R Notebook for the cystectomy study

Pascal Jerney

Last compiled on 26 Oktober, 2020

## Loading required package: pacman

## # Attaching packages  
## ✔ insight 0.10.0 ✔ bayestestR 0.7.5   
## ✔ performance 0.5.0 ✔ parameters 0.8.6   
## ✔ effectsize 0.4.0 ✔ modelbased 0.3.0   
## ✔ correlation 0.4.0 ✔ see 0.6.0

## The following packages have been unloaded:  
## MASS

## [conflicted] Will prefer DALEX::model\_performance over any other package

## [conflicted] Will prefer dplyr::filter over any other package

## [conflicted] Will prefer DALEX::explain over any other package

## [conflicted] Will prefer yardstick::spec over any other package

## [conflicted] Will prefer yardstick::sens over any other package

# Index

Welcome to the notebook!

## Table 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | no | yes | p | test |
| n | 798 | 370 |  |  |
| blood\_loss\_ml (median [IQR]) | 900.00 [700.00, 1245.00] | 1400.00 [1000.00, 2087.50] | <0.001 | nonnorm |
| blood\_loss\_ratio (median [IQR]) | 0.19 [0.14, 0.26] | 0.31 [0.22, 0.42] | <0.001 | nonnorm |
| oak = yes (%) | 23 (2.9) | 21 (5.7) | 0.030 |  |
| tcaggr = yes (%) | 87 (10.9) | 40 (10.8) | 1.000 |  |
| preop\_hb (median [IQR]) | 135.00 [123.00, 145.00] | 120.00 [104.00, 130.75] | <0.001 | nonnorm |
| preop\_tc (median [IQR]) | 251.50 [211.00, 309.00] | 267.00 [214.25, 338.75] | 0.002 | nonnorm |
| bmi (median [IQR]) | 25.56 [22.86, 28.72] | 25.40 [22.73, 28.51] | 0.710 | nonnorm |
| age (median [IQR]) | 67.08 [59.21, 74.29] | 71.05 [64.05, 77.41] | <0.001 | nonnorm |
| cci\_5plus (%) |  |  | <0.001 |  |
| 0 | 304 (38.1) | 106 (28.6) |  |  |
| 1 | 131 (16.4) | 48 (13.0) |  |  |
| 2 | 181 (22.7) | 90 (24.3) |  |  |
| 3 | 98 (12.3) | 56 (15.1) |  |  |
| 4 | 55 (6.9) | 37 (10.0) |  |  |
| 5 and more | 29 (3.6) | 33 (8.9) |  |  |
| gender = female (%) | 239 (29.9) | 140 (37.8) | 0.009 |  |
| p\_tumor (%) |  |  | <0.001 |  |
| 0 | 201 (25.2) | 59 (15.9) |  |  |
| 1 | 111 (13.9) | 38 (10.3) |  |  |
| 2 | 193 (24.2) | 75 (20.3) |  |  |
| 3 | 233 (29.2) | 129 (34.9) |  |  |
| 4 | 60 (7.5) | 69 (18.6) |  |  |
| p\_node\_pos = yes (%) | 167 (20.9) | 104 (28.1) | 0.009 |  |
| op\_year (%) |  |  | <0.001 |  |
| 2000 | 14 (1.8) | 35 (9.5) |  |  |
| 2001 | 25 (3.1) | 29 (7.8) |  |  |
| 2002 | 27 (3.4) | 19 (5.1) |  |  |
| 2003 | 54 (6.8) | 21 (5.7) |  |  |
| 2004 | 43 (5.4) | 26 (7.0) |  |  |
| 2005 | 31 (3.9) | 17 (4.6) |  |  |
| 2006 | 30 (3.8) | 26 (7.0) |  |  |
| 2007 | 46 (5.8) | 31 (8.4) |  |  |
| 2008 | 69 (8.6) | 19 (5.1) |  |  |
| 2009 | 47 (5.9) | 18 (4.9) |  |  |
| 2010 | 70 (8.8) | 22 (5.9) |  |  |
| 2011 | 69 (8.6) | 22 (5.9) |  |  |
| 2012 | 52 (6.5) | 29 (7.8) |  |  |
| 2013 | 74 (9.3) | 18 (4.9) |  |  |
| 2014 | 64 (8.0) | 15 (4.1) |  |  |
| 2015 | 53 (6.6) | 12 (3.2) |  |  |
| 2016 | 30 (3.8) | 11 (3.0) |  |  |
| op\_duration\_min (median [IQR]) | 390.00 [345.00, 425.75] | 408.50 [351.00, 454.00] | <0.001 | nonnorm |
| previous\_op = yes (%) | 326 (40.9) | 162 (43.8) | 0.378 |  |
| norepinephrine = yes (%) | 519 (65.0) | 203 (54.9) | 0.001 |  |
| crystalloids\_mlkgh (median [IQR]) | 4.50 [3.30, 6.01] | 5.30 [3.80, 7.06] | <0.001 | nonnorm |
| neoadj\_chemo = yes (%) | 101 (12.7) | 75 (20.3) | 0.001 |  |

