Predicting the incidence of postoperative nausea and vomiting

Ken Marciel, Project in Statistics (STAT 685), Texas A&M University

Spring 2022

1

Introduction

- PONV incidence is generally 20-40%
- Negative effects on health and well-being of patients
- Financially costly to healthcare providers
- Prophylaxis has negative side effects and financial costs

2

Challenge

- Develop a scoring system for accurate prescription
- Prophylaxis for patients at high risk of PONV
- Predictive model that is neither to conservative nor too liberal
- Balancing act known as "prevent-or-cure dilemma"

3

Scoring systems

- Several well-documented models for predicting PONV
- Guide prudent administration of anti-emetic prophylaxis
- Logistic regression and stepwise backward elimination typically used for variable selection

4

Performance evaluation

- Discrimination and calibration
  - Most common measures of validity
- AUC is most common measure of discriminating power
  - Area under receiver operating characteristic curve (ROC)
  - Values range from 0.61 to 0.785 across studies
- Slope and squared correlation (R^2) for line in calibration plot
  - Calibration slopes range from 0.3 to 1.71
  - Squared correlation ranges from 0.763 to 0.99

5

Purpose

- Analyze data set of 461 patients from anesthesiologist Jelena Velickovic, MD, in Belgrade, Serbia
- Develop predictive model for PONV
- Performance comparable to or better than previous models published in medical literature

6

Software

- R statistical computing
- R Markdown interface in R Studio
- Additional packages installed to R
  - dplyr - rename variables
  - alr4 - marginal model plots
  - leaps - regression subset plots
  - car - regression subset plots
  - rms - logistic regression
  - pROC - area under ROC curve
  - caret - data splitting, resampling

7

Data set

- Raw
  - No missing values
  - 916 rows and 26 columns
- Cleaned
  - Removed 93 duplicates
  - 823 rows and 26 columns
- Sample size
  - Removed 362 records of patients who received prophylaxis
  - 461 rows and 26 columns

8

Variables

- Original 26 variables encoded for analysis
  - ID variable, 9 response variables, 16 predictor variables
- Response variable selected
  - Y = PONV0to24 (binary) = incidence of PONV within 24 hours of operation
- Predictor variables selected
  - Excluded 8 variables
    - Anesthetic and postoperative risk factors
  - Remaining 8 variables for full model
    - Preoperative patient risk factors

9

Predictors

x_1 = Age (integer)

x_2 = Gender (binary)

$x_3 \ldots x_{27}$ = Diagnosis (categorical with 26 levels)

$x_{28} \ldots x_{34}$ = Surgery (categorical with 8 levels)

x_{35} = BMI (real)

x_{36} = Nonsmoker (binary)

x_{37} = Kinetosis history (binary)

x_{38} = PONV history (binary)

10

Full model

- 38 total variables
  - When factors with more than 2 levels are taken into full account
  - 25 dummy variables (for the 26 levels of *Diagnosis*)
  - 7 dummy variables (for 8 levels of *Surgery*)
- 10 variables for full model
  - ID variable
  - Response variable
  - 8 predictor variables
- Incidence of PONV for this data set is 37%

11

Exploratory data analysis

- Full logistic regression model
- Fitted using generalized linear method of least squares
- 3 predictors have estimated coefficients that are statistically significant
  - PONV history
  - Gender
  - Nonsmoker

12

Standardized deviance residuals

- Skewness present in response and all 8 predictors
  - Most are right-skewed
- Dependency of log odds on each skewed predictors
  - Linear function
  - Log transformation
- Residual plots are difficult to interpret for binary data
- Examined marginal model plots instead

13

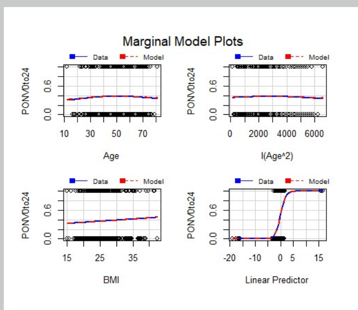
Marginal model plots

- Curves fitted for observed and predicted responses
- Reasonable agreement between both fits in each of the marginal model plots for *BMI* and *linear predictor*
- Lack of fit for *Age*, parabolic curvature for the observed response
- Resolved by adding a quadratic term for *Age*

