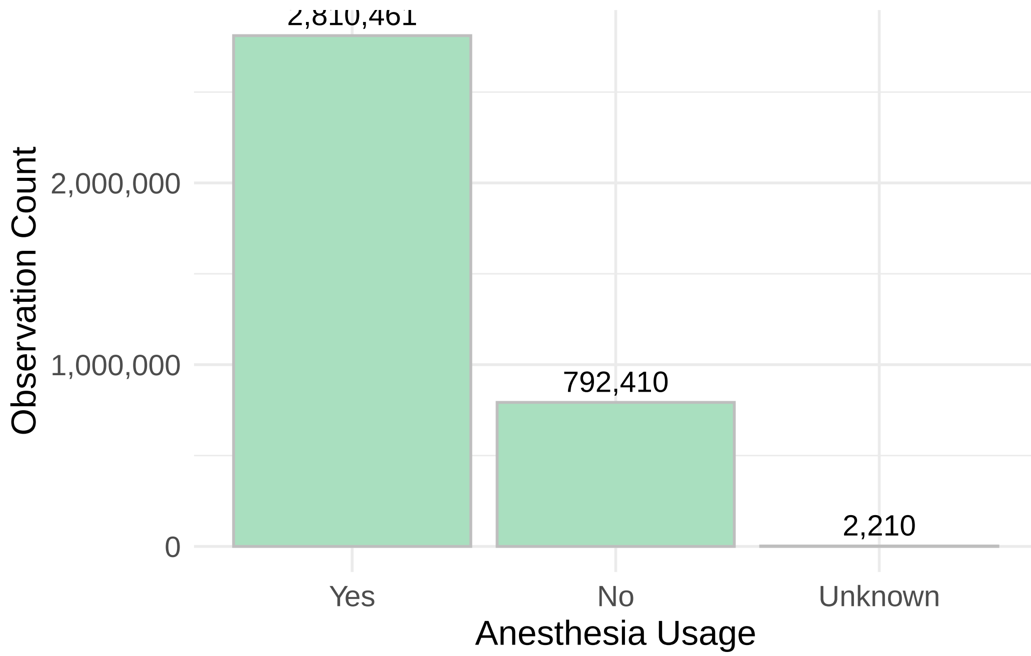


Figure 16: The distribution observations in the raw dataset based on whether the infant received anesthesia treatments or not

4.2 Results from the prediction model

4.2.1 Random Forest Model Results

Table 2 and Figure 29 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

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

5 Discussion

5.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

5.2 Second discussion point

Please don't use these as sub-heading labels - change them to be what your point actually is.

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Table 4: Explanatory models of flight time based on wing width and wing length