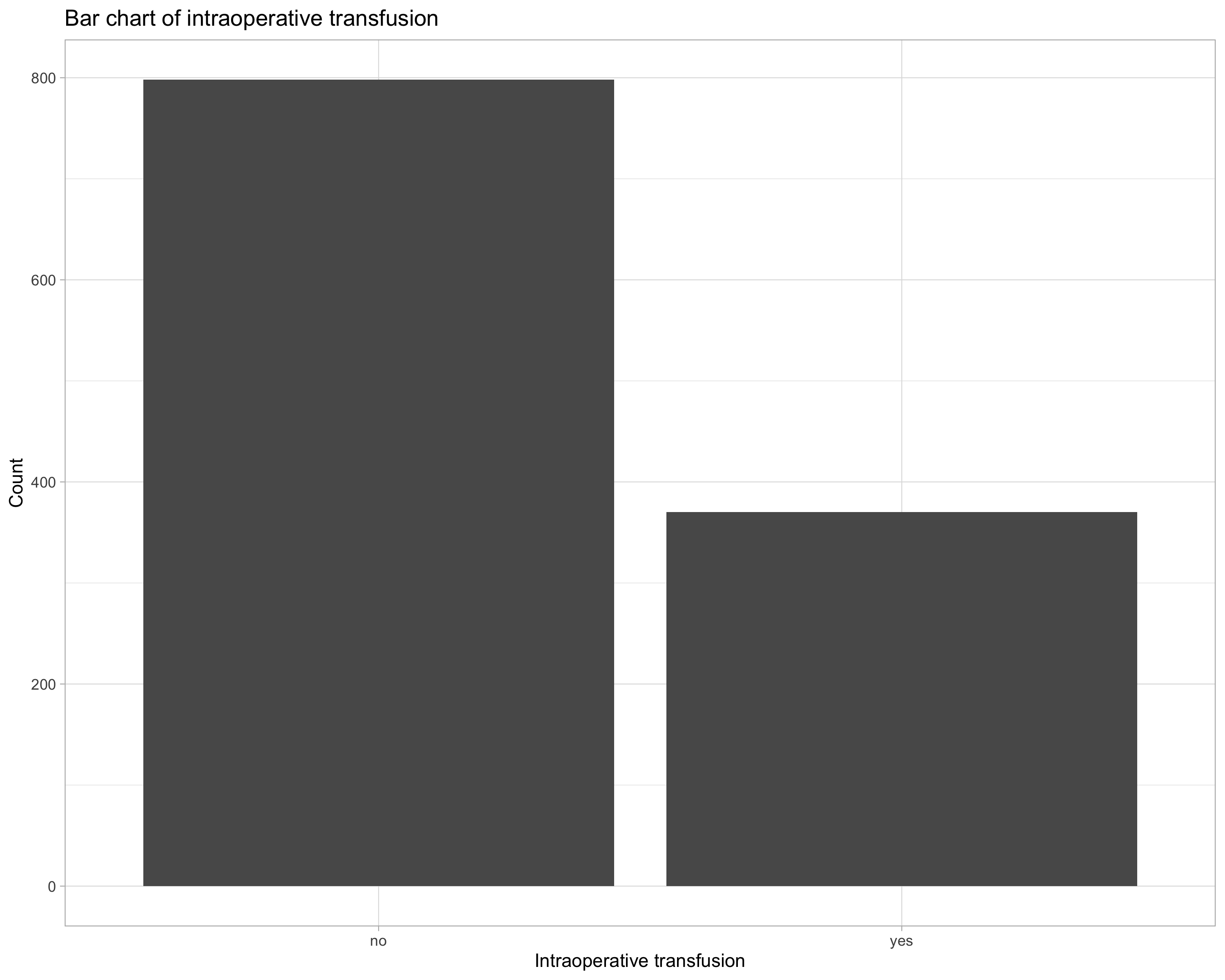
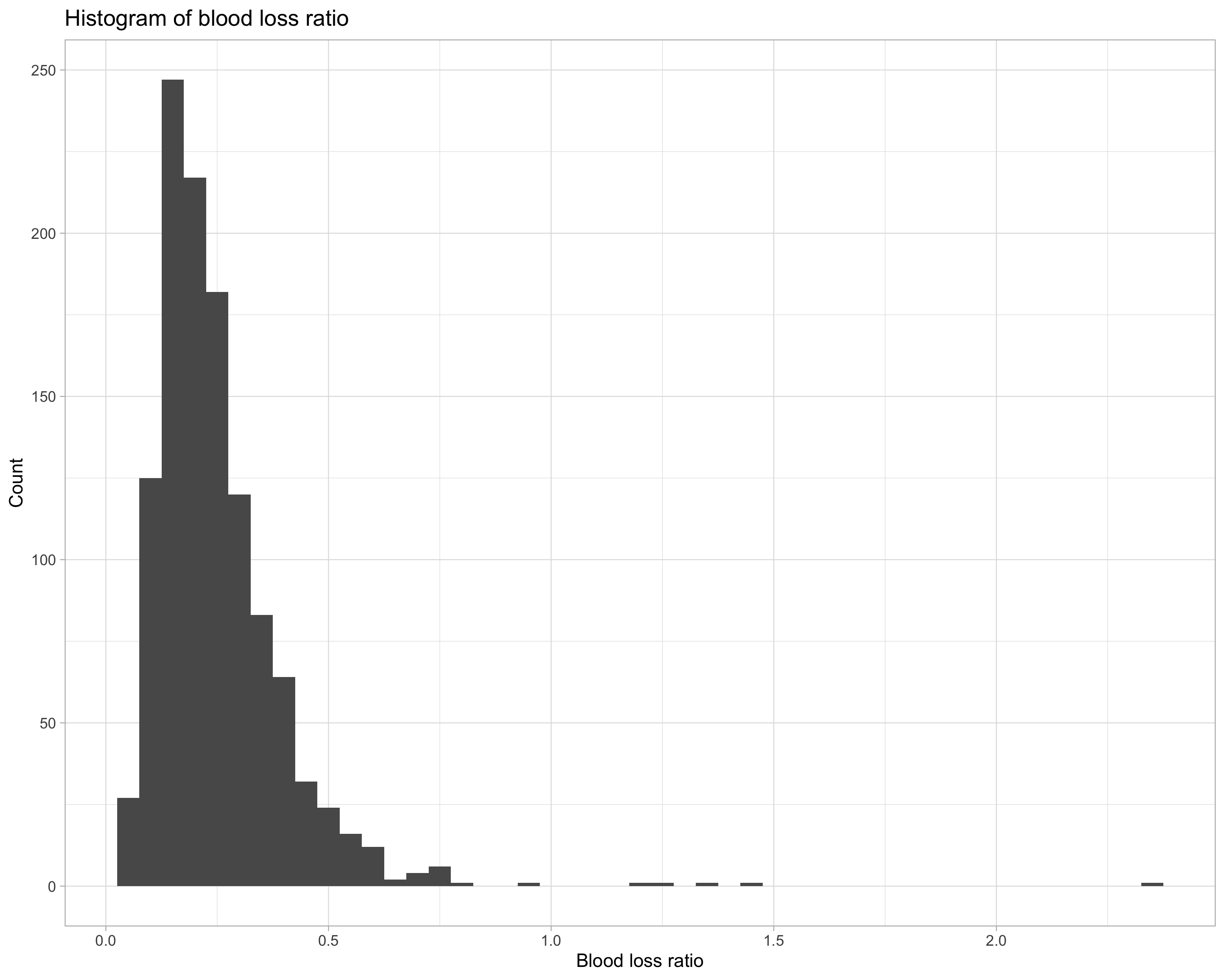
