Predicting Postoperative In-Hospital Mortality for Non-Emergency Procedures Using the INSPIRE Dataset.

Douglas Snyder, MD

Introduction

Preoperative evaluation is a critical component for patients undergoing anesthesia for both operative and non-operative procedures. The guidelines provided by the American Society of Anesthesiologists (ASA) give considerations for how patients should be evaluated prior to undergoing an anesthetic[1]. A complete review of the patient's history involves collection of data on pertinent medical history, physical exam findings, and labs. Much of these data are available in electronic medical records (EMR), which can be reviewed and provide potential insights on patient outcomes.

Although EMR data is typically rigorously secured to guard patient privacy, some institutions have undertaken the task of anonymization and publication of data to allow for research without the usual hurdles that accompany the use of an unredacted EMR. One such dataset is the Innovative Solutions for Pioneering Intensive Care Research (INSPIRE)[2,3]. INSPIRE includes 130,000 anesthetic cases from an academic medical center in Korea and contains information from pre-procedure evaluation to hospital discharge or death. Patient demographics, labs, medications, diagnoses, and vital sign measurements are provided. The following is a summary of in hospital deaths for procedures labeled as emergency or not, broken down by ASA physical status. ASA physical status gives a rough grade of overall patient health that often correlates with outcomes[4].

As seen in Table 1, in hospital mortality is not common among the healthiest patients (ASA 1 and 2) undergoing non-emergency procedures, with lest than 1% of patients dying in the hospital. Death occurs with greater frequency for patients with more severe disease as seen for patients who are designated as ASA 3 or 4 physical status. Emergency procedures cannot be delayed, by definition. Some of the deaths seen for non-emergent procedures, however, might be avoided if the factors that are influential mortality could be intervened upon. Knowing which features relate to mortality can also allow for better risk stratification and patient discussion for informed consent.

Methods

Cases were filtered to focus on elective procedures, to the degree possible using available data. Procedures that were classified in the operations table as emergency were excluded. To further reduce the number of likely non-elective cases, any patient who was recorded as being admitted to the ICU prior to entry to the OR was excluded. Vital sign records demonstrated that some patients had need of a ventilator, the use of extracorporeal membrane oxygenation (ECMO), or intra-aortic balloon pump (IABP) prior to surgery. These devices are only used in critically ill patients; any procedures done are extremely unlikely to be elective, so these patients were excluded from this review. ASA category 5 is typically reserved for patients who are critically ill and are not likely to survive without an operation; making their procedure almost by definition non-elective. Patients who are brain dead are classified as ASA 6 and are physically alive until such a time as organ harvesting can take place, so these patients were not considered. Since subsequent procedures and hospitalizations may represent complications of the initial procedure, which would not be in keeping with the focus on elective procedures, only the first procedure of the first hospitalization was considered.

Impossible values of heights and weights were removed, and BMI was calculated for use as a covariate. BMI values over 40 were only retained if the patient had an ICD-10 code related to being overweight or obese. Vital signs and laboratory values were selected from those that were reported prior to the time that patients entered the operating room. Missing vital signs and lab values were imputed with normal values under the assumption that values would be normal if there was no indication to check. Considerations on which diagnosis were considered were selected based on whether their chart time was entered prior to entry into the operating room to avoid making predictions based on events that occurred later in the patients hospital course. ASA 1 physical status is reserved for patients who are healthy and have medical comorbidities, so patients were considered to have complete diagnostic information if some diagnostic record was found for them or they were classified as ASA 1 physical status. Patients without diagnostic information were excluded. The data were split into a training and test set to allow for validation of models. Patients with incomplete data were excluded from analysis.