	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

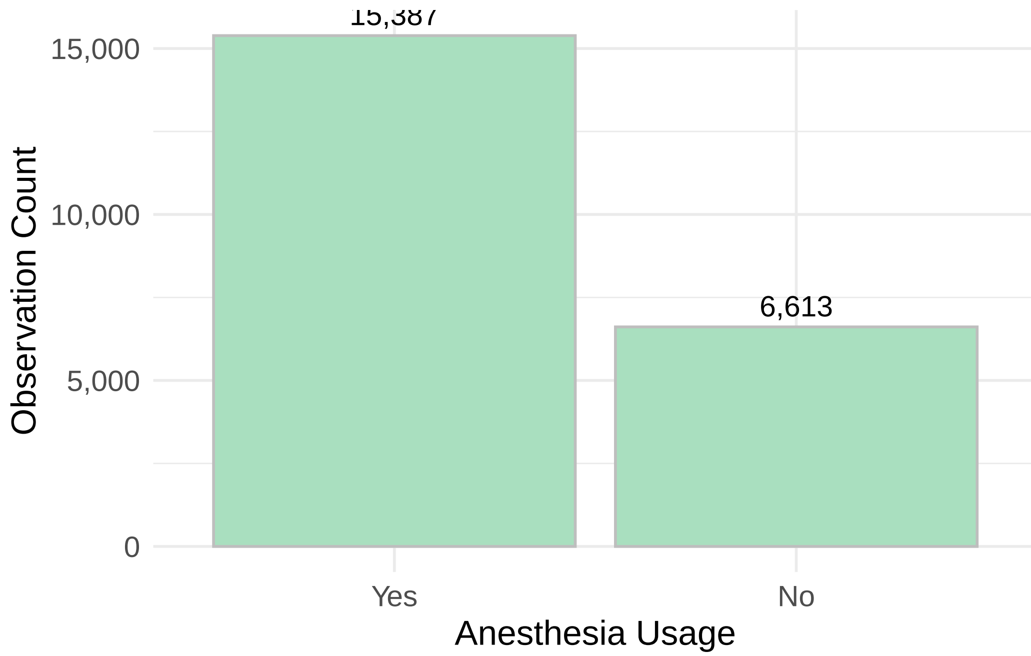


Figure 17: The distribution observations in the analysis datset based on whether the infant received anesthesia treatments or not

A Appendix

B Additional data details

C Model details

C.1 Posterior predictive check

In Figure 30a we implement a posterior predictive check. This shows...

In Figure 30b we compare the posterior with the prior. This shows...

C.2 Diagnostics

Figure 31a is a trace plot. It shows... This suggests...

Figure 31b is a Rhat plot. It shows... This suggests...

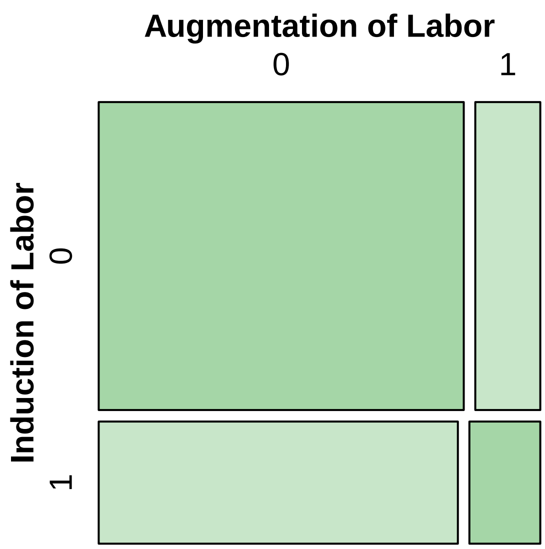


Figure 18: Mosaic plot showing the correlation between induction and augmentation of labor (**Note:** the shading intensity (from light green to dark green) represents the standardized residuals of the chi-squared test for the contingency table.)

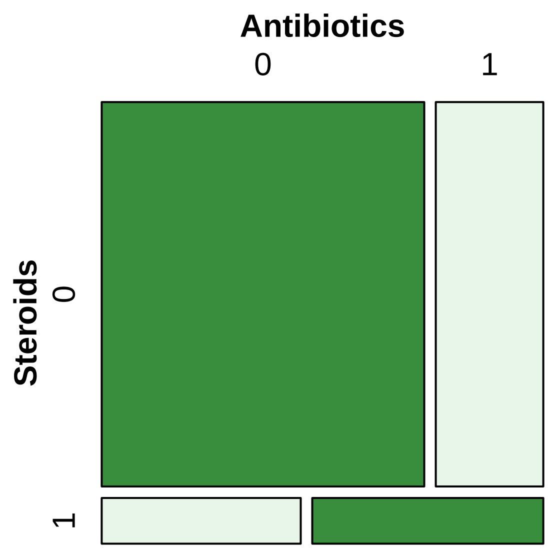


Figure 19: Mosaic plot showing the correlation between usage of steroids and antibiotics (**Note:** the shading intensity (from light green to dark green) represents the standardized residuals of the chi-squared test for the contingency table.)

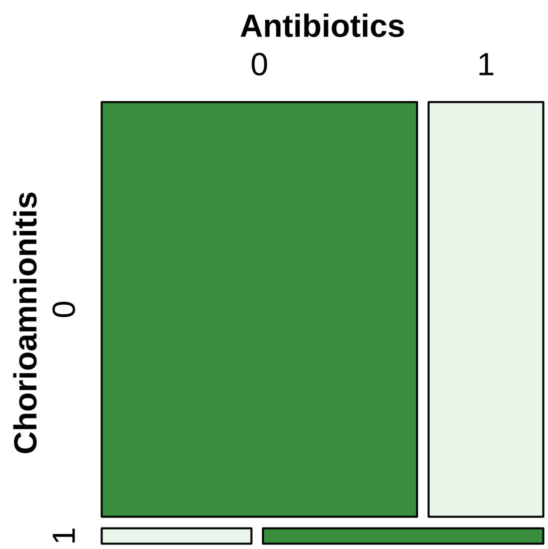


Figure 20: Mosaic plot showing the correlation between usage of chorioamnionitis and antibiotics (**Note:** the shading intensity (from light green to dark green) represents the standardized residuals of the chi-squared test for the contingency table.)