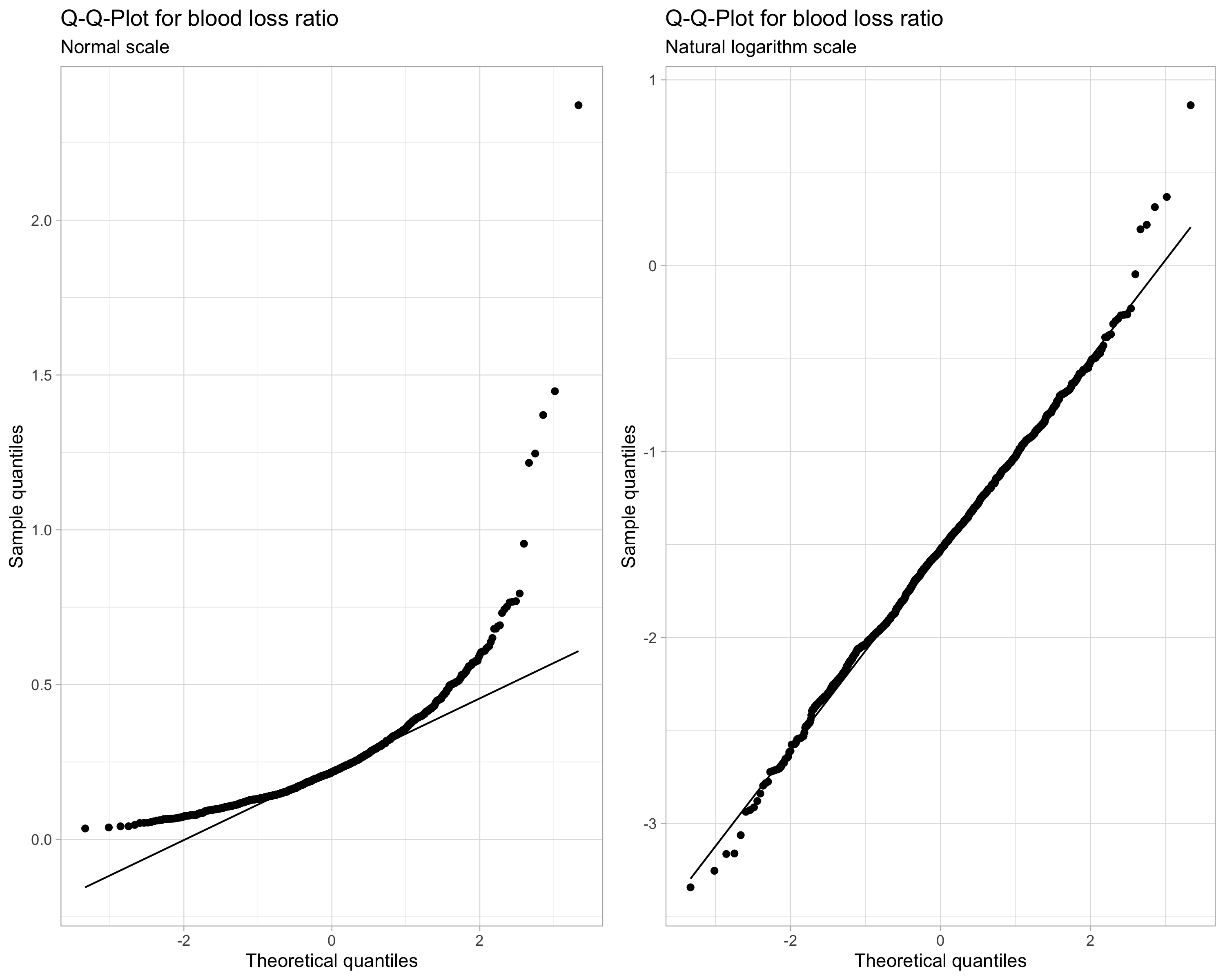
# Prerequisites

