DESIGN AND VALIDATION OF TWO PREDICTIVE MODELS FOR MORTALITY AND READMISSION FOLLOWING SURGERY IN PATIENTS WITH LIVER CIRRHOSIS

Monica L. Schmidt

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Health Policy and Management in the Gillings School of Global Public Health.

Chapel Hill
2015

Approved by:

Morris Weinberger

Andrea K. Biddle

Kristin Reiter

A.Sidney Barritt IV

Paul H. Hayashi

ABSTRACT

Monica L. Schmidt: Design and Validation of Two Predictive Models for Mortality and Readmission Following Surgery in Patients with Liver Cirrhosis
(Under the direction of Morris Weinberger)

Cirrhosis is the 12th leading cause of death in the United States. By 2020, it is expected to affect more than 1 million Americans. Cirrhosis is a costly, chronic condition requiring frequent hospitalizations and unplanned readmissions. Patients with cirrhosis often require routine surgeries including hernia repair, coronary artery by-pass surgery and orthopedic hip or knee replacements. These procedures present a greater risk of morbidity and mortality for cirrhotic patients, including a 8-fold increase in risk of mortality and higher hepatic decompensation after surgery. Predicting post-operative mortality prior to surgery or post-operative readmission would allow patients and clinicians to make informed decisions that optimize survival and reduce readmission costs. Currently, the MELD score is often used, inappropriately, to assess risk of procedures in patients with decompensated cirrhosis. However, no models exist that predict mortality or readmission among patients with cirrhosis. This study aimed to develop and validate two predictive models; one for mortality among all patients with cirrhosis undergoing surgery; the other will predict readmission among cirrhosis patients discharged alive after the index surgery. Each two model was then compared to the MELD score. The NSQIP Mortality Model was significantly better than the MELD score at predicting mortality ($p<0.001$) and had an area under the receiver operating characteristic (AUROC) curve of 0.84. The readmission model was also significantly better than the MELD score ($p<0.001$), with an AUROC of 0.75. Both models

provide the basis for developing two decision tools that can assist clinicians and patients in making informed decisions that optimize survival and reduce unplanned readmissions.

ACKNOWLEDGEMENTS

I am grateful for the support from both family and mentors during the last five years. I would like to first thank my husband for his continuous support during this endeavor. His kind words of encouragement kept me going when facing challenges. I would like to thank my mother and father for their support.

I have received unprecedented mentorship while at the University of North Carolina. My current chair, Morris Weinberger, has been instrumental in guiding me through the stages of the dissertation. His mentorship cannot begin to be repaid during my lifetime. He kept me focused and inspired. He offered professional mentorship outside of the requirements of the dissertation. I am forever in his debt.

In addition to my chair, other mentors have been instrumental in facilitating my growth as a researcher. First, I would like to thank Michael W. Fried. He took me under his wing early in my career and has supported my endeavors for the last five years. Andrea K. Biddle was my advisor, my friend and my confidant during tough times. Without her, I would not have had the opportunity to participate in the doctoral program. Drs. A. Sidney Barritt IV and Paul H. Hayashi are responsible for shooting my career forward at lightning speed. They guided me through my first publication experience with unimaginable expertise. Their deep desire to improve patient care inspires me daily. Kristin Reiter has been instrumental in sharing her knowledge with me during the dissertation process. Mark Holmes is amazing and without him, I would be lost in the methods. He truly cares about his students and takes time to ensure everyone succeeds. I am forever in his debt. I hope to carry the skills acquired in the Health Policy and Management

doctoral program into an exciting career that improves the quality of care for patients with liver diseases. Accepting such a challenge is truly the best way to repay my mentors for their time and guidance.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AASLD	American Association for the Study of Liver Diseases
ACS	American College of Surgeons
ALT	Alanine transaminase
ANOVA	Analysis of Variance
APRI	Aspartate transaminase to platelet Ratio
ASA Class	American College of Surgeons Classification
AST	Aspartate transaminase
AUROC	Area Under the Receiver Operating Characteristic Curve
CMS	Centers for Medicare and Medicaid Services
CPT	Current Procedural Terminology
DCA	Decision Curve Analysis
DNR	Do Not Resuscitate
eGFR	Estimated Glomerular Filtration Rate
FN	False Negative
FP	False Positive
IRB	Institutional Review Board
IRR	Incident Risk Ratio
MELD-Na	Model for End-Stage Liver Disease with Sodium
MELD	Model for End-Stage Liver Disease
NSQIP	National Surgical Quality Improvement Program
OR	Odds Ratio
PT INR	Prothrombin International Normalized Ratio