14

Marginal model plots

- Reasonable agreement between observed and predicted fits
  - *Age*, *Age*^{2}, *BMI*, and *linear predictor*
- Indicates that current model is an adequate fit for the data



15

Leverage values and standardized deviance residuals

- None of leverage points exceed 2.5 standard deviations
- Six points exceed two standard deviations
  - Should be investigated
  - Only 1% of 461 values in data set
- Continued with assumption that current model is an adequate fit for the data
- Proceeded next to variable selection

16

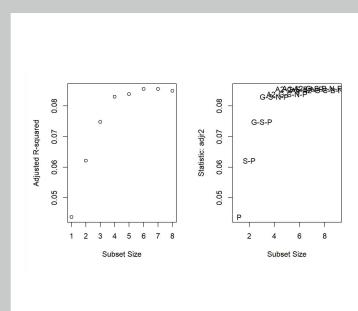
Modeling techniques for logistic regression

- Stepwise selection using all possible subsets of variables
- Repeated sampling using *k*-fold cross-validation
- Repeated sampling using bootstrap

17

Plots of adjusted R^2 against subset size

- Picks best subset of each size
- Squared and adjusted value of correlation coefficient
- Increases with number of predictors
- May lead to overfitting of model to data set
- Predictor with smallest *p*-value for its estimated coefficient was added to each subset to obtain the next subset



18

Stepwise selection using all possible subsets

Minimum BIC value corresponds to predictor subset of size two

- PONVhistory, Surgery

Minimum AIC and AIC_C values correspond to predictor subset of size four

- PONVhistory, Surgery, Gender, and Nonsmoker
- Includes same three predictors as full logistic regression
  - PONVhistory, Gender, and Nonsmoker
  - Before and after adding squared term for Age
- Has higher R^2 value than predictor subset of size two

Surgery excluded because none of its coefficients were statistically significant

19

Parsimonious logistic regression model from stepwise selection

$$Y = g(\beta_0 + \beta_1 PONVhistory + \beta_2 Gender + \beta_3 Nonsmoker + e)$$

where $e \sim \text{iid } N(0,1)$

Logit function used to model binary response variable:

$$g^{-1}(Y) = \left(\frac{\theta(Y)}{1-\theta(Y)}\right) = \beta_0 + \beta_1 PONVhist + \beta_2 Gender + \beta_3 Nonsmoker + e$$

where $\theta(Y) = \frac{(Y)}{1+(Y)} = \frac{1}{1+(-Y)}$

20

Estimated coefficients

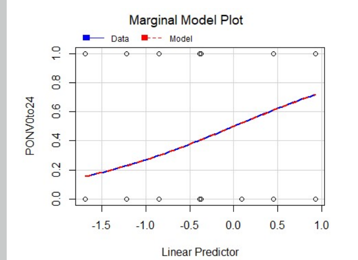
- Intercept and all three predictors are statistically significant

```
## Call:
## glm(formula = PONV0to24 ~ PONVhistory + Gender + Nonsmoker, family =
binomial,
##     data = ponv)
##
## Deviance Residuals:
##    Min      1Q  Median      3Q     Max
## -1.588  -1.023  -0.721   1.340   1.929
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -1.6911     0.3027  -5.587 2.32e-08 ***
## PONVhistory1   1.3029     0.3132   4.160 3.19e-05 ***
## Gender1        0.8401     0.2803   2.997  0.00273 **
## Nonsmoker1     0.4766     0.2159   2.208  0.02724 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 609.06  on 460  degrees of freedom
## Residual deviance: 573.95  on 457  degrees of freedom
## AIC: 581.95
##
## Number of Fisher Scoring iterations: 4
```

21

Linear predictor

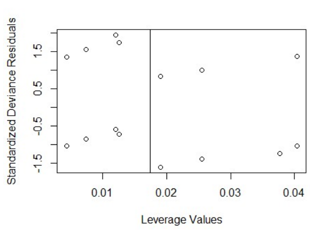
- Reasonable agreement between actual and predicted fits
- Model is an adequate fit for the data set