## Formulas

* Indexed blood volume[[1]](#footnote-24): [[2]](#footnote-26)
* Estimated blood volume: [[3]](#footnote-27)
* Blood loss ratio: [[4]](#footnote-28)
* Standardization method for age and bmi[[5]](#footnote-29): [[6]](#footnote-30)

# Results

## Data plots



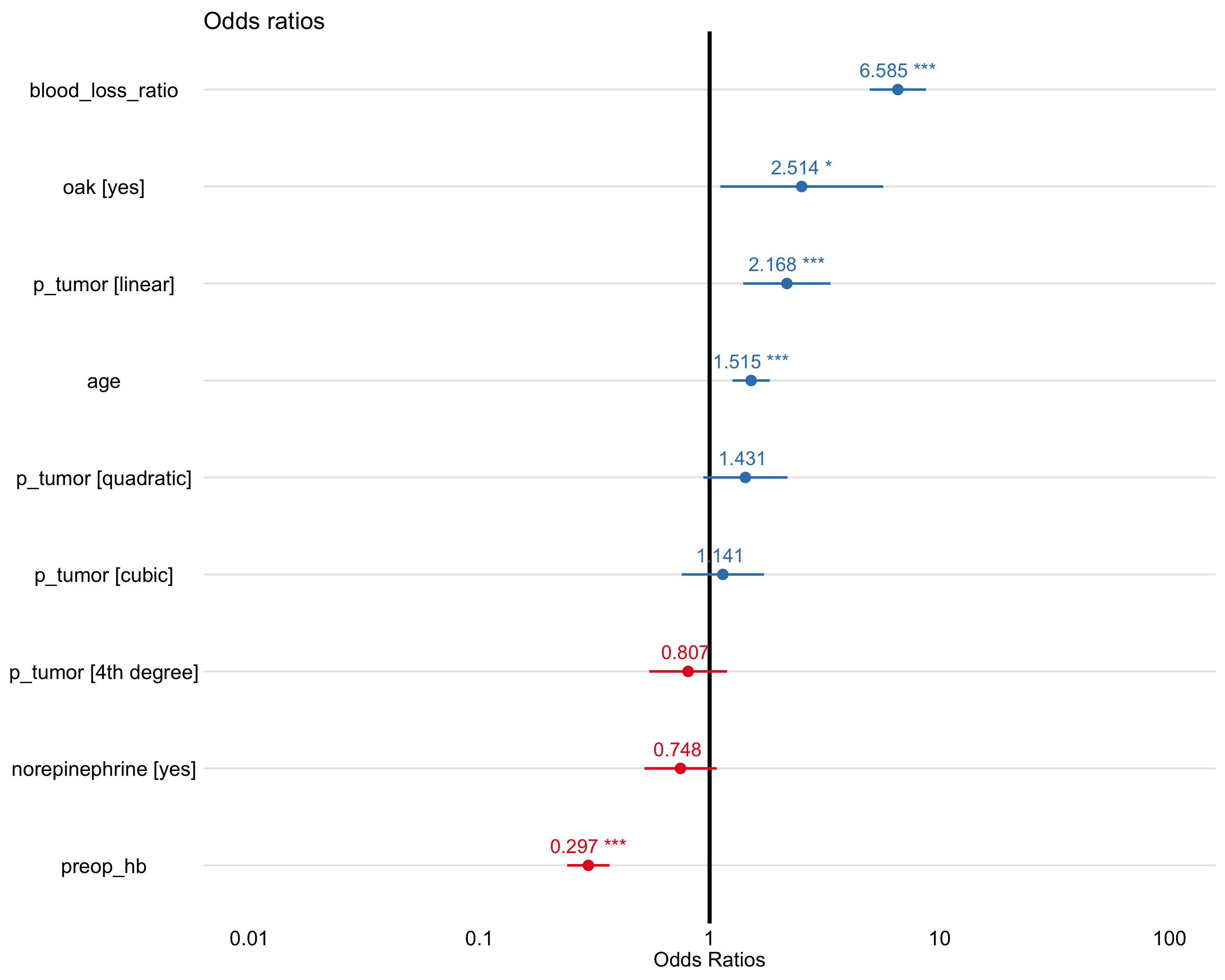
## Model outputs

### Models with intraoperative transfusion as response

#### Stepwise Logistic Regression

#> Warning: The `yardstick.event\_first` option has been deprecated as of yardstick 0.0.7 and will be completely ignored in a future version.  
#> Instead, set the following argument directly in the metric function:  
#> `options(yardstick.event\_first = TRUE)` -> `event\_level = 'first'` (the default)  
#> `options(yardstick.event\_first = FALSE)` -> `event\_level = 'second'`  
#> This warning is displayed once per session.  
#> Preparation of a new explainer is initiated  
#> -> model label : Stepwise Logistic Regression   
#> -> data : 115 rows 171 cols   
#> -> data : tibble converted into a data.frame   
#> -> target variable : 115 values   
#> -> predict function : yhat.workflow will be used ( [33m default [39m )  
#> -> predicted values : numerical, min = 0.00704458 , mean = 0.2955113 , max = 0.9999991   