SE	Standard Error
TIPS	Transjugular Intrahepatic Portosystemic Shunt
TN	True Negative
TP	True Positive
UNC	University of North Carolina

CHAPTER 1. SPECIFIC AIMS

Liver cirrhosis is a costly chronic condition requiring frequent hospitalizations, often for surgery. When undergoing surgery, patients with cirrhosis are at high risk for mortality and hepatic decompensation as well as for readmission to the hospital following discharge. Predicting their post-operative mortality as well as readmission post-discharge would allow patients and clinicians to improve: (1) treatment decision-making for patients and (2) process and outcomes of care by having hospitals implement strategies to reduce complications leading to readmissions and death. Currently, the Model for End-Stage Liver Disease (MELD) score is often used, inappropriately, to assess risk of procedures in patients with decompensated cirrhosis. However, no models exist that predict mortality or readmission among patients with cirrhosis.

The long-term objective of this research is to improve the quality of life for patients with cirrhosis. The immediate objective of this study is to develop and validate models that predict mortality for patients with cirrhosis undergoing surgery as well as their risk of readmission following hospital discharge. Each model will be compared to the MELD. The central hypothesis is that we can develop validated models that perform better than the MELD. The rationale for the proposed research is that these predictive models will be able to: 1) improve surgical decision making and 2) inform allocation of appropriate post-discharge resources within risk-stratified discharge plans that reduce readmissions.

Aim 1: The Mortality Model

To develop and validate a mortality predictive model and compare it to the currently used MELD score.

Aim 1a: To develop a model that predicts increased risk of in-hospital mortality after the index surgery.

Based on previous literature, age, gender, body mass index, functional status, ethnicity/race, American Surgical Association classification (ASA class), estimated glomerular filtration rate, serum albumin, platelet count, white blood cell count, total bilirubin, prothrombin time international normalization ratio (INR), emergent surgery, surgery type, and specific comorbid conditions will increase the risk of in-hospital mortality in patients with cirrhosis after the index surgery.

Aim 1b: Using results from Aim 1a, validate the mortality predictive model.

Hypothesis: The predictive model will have an area under the receiving operating characteristic curve (AUROC) of at least 0.75 when using cross-validation methods.

Aim 1c: To compare the mortality model to the MELD score.

Hypothesis: The predictive model will have significantly greater AUROC than the MELD score.

Aim 2: The Readmission Model

Aim 2a: To develop a model that predicts the risk of unplanned readmission within 30-days from discharge after the index surgery.

In addition to all variables used in the mortality model, post-operative variables will be added to include: pneumonia, sepsis, time on operating table, total time under anesthesia,

acute/progressive renal failure, urinary tract infection, ventilator wean time >48 hours after surgery, and discharge destination.

Aim 2b: Using results from Aim 1b, validate the readmission predictive model.

Hypothesis: The predictive model will have an area under the receiving operating characteristic curve (AUROC) of at least 0.75 when using cross-validation methods.

Aim 2c: To compare the readmission model to the MELD score.

Hypothesis: The predictive model will have significantly greater AUROC than the MELD score.

The two models can inform targeted measures to reduce inpatient mortality and readmission within 30 days of the index surgery. The mortality model would be used prior to surgery during the consultation between the hepatologist, surgeon, and patient to inform the risk associated with surgery. This information could be used to discuss the benefits and risks of alternative treatment options (including the decision to undergo surgery). The readmission model would be used prior to discharge to help a hospital's transitional care team provide appropriate discharge planning that may mitigate readmission, an increasingly important outcome to hospitals given the Centers for Medicare and Medicaid Services (CMS) Hospital Readmissions Reductions Program. Either of these models could feed decision support, tools that capitalize on increased access to clinical data from electronic medical records.