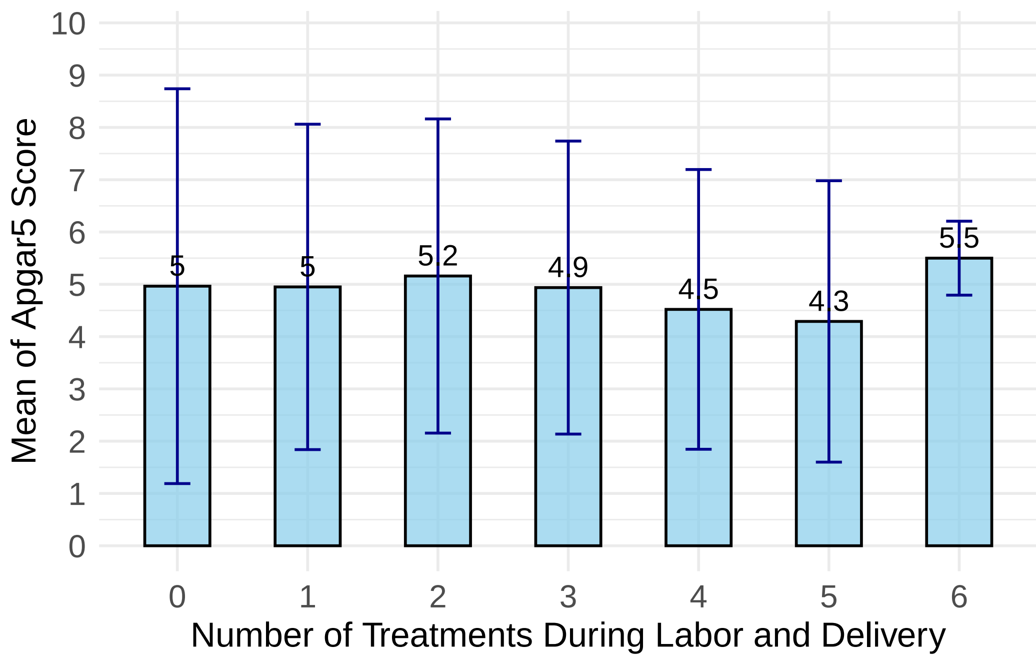


Figure 21: The mean of Apgar5 scores by the number of treatments received during labor and delivery (**Note:** The dark blue lines represent error bars, indicating the variability or uncertainty around the mean Apgar5 scores for each number of treatments.)