#> -> model\_info : package tidymodels , ver. 0.1.1 , task classification ( [33m default [39m )   
#> -> model\_info : type set to classification   
#> -> residual function : difference between y and yhat ( [33m default [39m )  
#> -> residuals : numerical, min = -0.7274198 , mean = 0.0175322 , max = 0.9738698   
#> [32m A new explainer has been created! [39m

##### Coefficients



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| term | estimate | std.error | statistic | p.value | conf.low | conf.high |
| (Intercept) | 0.3969 | 0.1491 | -6.1960 | 0.0000 | 0.2949 | 0.5294 |
| oakyes | 2.5145 | 0.4159 | 2.2171 | 0.0266 | 1.0985 | 5.6550 |
| preop\_hb | 0.2971 | 0.1083 | -11.2100 | 0.0000 | 0.2388 | 0.3652 |
| age | 1.5154 | 0.0949 | 4.3787 | 0.0000 | 1.2617 | 1.8316 |
| norepinephrineyes | 0.7476 | 0.1845 | -1.5770 | 0.1148 | 0.5207 | 1.0740 |
| p\_tumor.L | 2.1679 | 0.2229 | 3.4712 | 0.0005 | 1.4032 | 3.3660 |
| p\_tumor.Q | 1.4311 | 0.2149 | 1.6684 | 0.0952 | 0.9394 | 2.1831 |
| p\_tumor.C | 1.1414 | 0.2105 | 0.6285 | 0.5297 | 0.7525 | 1.7194 |
| p\_tumor^4 | 0.8068 | 0.1984 | -1.0821 | 0.2792 | 0.5461 | 1.1902 |
| blood\_loss\_ratio | 6.5846 | 0.1437 | 13.1201 | 0.0000 | 5.0131 | 8.8088 |