CHAPTER 2. BACKGROUND & STUDY OVERVIEW

Background

Cirrhosis is the twelfth-leading cause of death in the United States[1]. By 2020, it is expected to affect more than 1 million Americans[2]. Cirrhosis is a costly, chronic condition requiring frequent hospitalizations and unplanned readmissions[3, 4]. The average-length stay for patients hospitalized with cirrhosis is eight days with a mean charge of $46,663 per stay if decompensated[5]. Cumulative one- and five-year risks for readmission were 45% and 83% in a study of 200 patients with cirrhosis.[3].

Patients with cirrhosis often require routine surgeries to maintain their quality of life. Hernia repair, coronary artery by-pass surgery, and orthopedic hip or knee replacements are common as these patients age[6]. These procedures present a greater risk of morbidity and mortality for cirrhotic patients[6, 7], including an eight-fold increase in risk of mortality[8] and increased hepatic decompensation[6, 9]. A patient with liver disease who has any of the following conditions is decompensated: ascites (fluid) in the abdominal cavity, hepatic encephalopathy (swelling of the brain), portal hypertension, or variceal hemorrhage (bleeding in esophagus). Decompensation increases the risk of in-hospital mortality as well as unplanned readmissions among those who survive to discharge[10-12].

Predicting post-operative mortality prior to surgery would allow patients and clinicians to make informed decisions that optimize survival[13]. Patients who die in-hospital after a surgical procedure may incur a prolonged stay in acute care. The cost of a prolonged stay would not be fully covered if the bundled service payment allows for a short stay in acute care after surgery.

Predictive models that allow for preventative services to be delivered to patients at greater risk of in-hospital mortality (i.e. prolonged stay) or readmission may mitigate financial losses experienced by hospitals under the bundled payment initiative[14, 15].

Readmissions are problematic for patients and may represent a problem in the quality of care during the index admission or the post-discharge period. The urgency to reduce readmissions is fueled by financial penalties imposed on hospitals by the Centers for Medicare and Medicaid Services (CMS) for Medicare patients with 30-day readmissions [16, 17]. As the cirrhotic population ages, more will have Medicare as their primary payer and expose hospitals to the risk for significant financial penalties [17]. Other payers will likely adopt CMS's financial incentives to reduce 30-day readmissions. As CMS begins its bundled payments for care , it will be critical to a hospital's financial health to know the risk of a prolonged stay or complication prior to the service provided (e.g., surgical procedure). The ability to predict the risk of readmission among those who survive surgery would allow for appropriate discharge planning[18].

Predictive Models in Medicine to Improve the Quality of Care

Making decisions using complex information is standard practice in medicine. Clinicians are expected to make decisions by evaluating immense quantities of data and, using evidence-based guidelines, apply those data to a specific patient. For example, when choosing to take a patient with cirrhosis to surgery, hepatologists and surgeons consider laboratory, history, social, and clinical factors that might impact outcomes[19, 20]. There is a need for a model that can deliver an evidence-based, patient-centered risk assessment of the cirrhotic patient to assist with this difficult clinical decision.

Electronic health records make it more feasible to incorporate patient data into predictive models that can give the clinician a quantitative risk score for readmission or mortality[21]. Predictive models may assist physicians when making decisions that involve complex patients instead of relying solely on their prior experiences[22, 23]. Ultimately, these models can be used as the foundation for shared decision-making and targeted discharge planning[24]. In an era when population health outcomes are linked to reimbursement, there is a need for predictive models to inform decisions and allocate resources [25].

The MELD Score

The Model for End-Stage Liver Disease (MELD) score was originally developed and validated to predict mortality in patients with cirrhosis undergoing transjugular intrahepatic portosystemic shunt (TIPS) procedures. It was validated with 231 patients undergoing elective TIPS procedures to prevent hemorrhage or to treat refractory ascites[26, 27]; 70 patients died within 90 days of the procedure. Variables included in the MELD score are serum creatinine, total bilirubin, and prothrombin INR as follows:

MELD Score = (0.957 * ln(serum creatinine) + 0.378 * ln(serum bilirubin) + 1.120 * ln(PT INR) + 0.643) * 10 (if hemodialysis, value for creatinine is automatically set to 4.0)

Although the MELD score is attractive because it uses objective, commonly available laboratory values, there are many weaknesses. First, inter-laboratory variability in creatinine measurement causes poor predictive power of the MELD score[28-30]. Replacing creatinine with estimated glomerular filtration rate (eGFR) could mitigate this problem by offering more robust measures of liver biosynthetic function[31, 32]. Second, there is evidence that sodium, albumin, and platelet counts are important predictors of liver function and should be included in the predictive model[33-39]. In fact, the MELD score was modified to include sodium in 2006.