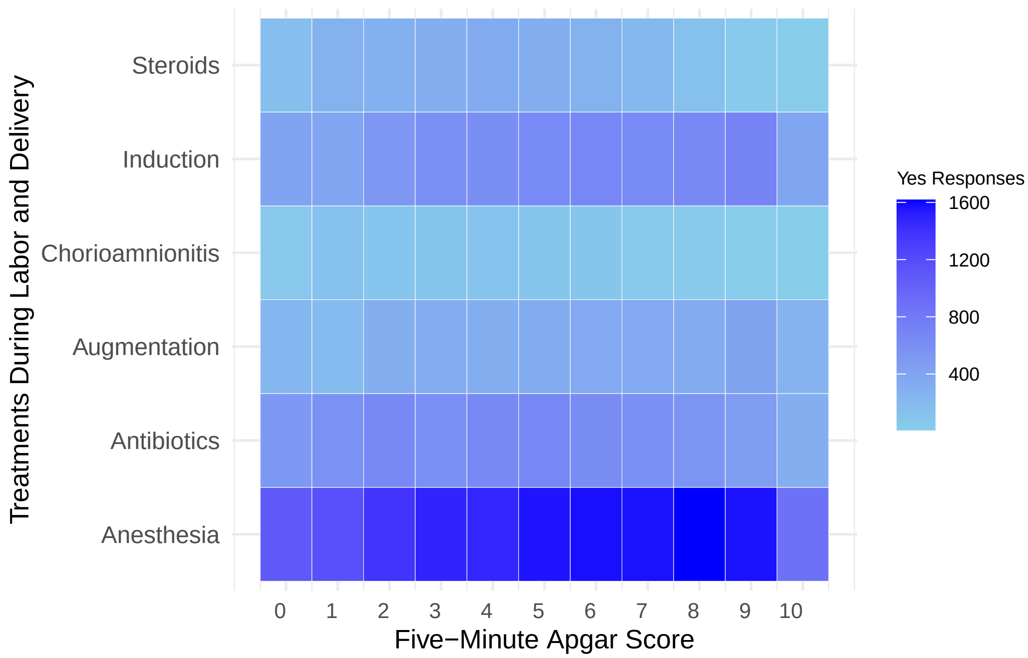


Figure 22: Heatmap of the overall correlation between treatments administered during labor/delivery and Apgar5 scores

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/.

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

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.

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). n.d. *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.

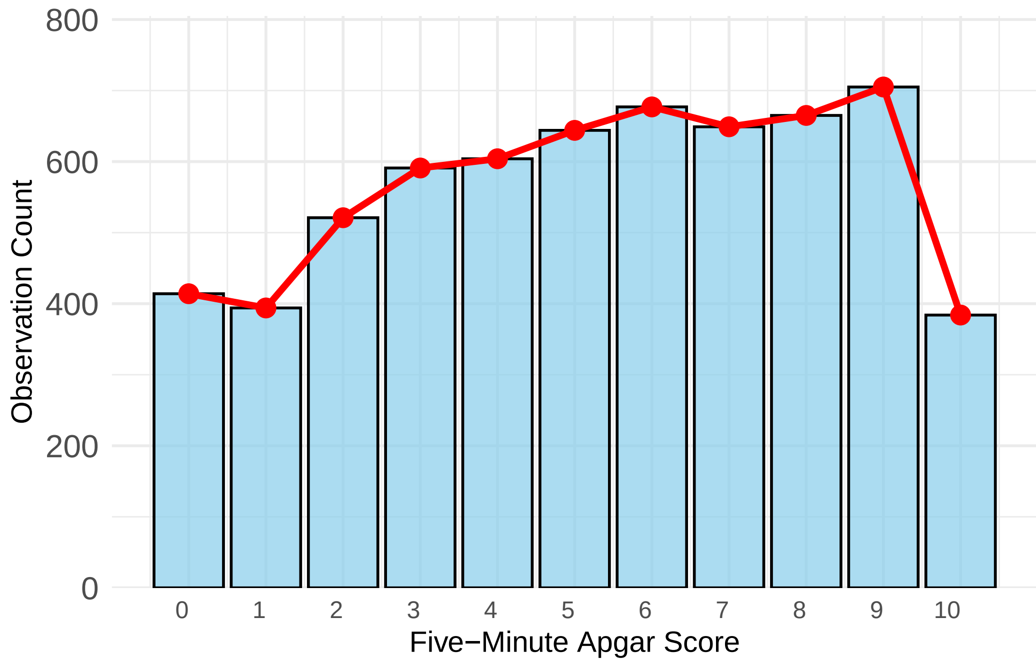


Figure 23: Apgar5 score distribution for the usage of induction labor (**Note:** The red line represents a visual summary of how observation counts change along the score range.)