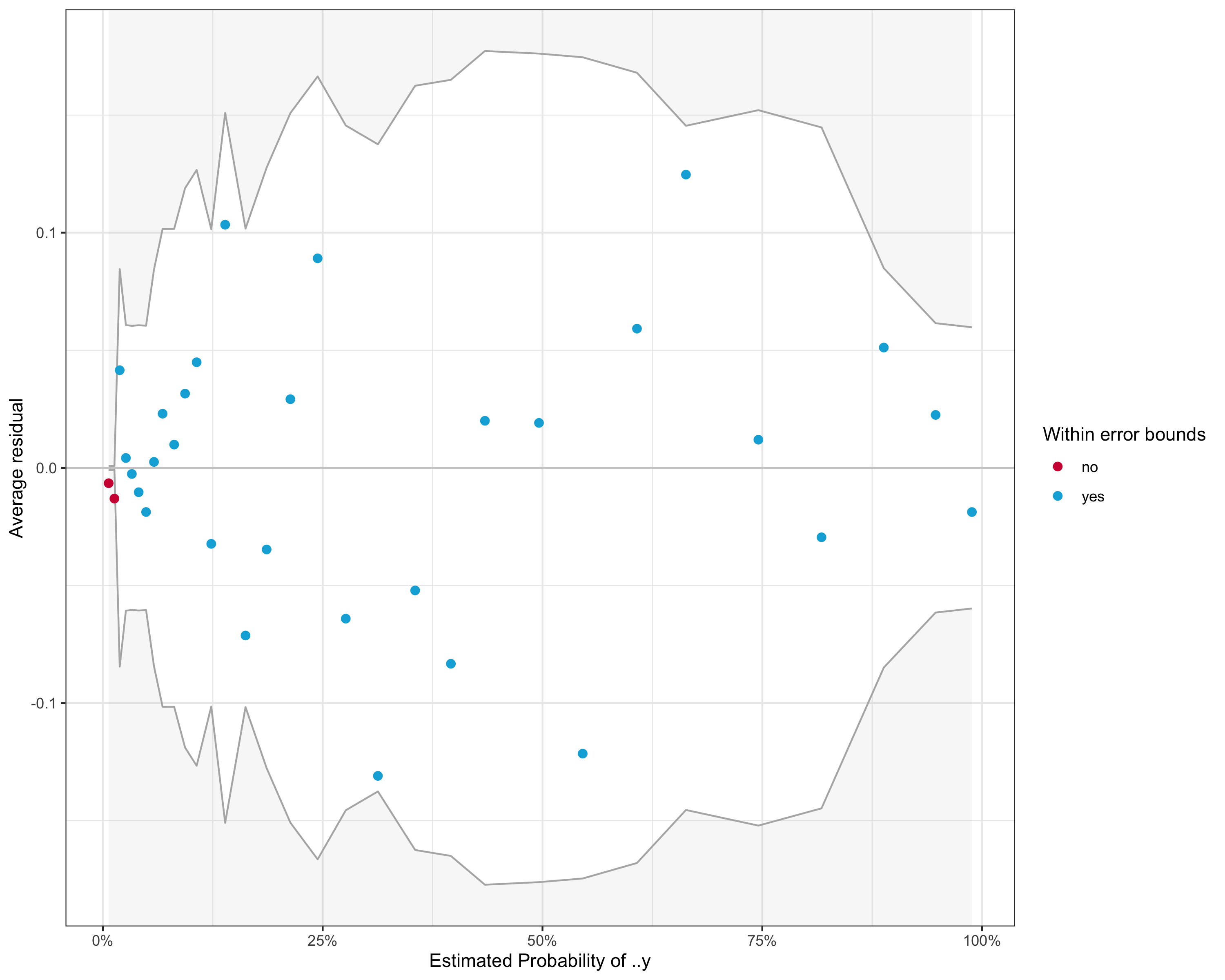
##### Estimated marginal means

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| term | contrast | null.value | odds.ratio | std.error | df | z.ratio | adj.p.value |
| p\_tumor | 1 / 0 | 0 | 1.2350 | 0.3960 | Inf | 0.6583 | 0.9186 |
| p\_tumor | 2 / 1 | 0 | 0.8258 | 0.2602 | Inf | -0.6075 | 0.9371 |
| p\_tumor | 3 / 2 | 0 | 1.6710 | 0.4088 | Inf | 2.0988 | 0.1239 |
| p\_tumor | 4 / 3 | 0 | 1.6977 | 0.4871 | Inf | 1.8446 | 0.2133 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| term | contrast | null.value | odds.ratio | std.error | df | z.ratio | adj.p.value |
| p\_tumor | 0|1 | 0 | 1.5786 | 0.3549 | Inf | 2.0306 | 0.1259 |
| p\_tumor | 1|2 | 0 | 1.5417 | 0.3067 | Inf | 2.1760 | 0.0907 |
| p\_tumor | 2|3 | 0 | 2.0562 | 0.4005 | Inf | 3.7008 | 0.0007 |
| p\_tumor | 3|4 | 0 | 2.3903 | 0.6410 | Inf | 3.2496 | 0.0040 |

##### Binned residuals

#> Warning: About 94% of the residuals are inside the error bounds (~95% or higher would be good).