22

Leverage values and standardized deviance residuals

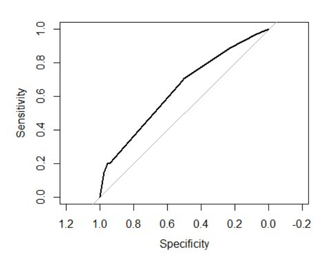
- All points within two standard deviations
- No bad leverage points
- Valid model for prediction



23

Area under ROC curve

- Receiver operating characteristic
- The area under ROC curve (AUC) is 0.639
- ROC curve is insensitive to class imbalance
- ROC curve is a function of both sensitivity and specificity



24

Confusion matrix

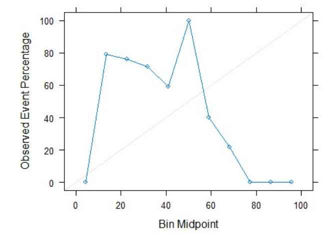
- Sensitivity of 0.94
  - Good at classifying high risk patients
- Specificity of 0.20
  - Poor at classifying low risk patients
- Reflected in the calibration plot

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction   0   1
##          0 272 138
##          1  17  34
##
##                Accuracy : 0.6638
##                  95% CI : (0.6186, 0.7068)
##     No Information Rate : 0.6269
##     P-Value [Acc > NIR] : 0.05527
##
##                   Kappa : 0.1619
##
##  Mcnemar's Test P-Value : < 2e-16
##
##             Sensitivity : 0.9412
##             Specificity : 0.1977
##          Pos Pred Value : 0.6634
##          Neg Pred Value : 0.6667
##              Prevalence : 0.6269
##          Detection Rate : 0.5900
##    Detection Prevalence : 0.8894
##       Balanced Accuracy : 0.5694
##
##        'Positive' Class : 0
##
## Setting levels: control = 0, case = 1
## Setting direction: controls < cases
```

25

Calibration plot

- Slope
  - Underestimates patients with low PONV risk
  - Overestimates patients with high PONV risk



26

Repeated sampling using *k*-fold cross-validation

- 10-fold cross-validation
  - Each training set contains 90% of data set
  - Each test set contains the other 10% of data (left out)
  - Each sample size between 414 and 416
    - Large training sets avoid potential bias and variance issues
- 5 repetitions of process
  - Generate 50 different holdout sets for estimating model accuracy
  - Increase precision of estimates while maintaining small bias

27

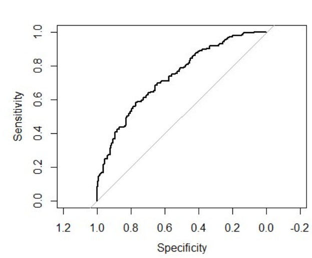
Repeated sampling using *k*-fold cross-validation

- Performed using the full logistic regression model
- 3 predictor coefficients were statistically significant ($p < .05$)
  - PONV history, Gender, and Nonsmoker
  - Same as model fitted using all possible subsets on full data set

28

Area under ROC curve

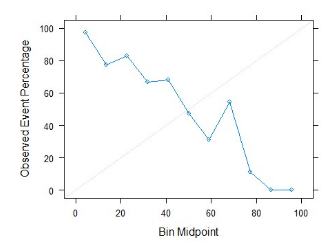
- 10-fold cross-validation
- AUC = 0.7358
- Improvement over AUC of 0.64 for baseline model obtained from all possible subsets



29

Calibration plot

- 10-fold cross-validation
- Sensitivity of 0.31
  - Poor at classifying high risk patients
- Specificity of 0.83
  - Good at classifying low risk patients
- Not reflected in calibration plot



30

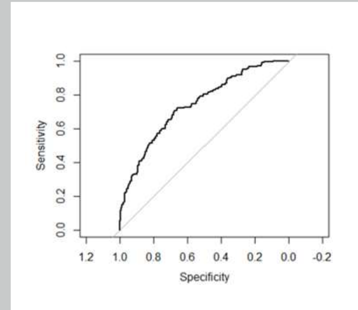
Repeated sampling using the bootstrap

- Random sample equal to size of data set
- Taken *with replacement*
- Repeated 25 times
- Performed using the full logistic regression model
- 3 predictor coefficients were statistically significant ($p < .05$)
  - PONV history, Gender, and Nonsmoker
  - Same as model fitted using all possible subsets on full data set

31

Area under ROC curve

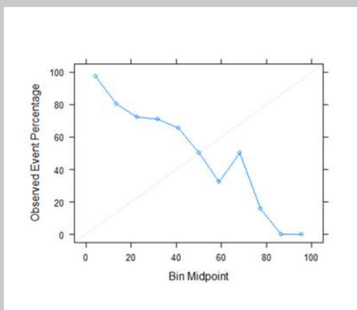
- Bootstrap
- AUC = 0.7408 for bootstrap
- Improvement over AUC of 0.64 for baseline model obtained from all possible subsets



32

Calibration plot

- Bootstrap
- Sensitivity of 0.34