However, this modified MELD-NA has not been fully accepted [35]. Third, MELD score performs poorly when assessing the immediate risk of mortality after surgery. In the largest study to date, Teh and colleagues used 772 patients having any major surgical procedure and cirrhosis to assess the MELD score as a model for predicting post-surgical mortality [7]. Although they found the MELD score to predict 30-day post-surgical mortality with some accuracy, it was not accurate zero to 7 days after surgery[7]. Moreover, MELD scores fail to include important predictors of mortality such as age [7]. In addition, gender influences the MELD score's ability to predict short-term mortality in patients awaiting transplant[40, 41]. Differences in creatinine levels and height between males and females result in increased MELD scores for women rather than the severity of liver dysfunction[40, 41]. Finally, the MELD score is a poor predictor of mortality in the presence of complications of cirrhosis such as persistent ascites and hepatorenal syndrome [28]. These factors limit the usefulness of the MELD. Thus, it is critical to develop and validate a predictive model that is robust when predicting post-surgical mortality across different ages, genders, and varying degrees of liver dysfunction.

Beyond the weakness in using the MELD score to predict mortality, little is known about its ability to predict readmissions in patients with decompensated cirrhosis[4]. The only two single-site studies to date found that MELD predicted readmission in patients with decompensated cirrhosis when controlling for serum sodium, number of medications, gender, number of comorbidities, transplant list status, and discharge destination[4, 12]. However, both studies were small and had limited generalizability. Therefore, there is a need to develop and validate a model to predict hospital readmission for post-surgical patients with either compensated or decompensated cirrhosis.

Study Overview

Currently, clinicians use the MELD score to guide surgical treatment decisions [42, 43]. To date, there are no valid predictive models to assess risk of mortality or readmission for patients with cirrhosis. My dissertation will address this gap in the literature by developing and validating two predictive models that will accurately stratify patients with cirrhosis by their risk of in-hospital mortality and 30-day readmission after surgery.

The Proposed Predictive Models: Theory and Selection of the Measures

A thorough literature review and discussion with hepatologists and surgeons allowed a priori selection of variables for consideration in the predictive models for both in-hospital mortality and readmission (Table 1) for common surgical procedures undergone by patients with cirrhosis (Table 2).

Institutional Review Board Approval

This study has been given an exemption (IRB No. 13-3559) by the Office of Human Research Ethics at the University of North Carolina. It is a limited dataset and does not involve subject contact. The study was deemed exempt and updated to the American College of Surgeons (ACS) National Surgical Quality Improvement Program (NSQIP) on November 11, 2013, and renewal is not required.

CHAPTER 3. THE NSQIP MORTALITY MODEL

Introduction

The ability to predict post-operative mortality prior to surgery would allow patients and clinicians to make informed decisions that optimize survival[13]. Patients who die in-hospital after a surgical procedure may incur a prolonged stay in acute care. The cost of a prolonged stay would not be fully covered if the bundled service payment only allows for a short stay in acute care after surgery. Predictive models that allow for preventative services to be delivered to patients at greater risk of in-hospital mortality (i.e., prolonged stay) may mitigate financial losses experienced by hospitals under the bundled payment initiative[14, 15]. This study aimed to develop, validate, and compare a mortality model to the MELD score.

Methods

Data Source and Sample Selection

This study used American College of Surgeons (ACS) National Surgical Quality Improvement Program (NSQIP) data from 2011–2013[44]. NSQIP is a national voluntary program to help hospitals reduce surgical morbidity and mortality. The program offers three levels of participation that vary by labor intensity to allow smaller hospitals with limited resources to participate. To date, more than 300 hospitals nationwide participate in the NSQIP[44].

Data are collected through a combination of electronic and manual chart abstraction at each site. More than 150 variables are collected on surgical cases. Each year, hospitals rotate through 46 eight-day cycles of data collection to capture a random sampling of surgical cases.