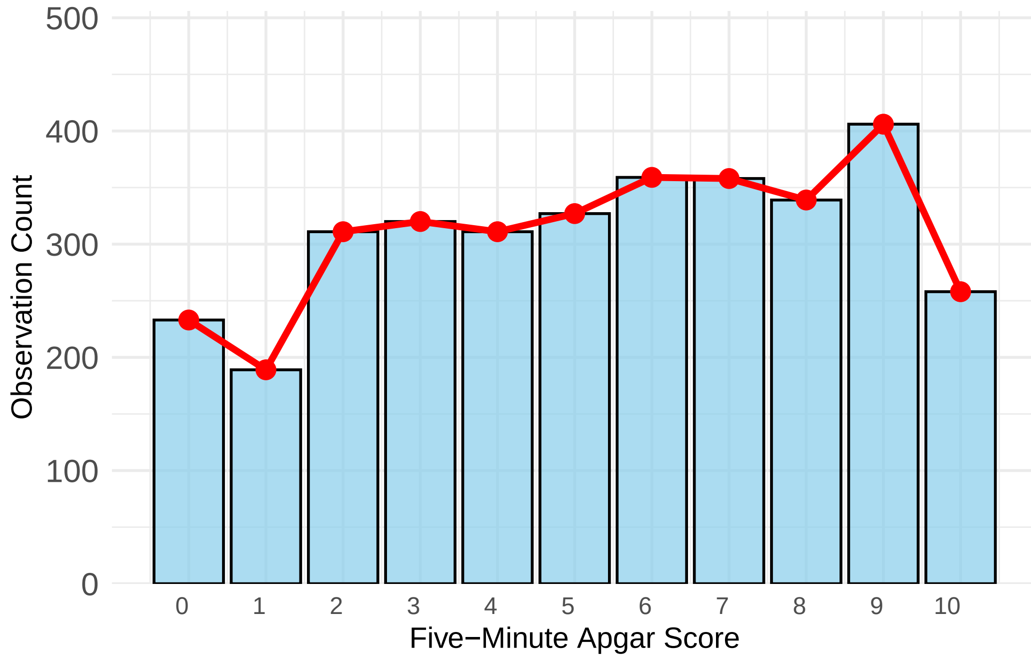


Figure 24: Apgar5 score distribution for the usage of augmentation labor (**Note:** The red line represents a visual summary of how observation counts change along the score range.)

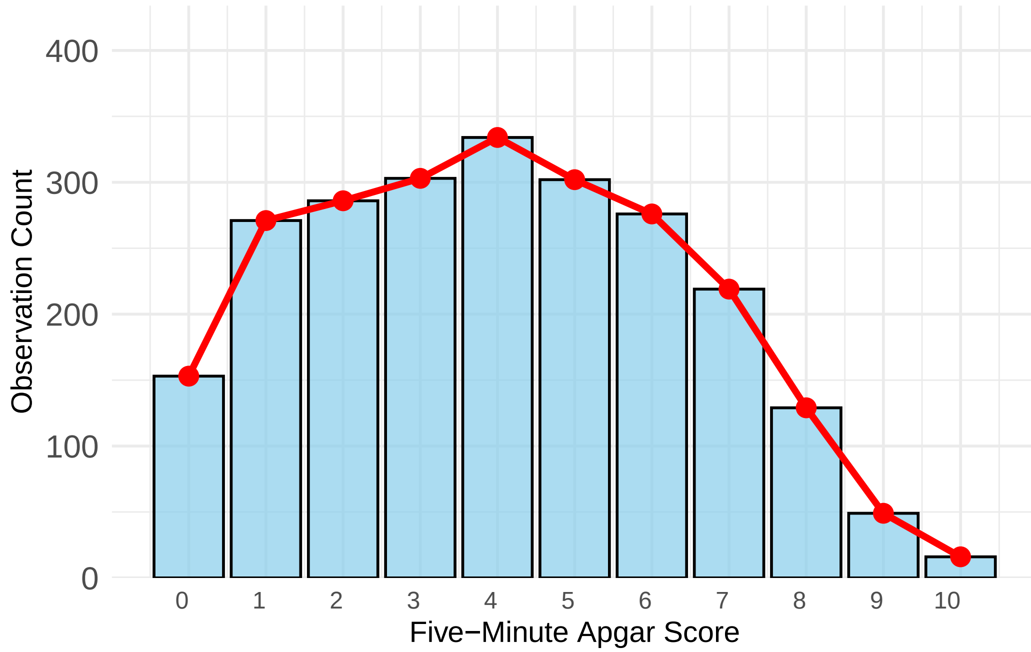


Figure 25: Apgar5 score distribution for the usage of steroids (**Note:** The red line represents a visual summary of how observation counts change along the score range.)

———. 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.

Cleveland Clinic. n.d. "Apgar Score." https://my.clevelandclinic.org/health/diagnostics/23094-apgar-score.

GeeksforGeeks. n.d. "Feature Encoding Techniques - Machine Learning." https://www.geeksforgeeks.org/feature-encoding-techniques-machine-learning/.

———. n.d. "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/.

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

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.