Although the ICD-10 codes for procedures and diagnoses were truncated for the purposes of anonymization in INSPIRE, there were still over 1000 unique diagnostic codes and over 2000 unique procedure codes. A random forest with 200 trees was generated to determine the importance of features as a way to narrow down this list for further consideration. Cofactors were given further consideration if they had some predictive utility in predicting both the outcome of death or not death according to the importance measure of the random forest. In order to create a model with interpretable parameters, logistic regression models were created. An initial model was created based on those features that were identified as important by the random forest tree. Further selection of parameters was based on clinical judgment, assessment of collinearity, and statistical significance. Due to the rareness and severity of the outcome, a Bonferroni correction was not employed when selecting features for the final model to allow for a larger number of possible predictors of mortality. Descriptions for the ICD codes from the final model were referenced manually using an online database[5]. In the final model, laboratory values were considered in comparison with normal reference ranges[6] to aid in the interpretation.

Results

After filtering the data to consider cases determined to be elective and including only a patient's first procedure, there were 90,572 patients. Missing data was found in 3.8 percent of these, leaving 87,117 patients, which were divided into 69,740 for the training set and 69,740 for the test set. Demographics for these patients are summarized in Table 2.

There were a total of 2,709 predictors considered when generating the random forest model. Figure 1 illustrates the receiver operating characteristic (ROC) curves for the random forest model when applied to the training and test sets. Notably, the area under the ROC (AUC) on the training set was 1.0, while the AUC on the test set was 0.815, indicating that there was substantial overfitting of the random forest on the training set. Using the `importance` function from the `randomForest` package, 117 were selected for further consideration. Of these, 6 were specific ICD10 procedure codes and 55 were ICD10 diagnosis codes, a significant reduction from the several thousand ICD10 codes under initial consideration.

When using the important features found by the random forests model, there is a notable improvement in the AUC applied to the test set to 0.870, at the cost of a reduction in training set AUC to 0.891, which is depicted in Figure 2. When creating the final model that would be the most interpretable, it was noted that there were still many correlations between variables and several covariates had variance inflation factors (VIF) far above 2. Based on this, several cofactors were eliminated. For example hematocrit, hemoglobin concentration, and red blood cell count all essentially measure the same health feature of a patient's oxygen carrying capacity. Based on the association with death according to the first logistic regression model, the number of covariates was reduced to 31 in the final model. BMI was divided into categories of Underweight, Normal, Overweight, and Obese to account for possible non-linearity of BMI, but without losing interpretability that would occur with the use of polynomial terms or splines. Anesthesia type of Neuraxial (which refers to epidural and spinal anesthetics) vs the other categories of General Anesthesia, Regional Anesthesia (using local anesthesia to block a peripheral nerve), and monitored anesthesia care (MAC; typically sedation to the level that does not require an endotracheal tube or a laryngeal mask). No vital sign was found to have a statistically significant association with in hospital mortality; 5 laboratory values, 4 procedures, and 16 diagnostic categories were found to be relevant according to these data, and included in the final model.

The final logistic regression model contained 35 parameter estimates (including the intercept, 2 for ASA, and 3 for BMI). The ROC curves are depicted in Figure 3. When compared to the linear model using all important features, the AUC reduces from 0.892 to 0.864 in the training set. VIF values were below two for all covariates in the final model. In the test set, the change in AUC is from 0.866 to 0.858. The odds ratio with 95% confidence intervals for each of these estimates are listed in Table 3. When considering ASA physical status, ASA 2 patients are defined as having well controlled systemic disease and are considered healthy; there did not appear to be a difference between ASA 1 and ASA 2 patients in terms of survival; these two categories were combined for comparison. ASA 3 patients had 2.08 times odds of in hopspital mortality when compared to ASA 1 or 2 patients (95% CI: 1.58 to 2.73), while ASA 4 patients showed an odds ratio of 3.11 (95% CI: 1.20 to 8.08). Age should be interpreted as the odds ratio for a one unit change: the odds of in hospital death after surgery for a one year increase in age is 1.03 times the odds of death of of the current age, holding all other covariates constant, according to these data (95% CI: 1.02 to 1.04). When considering gender, being categorized as male demonstrated a 1.80 fold odds of in hospital death compared to being categorized as female (95% CI: 1.42 to 2.28). An underweight (< 18.5) BMI is associated with with an increased odds of death compared to normal BMI (OR: 2.21, 95% CI: 1.62 to 3.01), while being overweight showed a lower association (OR: 0.49, 95% CI 0.36 to 0.67); an Obese BMI did not show a statistically significant difference from normal BMI (OR: 0.84, 95% CI 0.47 to 1.52). The diagnostic category for that showed the second highest odds ratio for in hospital mortality was "Artificial opening (Z93)", which refers to a patient having an artificially created opening such as a tracheotomy, cystostomy, or colostomy.