##### Variance inflation factor

#> GVIF Df GVIF^(1/(2\*Df))  
#> oak 1.031610 1 1.015682  
#> preop\_hb 1.252277 1 1.119052  
#> age 1.070004 1 1.034410  
#> norepinephrine 1.071535 1 1.035150  
#> p\_tumor 1.096664 4 1.011601  
#> blood\_loss\_ratio 1.342430 1 1.158633

#### Elastic Net Logistic Regression

#> Preparation of a new explainer is initiated  
#> -> model label : Elastic Net Logistic Regression   
#> -> data : 115 rows 171 cols   
#> -> data : tibble converted into a data.frame   
#> -> target variable : 115 values   
#> -> predict function : yhat.workflow will be used ( [33m default [39m )  
#> -> predicted values : numerical, min = 0.01726955 , mean = 0.2951403 , max = 0.9999944   
#> -> model\_info : package tidymodels , ver. 0.1.1 , task classification ( [33m default [39m )   
#> -> residual function : difference between y and yhat ( [33m default [39m )  
#> -> residuals : numerical, min = -0.7034869 , mean = 0.01790322 , max = 0.9577324   
#> [32m A new explainer has been created! [39m

#### Random Forest

#> Preparation of a new explainer is initiated  
#> -> model label : Random Forest   
#> -> data : 115 rows 171 cols   
#> -> data : tibble converted into a data.frame   
#> -> target variable : 115 values   
#> -> predict function : yhat.workflow will be used ( [33m default [39m )  
#> -> predicted values : numerical, min = 0.00836172 , mean = 0.2907463 , max = 0.8754321   
#> -> model\_info : package tidymodels , ver. 0.1.1 , task classification ( [33m default [39m )   
#> -> residual function : difference between y and yhat ( [33m default [39m )  
#> -> residuals : numerical, min = -0.6476345 , mean = 0.02229717 , max = 0.9594582   
#> [32m A new explainer has been created! [39m

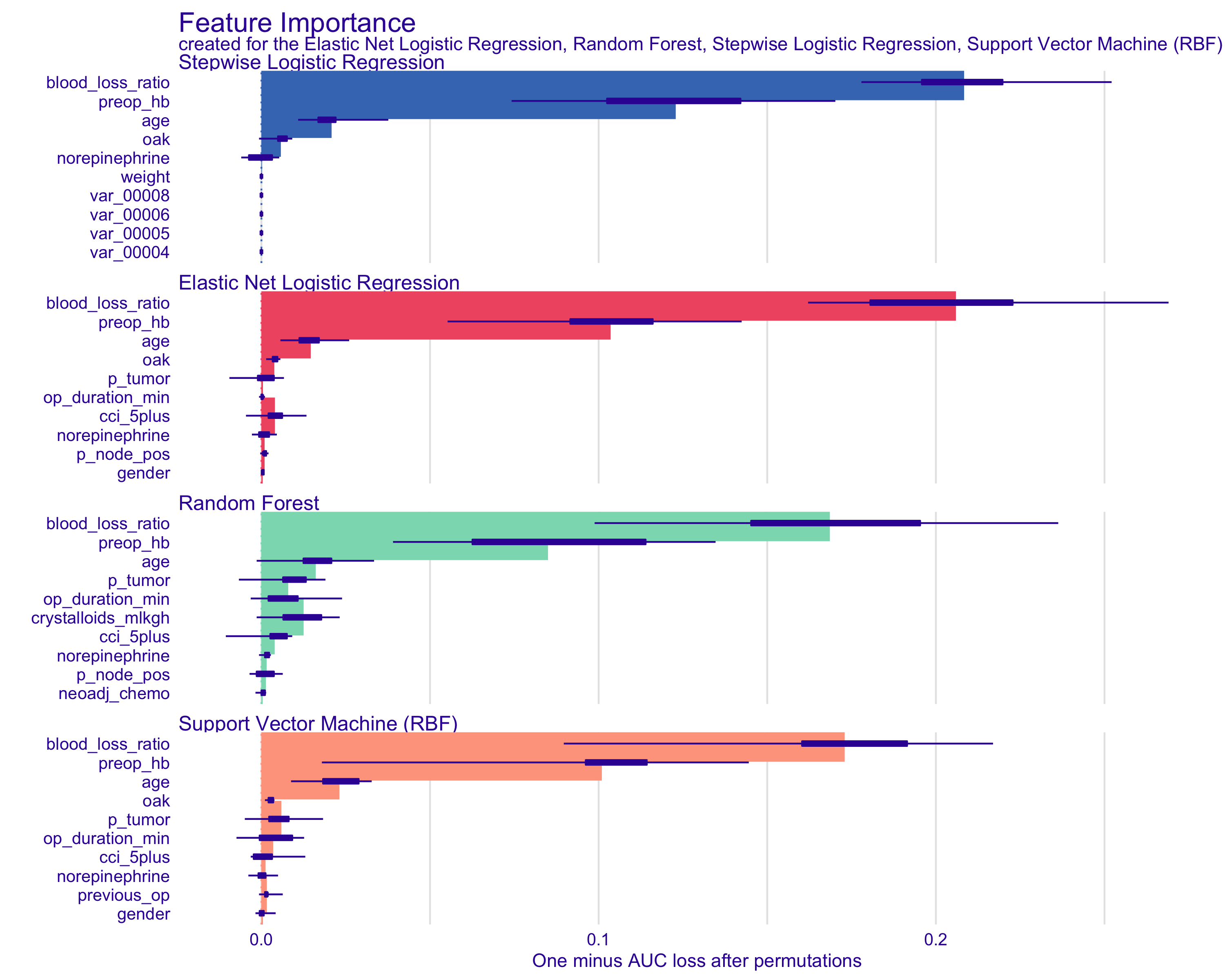
#### Support Vector Machine (RBF)