Included are patients having a major surgery, defined by Current Procedural Terminology (CPT) code, during one of the eight-week data collection cycles. Both emergent and non-emergent surgeries are captured. Trauma and transplant surgeries are excluded[44, 45].

Data are collected at the patient level. Pre-, intra-, and post-operative variables are collected for each patient. If a patient died prior to discharge, he or she is captured in the discharge status variable with an "expired" status[44].

Identifying Patients with Cirrhosis

I selected patients with liver disease, defined as having an esophageal hemorrhage or ascites as identified by the ACS NSQIP [44]. The ACS NSQIP definition for these variables requires a history of liver disease on record concurrent with an esophageal varices or ascites diagnosis[44, 45]. For these patients, previously validated laboratory values were used to define liver cirrhosis: platelet value of $\leq$ 140,000/mm^3 and albumin <3.5 g/dL and prothrombin INR>1.5 [45-47]. Finally, patients with disseminated cancer, an outpatient surgery, or a do not resuscitate (DNR) order on file, as defined by a variable in the NSQIP, were excluded. This resulted in a final cohort of patients with liver cirrhosis (n=4,916).

Mortality Model Cohort

For aims 1a, 1b, and 1c, all patients with cirrhosis were included. After excluding 273 patients with missing laboratory data, I used variables from the NSQIP to identify patients who died while hospitalized and were not discharged after the index surgery (Figure 1). Of those with cirrhosis, 608 died while hospitalized and 4,308 were discharged alive. Our final cohort for analysis in aims 1a, 1b, and 1c will include patients with laboratory values available to generate the MELD score and mortality model (died: n=543) and (n=4,100 discharged alive).

Bivariate Analyses

Patients with and without an in-hospital death were compared using analysis of variance (ANOVA), chi-square analyses, and non-parametric test for trends in ordered variables. Eligible patients with and without pre-operative laboratory values were compared by ANOVA or chi-square analysis to determine if their baseline characteristics differ significantly ($p<0.05$). Those with and without laboratory data by surgery type were evaluated to further understand any bias that may be introduced into the model. It may be the less invasive surgery types had fewer labs ordered by the surgeon due to the perception of lower risk of mortality. Any significant differences are noted in limitations. It is not possible to control for selection bias in this sample when we do not have physician- or hospital-level factors necessary to generate an appropriate selection model.

Development of the Multivariable Mortality Model

A Poisson model with binary in-hospital mortality as the outcome was used. This model is has two advantages over logistic regression [48, 49]. First, it computes incident risk ratios (IRR) as relative risks, which facilitates interpretation[49, 50]. Second, when cross-sectional data have frequent outcomes of interest, odds ratios (OR) may overestimate the prevalence in the sample[50]. Poisson is appropriate for a binary outcome model when relative risk is desired and events of interest are frequent. The Poisson model requires Huber-White standard error adjustment (robust standard errors) to avoid under-dispersion[49, 50]. If a variable was not significant, I tested for joint significance ($p<0.05$). If there was joint significance with other variables in the model, I included these measures in the final model for validation.

Validation of the Mortality Model

After determining the predictors of mortality using a Poisson model, a logistic regression maximum likelihood model with a binary outcome (mortality) was used to obtain in-range predictions. Predictions for Poisson models (used for development) are often not appropriate for validation methods discussed earlier and shown in Table 1 [51]. I will report ORs for the mortality estimates. Predictions from the model will be compared to those that actually died in-hospital after the index surgery. I am interested in the predictions from the model and how they compare to the actual outcome. I will obtain the predicted probability of death for each patient in the sample. The actual outcome is binary with 1 equal to death. Having a binary outcome makes it difficult to compare predictions that are continuous between 0 and 1. It will be necessary to perform a cut-point (i.e., threshold) analysis to determine where along the range of predicted probabilities the sensitivity and specificity are optimized (e.g., where do you call a "positive" result=death versus a negative result=survive)[52, 53].