Discussion and Conclusion

The three different models created here highlight the bias variance trade off. The random forest model showed no bias on the training set, but the result was an ROC that was worse on the test set than simply using a subset of the predictors on a simpler model. Moving to the final model that prioritized interpretability did cause a further reduction in ROC, but this reduction was modest and allowed for a model that is relatively easy to understand. Using other techniques that balance bias and variance would likely have resulted in a model with much better predictive power on the test set than any of the models created here, but that was not explored at this time.

The goal of the present project was to focus on elective surgery. It is unclear, however, whether this was truly the group of patients represented here, and whether it is possible to properly make this selection based on information made available by INSPIRE. Simply relying on the emergency indicator that was present in the `operations` table was insufficient, given that it was necessary to further exclude patients who were using ECMO and other highly invasive technologies prior to surgery. In the final data, one of the predictors of mortality was the presence of complications from transplant. There did not appear to be a way to determine if this was an indication for the patient's procedure; a patient going to the operating room due to organ rejection is not elective. On the other hand, a patient who had undergone a transplant in the past that was complicated by rejection may later present for elective surgery. Performing a similar project as this, but either on an anonymized data set that provides indicators with more clarity with respect to elective surgery, or with direct access to a full EMR might provide different results.

Many of the covariates found to be associated with mortality in this data set appear to relate to severe disease such as having cancer, liver failure, or other disorders that cannot be reasonably reversed. The procedures on these patients may not have been emergent, but it would not be reasonable to delay care for a patient in the hopes that they will be cured, particularly if the procedure is for palliation or managing the complications of a disease.

This project should also be considered in its clinical context. The INSPIRE dataset is from a single center in Korea. As a consequence, all of the patients in this study were classified as "Asian". The rate of Obesity in Korea, and in this dataset, is much lower than that found in the United States or other western countries, which may limit the generalizability. There were some prognostic features that were not provided, such as smoking status, that would likely have been useful in making predictions and may have confounded the relationship between some of the covariates, such as BMI, and the outcome of operative mortality. Smoking status was not documented in INSPIRE, however.

"Vital Signs are vital" is a common phrase in medicine. There did not, however, appear to be a relevant relationship between vital signs and in hospital mortality, and they were not included in the final model. First, a lack of evidence does not indicate evidence for a lack of effect. Even if it were, it should not be concluded that vital signs are not actually vital. Given the attempt to remove emergency procedures from consideration, it is more likely that the lack of evidence for an effect on vital signs is that the personnel in Korea have the good judgment not to take a person to surgery in the presence of severe hypertension, hypoxemia, or pyrexia. The absence in utility for commonly considered laboratory values may have a similar explanation.

In summary, this study highlights many factors that are relevant to in-hospital mortality following non-emergency procedures and demonstrates the potential of EMR data in considering perioperative outcomes, but may be limited in generalizability and applicability to truly elective procedures.

Tables and Figures

Table 1: In Hospital Death Rate (%)

ASA Physical Status	Non-Emergency Procedure	Emergency Procedure
1	0.2	1.3
2	0.4	3.2
3	3.2	12.5
4	9.5	23.3
5	NA	40.0
6	25.0	100.0