#> Preparation of a new explainer is initiated  
#> -> model label : Support Vector Machine (RBF)   
#> -> data : 115 rows 171 cols   
#> -> data : tibble converted into a data.frame   
#> -> target variable : 115 values   
#> -> predict function : yhat.workflow will be used ( [33m default [39m )  
#> -> predicted values : numerical, min = 0.0129083 , mean = 0.2976744 , max = 0.9999929   
#> -> model\_info : package tidymodels , ver. 0.1.1 , task classification ( [33m default [39m )   
#> -> residual function : difference between y and yhat ( [33m default [39m )  
#> -> residuals : numerical, min = -0.7628738 , mean = 0.01536906 , max = 0.9726944   
#> [32m A new explainer has been created! [39m

#### Model performance

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Accuracy | AUC | Brier |
| Stepwise Logistic Regression | 0.800 | 0.859 | 0.131 |
| Random Forest | 0.791 | 0.858 | 0.135 |
| Elastic Net Logistic Regression | 0.809 | 0.853 | 0.132 |
| Support Vector Machine | 0.791 | 0.852 | 0.136 |

#### Feature importance



#### Optimal cutpoint

(#tab:ml\_cutpoint\_table)Optimal probability threshold by absolute distance to optimal model (sensitivity and specificity = 1

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Threshold | Sensitivity | Specificity | Distance |
| Stepwise Logistic Regression | 0.3201714 | 0.7923708 | 0.7908440 | 0.0082724 |
| Elastic Net Logistic Regression | 0.3174359 | 0.7930836 | 0.7918560 | 0.0091146 |
| Support Vector Machine | 0.3149697 | 0.7945120 | 0.7951261 | 0.0092939 |
| Random Forest | 0.3494245 | 0.7849229 | 0.7846765 | 0.0131046 |

# Methods

Baseline, intra-operative and postoperative variables were compared between patients who received intraoperative blood product (packed red blood cells, fresh frozen plasma or platelet concentrate) transfusions and those who did not. Data were expressed as median with interquartile range for continuous variables and frequencies for categorical ones. We performed exploratory landmark analyses for categorical data using the test, and for continuous data the Kruskal-Wallis test.

For model validation we held out 20% of the data. We repeatedly split the remaining 80% into ten folds for cross-validation for ten repetitions overall. Multiple logistic regression analyses using a stepwise selection procedure (minimizing Akaike’s Information Criterion AIC) as well as using repeated 10-fold cross-validation for averaging ROC-AUC were applied to identify independent risk factors for intraoperative blood product transfusion and reported as Wald *P* value and adjusted odds ratio (OR) with 95% confidence intervals (CI). Factors that were selected a priori based on their potential association with intraoperative blood product transfusion were blood loss ratio, use of oral anticoagulants or platelet inhibitors, preoperative Hb values, BMI, age, sex, Charlson comorbidity index (CCI), tumor stage, positive nodal stage, duration of surgery, presence of previous surgery, use of noradrenaline, intra-operative volumes of crystalloids administered and presence of neoadjuvant chemotherapy.

1. Lemmens, H. J. M., Bernstein, D. P., & Brodsky, J. B. (2006). Estimating blood volume in obese and morbidly obese patients. Obesity Surgery, 16(6), 773–776. <https://doi.org/10.1381/096089206777346673> [↑](#footnote-ref-24)
2. = Indexed blood volume, = Body mass index [↑](#footnote-ref-26)
3. = Estimated blood volume [↑](#footnote-ref-27)
4. = Blood loss ratio, = Absolute blood loss [↑](#footnote-ref-28)
5. Iglewicz, B., & Hoaglin, D. C. (1993). How to detect and handle outliers. Milwaukee, Wis: ASQC Quality Press. [↑](#footnote-ref-29)
6. = Modified Z-score, = Median of , = Median absolute deviation [↑](#footnote-ref-30)