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

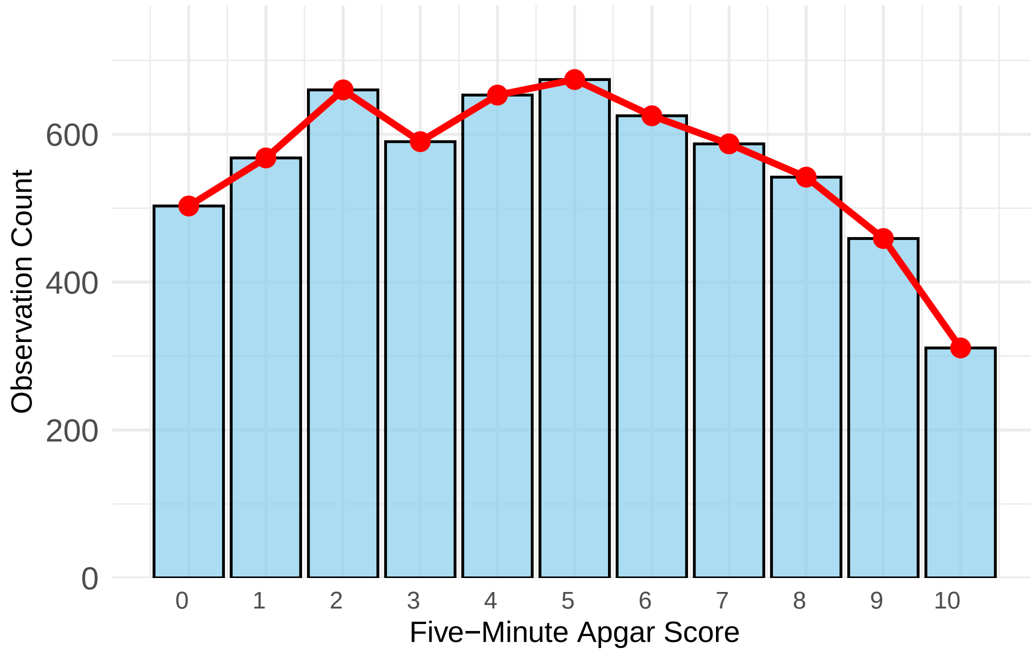


Figure 26: Apgar5 score distribution for the usage of antibiotics (**Note:** The red line represents a visual summary of how observation counts change along the score range.)

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

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

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/.

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.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing.* Vienna,

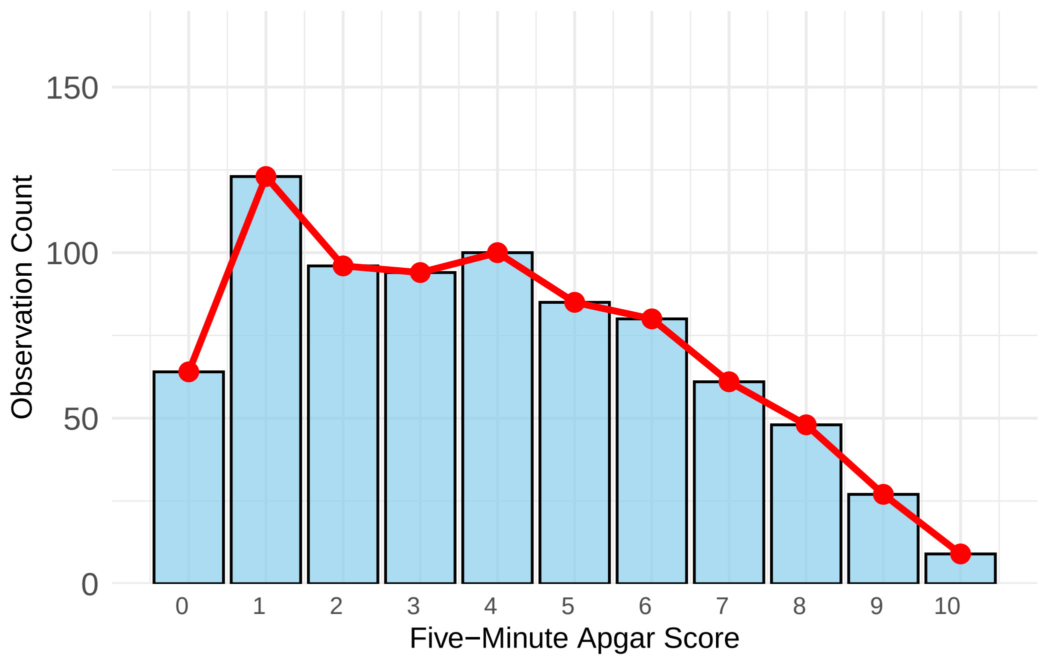


Figure 27: Apgar5 score distribution for the usage of chorioamnionitis (**Note:** The red line represents a visual summary of how observation counts change along the score range.)