Table 2: Demographic Information

	Overall	Train	Test
n	87117	69740	17377
Death = Y (%)	439 (0.5)	351 (0.5)	88 (0.5)
Age (mean (SD))	55.01 (15.79)	55.04 (15.80)	54.93 (15.75)
Sex = M (%)	37611 (43.2)	29943 (42.9)	7668 (44.1)
Race = Asian (%)	87117 (100.0)	69740 (100.0)	17377 (100.0)
Height in cm (mean (SD))	161.93 (8.88)	161.88 (8.86)	162.13 (8.93)
Weight in kg (mean (SD))	62.92 (11.79)	62.89 (11.79)	63.02 (11.78)
BMI (mean (SD))	23.92 (3.55)	23.92 (3.56)	23.90 (3.51)
ASA Physical Status (%)			
1	34921 (40.1)	27959 (40.1)	6962 (40.1)
2	46955 (53.9)	37625 (54.0)	9330 (53.7)
3	5099 (5.9)	4045 (5.8)	1054 (6.1)
4	142 (0.2)	111 (0.2)	31 (0.2)
Anesthesia Type (%)			
General	70387 (80.8)	56310 (80.7)	14077 (81.0)
MAC	8817 (10.1)	7087 (10.2)	1730 (10.0)
Neuraxial	7841 (9.0)	6289 (9.0)	1552 (8.9)
Regional	72 (0.1)	54 (0.1)	18 (0.1)

Table 3: Odds Ratios of In Hospital Mortality

Covariate	Odds Ratio	95% CI	p value
Age	1.03	(1.02 to 1.04)	1.44e-11
Male Sex	1.80	(1.42 to 2.28)	1.16e-06
ASA 3 (vs 1 or 2)	2.08	(1.58 to 2.73)	1.86e-07
ASA 4 (vs 1 or 2)	3.11	(1.20 to 8.08)	0.020
Obese BMI (vs Normal)	0.84	(0.47 to 1.52)	0.572
Overweight BMI (vs Normal)	0.49	(0.36 to 0.67)	9.07e-06
Underweight BMI (vs Normal)	2.21	(1.62 to 3.01)	5.16e-07
Neuraxial Anesthesia (vs General/Regional/MAC)	0.42	(0.24 to 0.73)	2.13e-03
Cardiothoracic Surgery (vs All Other Departments)	1.88	(1.37 to 2.59)	8.93e-05
Low Albumin (< 3.5 g/dL)	2.77	(2.06 to 3.73)	1.64e-11
Low Chloride (< 98 mmol/L)	2.75	(1.94 to 3.90)	1.33e-08
High Fibrinogen (> 450 mg/dL)	2.36	(1.71 to 3.26)	1.81e-07
Low Lymphocyte Count (< 4.0/ng)	2.80	(1.22 to 6.44)	0.015
Low Platelet Count (< 150/ng)	1.68	(1.23 to 2.28)	9.51e-04
Open Inspection of Peritoneal Cavity (0WJG0)	5.44	(2.99 to 9.89)	2.75e-08
Open Excision of Cervical Vertebra (0PB30)	4.91	(1.47 to 16.40)	9.67e-03
Percutaneous Excision of Bladder (0TBB4)	6.32	(1.98 to 20.23)	1.88e-03
Percutaneous Stomach Bypass (0D164)	6.96	(2.26 to 21.41)	7.16e-04
Angina pectoris (I20)	2.05	(1.38 to 3.06)	4.06e-04
Osteoporosis (M81)	3.17	(1.62 to 6.21)	7.54e-04
Hematuria (R31)	2.34	(1.27 to 4.31)	6.11e-03
Artificial opening (Z93)	20.34	(7.01 to 59.01)	2.95e-08
Malignant neoplasm of kidney (C64)	3.54	(1.73 to 7.24)	5.49e-04
Malignant neoplasm of liver (C22)	1.93	(1.22 to 3.06)	5.22e-03
Malignant neoplasm of brain (C71)	7.91	(3.61 to 17.31)	2.31e-07
Rheumatic tricuspid valve disease (I07)	13.34	(5.26 to 33.85)	4.91e-08
Hypertensive heart disease (I11)	2.58	(1.13 to 5.90)	0.025
Complications of cardiac and vascular prosthetic devices (T82)	4.94	(2.22 to 10.98)	9.03e-05
Nephritic syndrome (N05)	4.34	(1.57 to 12.00)	4.73e-03
Disorders of fluid, electrolyte and acid-base balance (E87)	4.04	(1.22 to 13.31)	0.022
Lichen simplex chronicus and prurigo (L28)	7.11	(1.96 to 25.77)	2.84e-03
Myeloid leukemia (C92)	10.02	(3.38 to 29.73)	3.29e-05
Malignant neoplasm of retroperitoneum and peritoneum (C48)	9.85	(2.33 to 41.70)	1.89e-03
Complications of transplanted organs and tissue (T86)	21.92	(6.66 to 72.11)	3.73e-07