  - Poor at classifying high risk patients
- Specificity of 0.78
  - Good at classifying low risk patients
- Not reflected in calibration plot



33

Predictions of PONV

- Logistic regression model obtained from all possible subsets consists of three predictors which are all binary variables
- Lacks numerous dummy variables of two models trained with resampling techniques
- Chosen as parsimonious model to make predictions
- Applied to examples of hypothetical patients

34

Predictions of PONV

PONV history	Gender	Nonsmoker	PONV probability
Yes	Female	Yes	0.7168
Yes	Female	No	0.6111
Yes	Male	Yes	0.5221
Yes	Male	No	0.4042
No	Female	Yes	0.4075
No	Female	No	0.2992
No	Male	Yes	0.2289
No	Male	No	0.1556

35

Predictions of PONV

- The more significant predictors that a patient has, the more likely the patient will have PONV
- A patient with all three risk factors in the parsimonious model has a 72% probability of experiencing PONV
- A patient with none of the three risk factors has a 16% probability of experiencing PONV
- Previously determined to be a valid predictive model

36

Conclusion

Simple model may have practical application for patient populations with characteristics like this data set

To choose a threshold for prescription of prophylaxis, healthcare professionals could select one of the two hypothetical predictions in our example having a probability greater than 60%

37

Acknowledgments

This semester-long investigation was conducted under the guidance of Professor Brani Vidakovic, PhD

In collaboration with graduating classmates (MS in Statistics) Scott Tillet and Nguyet Vu, in Spring 2022, for the Project in Statistics (STAT 685) graduate course at Texas A&M University

38

Literature references

Apfel, C. C., Kranke, P., Eberhart, L. H. J., Roos, A., and Roewer, N. (2002), "Comparison of Predictive Models for Postoperative Nausea and Vomiting," British Journal of Anaesthesia, 88 (2), 234-40.

Eberhart, L. H. J., Hogel, J., Seeling, W., Staack, A.M., Geldner, G., and Georgieff, M. (2000), "Evaluation of Three Risk Scores to Predict Postoperative Nausea and Vomiting," Acta Anaesthesiologica Scandinavica, 44, 480–488.

Sinclair, D. R., Chung, F., and Mezei, G. (1999), "Can Postoperative Nausea and Vomiting Be Predicted?" Anesthesiology, 91, 109-118.

Thomas, R., Jones, N. A., and Strike, P. (2002), "The Value of Risk Scores for Predicting Postoperative Nausea and Vomiting when Used to Compare Patient Groups in a Randomised Controlled Trial," Anaesthesia, 57, 1119-1128.

van den Bosch, J.E., Kalkman, C. J., Vergouwe, Y., Van Klei, W. A., Bonsel, G. J., Grobbee, D. E., and Moons, K. G. M. (2005), "Assessing the Applicability of Scoring Systems for Predicting Postoperative Nausea and Vomiting," Anaesthesia, 60, 323-331.

39

Textbook references

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2021), An Introduction to Statistical Learning (2nd ed.), New York, NY: Springer Science+Business Media, LLC.

Kuhn, M., and Johnson, K. (2013), Applied Predictive Modeling, New York , NY: Springer Science+Business Media, LLC.

Pampel, F. C. (2021), Logistic Regression (2nd ed.), Thousand Oaks, CA: SAGE Publications, Inc.

Sheather, S. J. (2009), A Modern Approach to Regression with R, New York, NY: Springer Science+Business Media, LLC.

Vidakovic, B. (2017), Engineering Biostatistics, Hoboken, NJ: John Wiley & Sons Ltd.

40