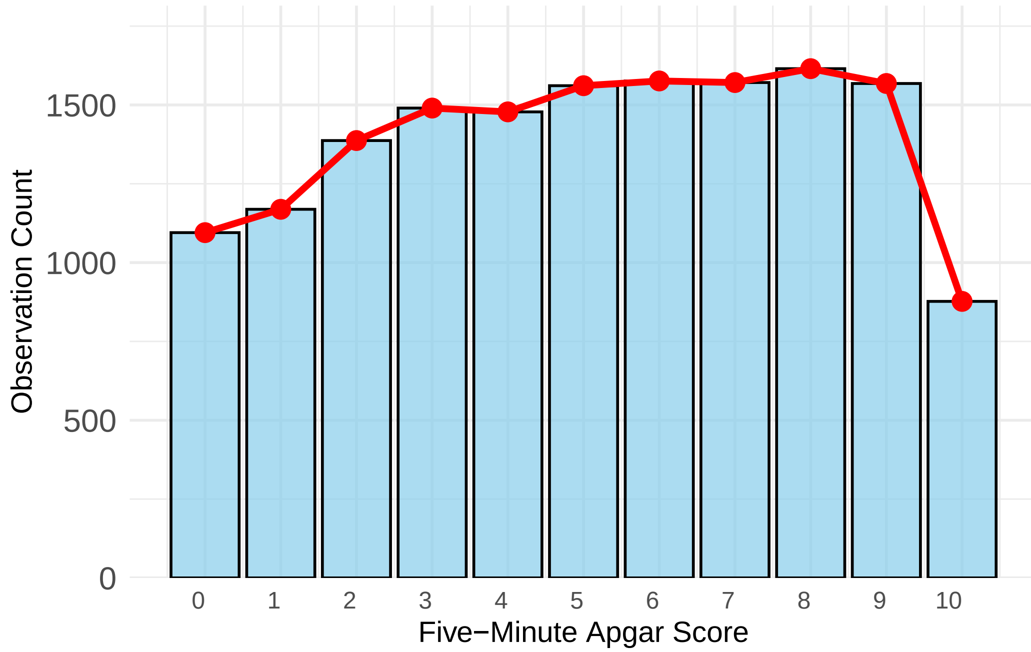


Figure 28: Apgar5 score distribution for the usage of anesthesia (**Note:** The red line represents a visual summary of how observation counts change along the score range.)

Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

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

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

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.