Figure 1: Random Forest ROC curves

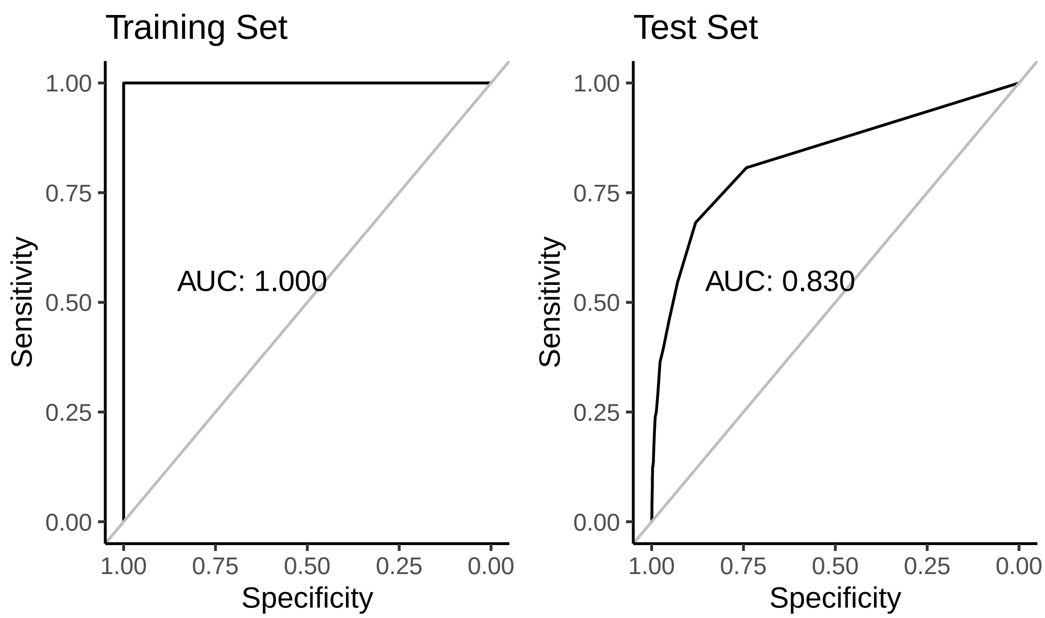


Figure 2: ROC Curves for Logistic Model With All Important Features

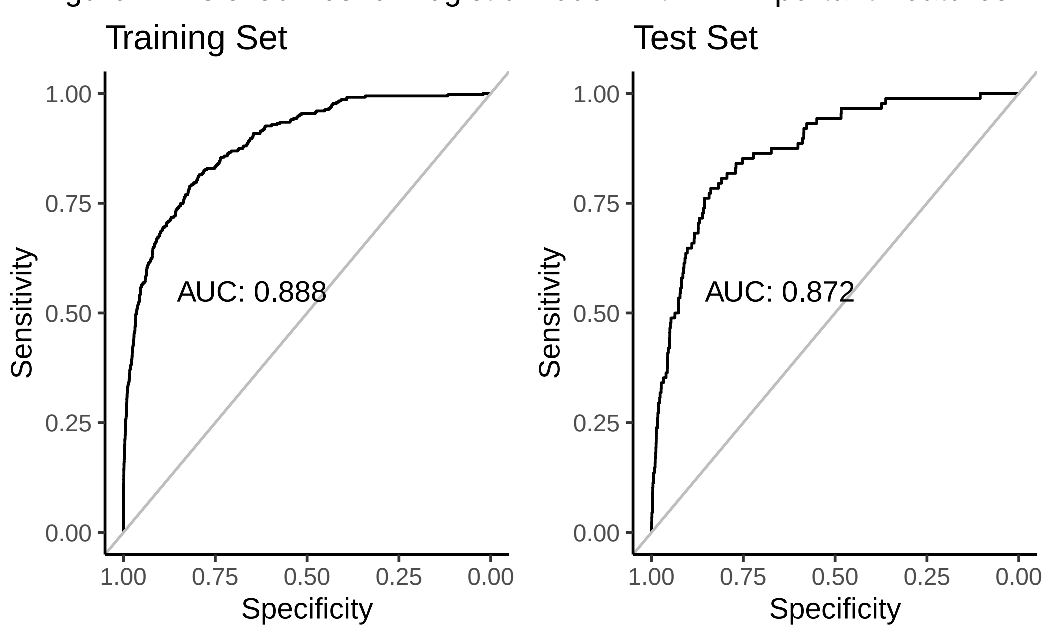
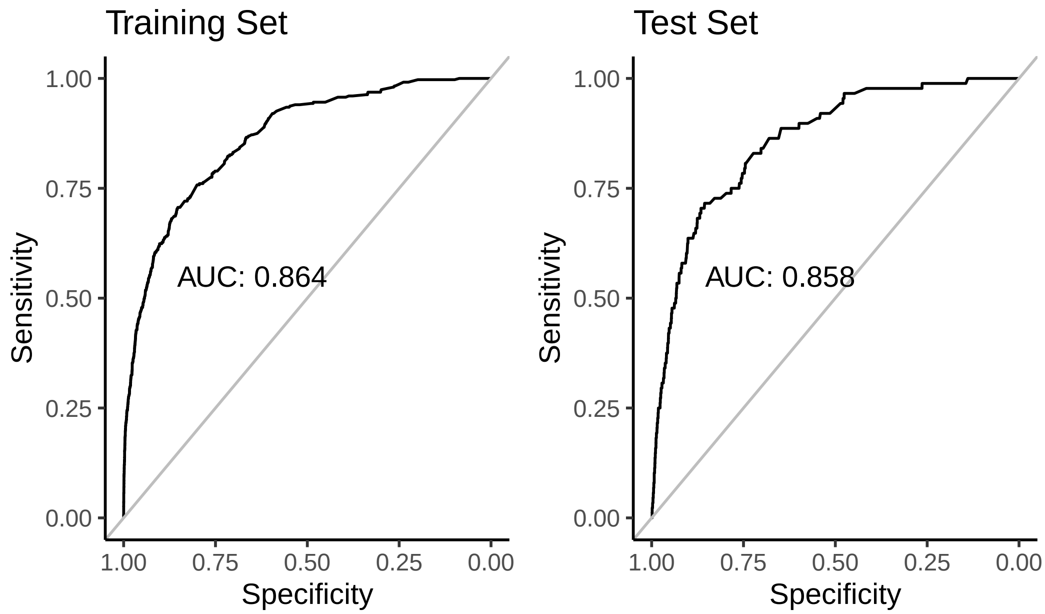


Figure 3: ROC Curves for Final Logistic Regression Model



References

1. Committee on Standards and Practice Parameters *et al.* Practice advisory for preanesthesia evaluation: An updated report by the american society of anesthesiologists task force on preanesthesia evaluation. *Anesthesiology* **116**, 522–538 (2012).

2. Lee, H.-C. & Lim, L. INSPIRE, a publicly available research dataset for perioperative medicine. doi:10.13026/JYZB-EZ61.

3. Goldberger, A. L. *et al.* PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* **101**, E215–220 (2000).

4. Horvath, B., Kloesel, B., Todd, M. M., Cole, D. J. & Prielipp, R. C. The evolution, current value, and future of the american society of anesthesiologists physical status classification system. *Anesthesiology* **135**, 904–919 (2021).

5. ICD10Data.com. https://www.icd10data.com/.

6. American College of Physicians. Laboratory values. https://annualmeeting.acponline.org/sites/default/files/shared/documents/for-meeting-attendees/normal-lab-values.pdf.