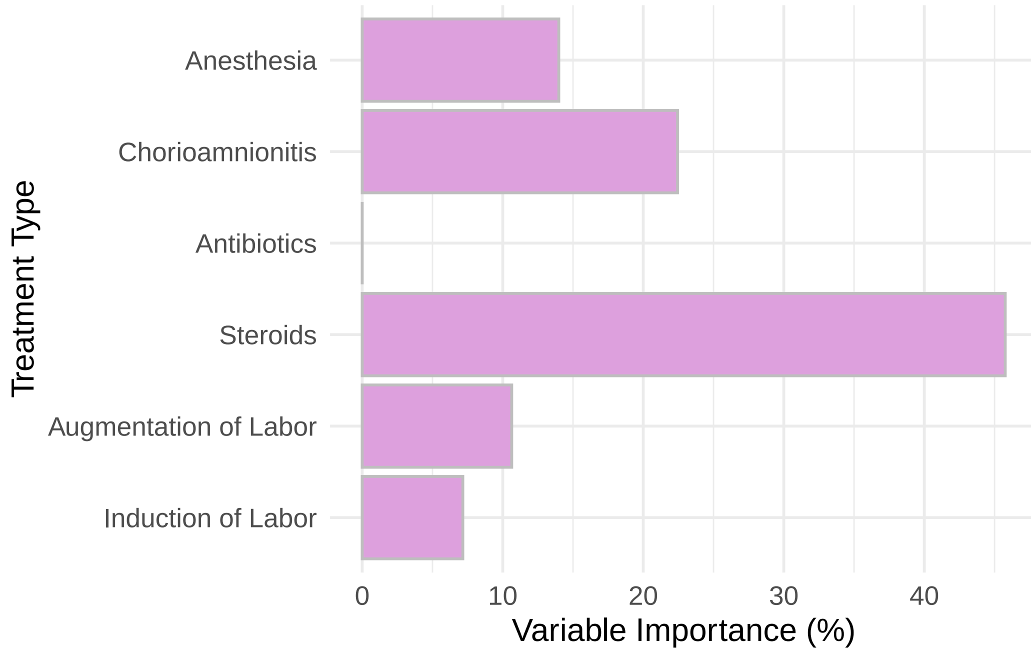
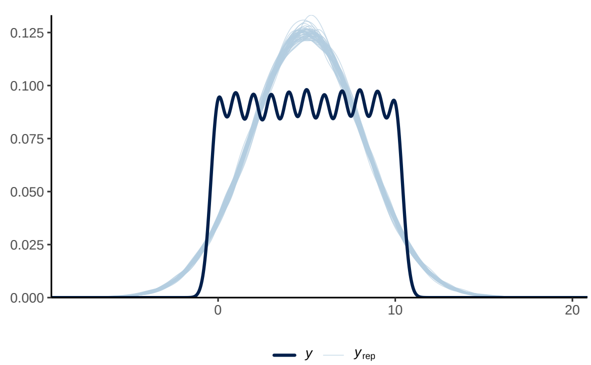
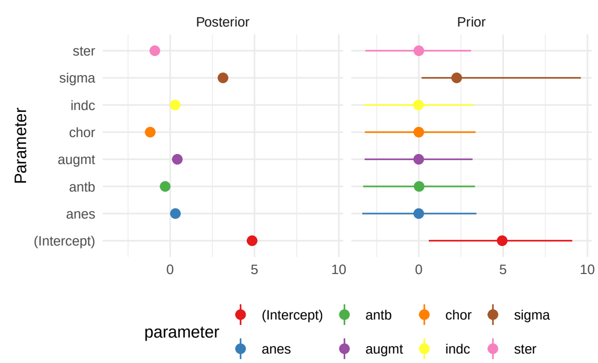


Figure 29: Variable importance of treatments during labor and delivery on infant health, expressed in bar plots

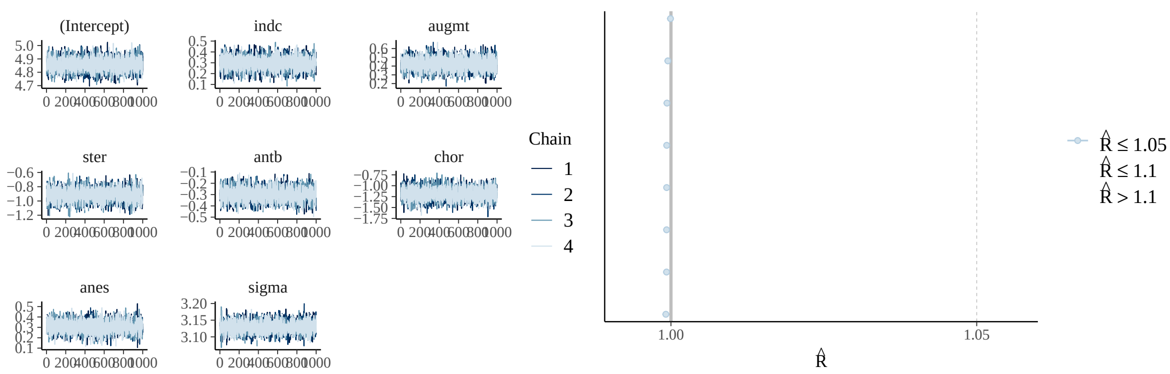


(a) Posterior prediction check



(b) Comparing the posterior with the prior

Figure 30: Examining how the model fits, and is affected by, the data



(a) Trace plot

(b) Rhat plot

Figure 31: Checking the convergence of the MCMC algorithm