Assessment of Mortality Model Performance

Overall model performance will be assessed with the Brier score [52, 53] (Table 3). Brier scores measure the distance between the actual outcome and the predicted outcome[53]. Brier scores are similar to the Pearson's R^2 statistic. It is a quadratic scoring equation that uses the squared differences between the actual binary outcome (died/survive) and the predictions (p). [Brier score=$(Y-p)^2$] [52]. The lower the Brier score, the better the predictions are calibrated to the actual outcome; for a perfect model with optimal fit, the Brier score equals zero. The Brier score is affected by the incidence of the outcome of interest in the population. When the incidence of the outcome is lower in the sample, the upper limit of the Brier score is lower.

I calculated the Brier score's upper limit using incidence of mortality in the NSQIP sample. I have 30-day mortality incidence of 12.36%. This means our maximum Brier score would be: $(\text{Incidence-1})^2$, which gives an upper limit of 0.7680 for an uninformative model. For this study, I will calculate the Brier score for all patients in the sample and use the mean Brier score to assess the model's accuracy.

It is not necessary to know the point where we call the prediction positive or negative (cut-point) for computing the Brier score. We will need to know the cut-point when evaluating performance at different areas under the receiving operating characteristic (AUROC) curve discussed later. Calibration (model fit) of a predictive model refers to the process of creating the model and running it on a population that is not used to generate the predictions[53].

Cross-validation will be used to predict the performance of the model in an external dataset [52, 54] (Table 3). The algorithm can be assessed for its performance in an external dataset by folding the data repeatedly on a subset of the observations (multiple training sets) and applying it to the remaining observations multiple times (validation sets). K-fold cross-validation has been shown to be more efficient and a good substitute for training set and validation methods when sample size is limited [52, 54]. Cross-validation was chosen over creating a single training and validation set of patients due the rapidly changing surgical methods and procedures in our data. For example, if we reserve year 2011 for training and apply that to data from 2012–2013, the algorithm may fail due to newer procedures that reduce mortality or shifting populations (aging baby boomers in the cirrhotic cohort). Therefore, it is optimal to train the algorithm on multiple samples selected from different years then apply it to multiple validation samples from the same years (at least 10 k-folds).

The ability of the model to discriminate between those who survived or not will be evaluated using AUROC curves [52] (Table 3). Calibration will be assessed by the calibration-in-the-large and calibration-in-the-small measures. Both measures are decomposed from the Brier score. Calibration-in-the-large measures the discrepancy between the mean predicted probability and observed fraction of positive outcomes. Taking the square root of this value gives you the percentage difference between the prediction and the actual positive outcome. Calibration-in-the-small measures the difference between the predictive probability and actual outcome within the groups[52].

To determine the cut-point for the predicted probability of death that best agrees with the actual outcome, decision curve analysis will be performed[53] (Table 3). This analysis compares the net number of true positives gained by using the model plotted against a range of thresholds (probability of death cut-points). The cut-point that results in the greatest number of true positives will be chosen for the model.

Comparison of the Mortality Model to the MELD Score

For this analysis, we will use only patients that have laboratory values to generate both the MELD score and the mortality model. Those with and without laboratory values will be compared. If significant differences are observed, they will be listed as a limitation and addressed in future validation efforts.

I will use a chi-square statistic distributed with one degree of freedom to compare the AUROC for each model to the MELD score [53]. The MELD score will be calculated for each patient in the cohort using the Mayo Clinic MELD score algorithm:

MELD Score = (0.957 * ln(serum creatinine) + 0.378 * ln(serum bilirubin) + 1.120 * ln(PT INR) + 0.643) * 10 (if hemodialysis, value for creatinine is automatically set to 4.0)

The AUROC for the MELD score will be calculated and compared to the new predictive model [22, 53]. I expect the mortality model to have a greater AUROC than the MELD score when predicting post-surgical in-hospital mortality. I expect this difference to be significant with $p \leq 0.05$.

Sample Size and Power Determination

To estimate sample size and power, I use methods that are appropriate for diagnostic tests[55] (Table 4). A predictive model is similar to a diagnostic where sensitivity and specificity of the predictions determine the usefulness of the model. We want to optimize true positive and true negative predictions while minimizing false positives and false negatives. Jones et al. propose two separate calculations to determine sample size, one for optimizing sensitivity and one for optimizing specificity[55]. It is desirable to have the sample size to meet both requirements. We estimate power and sample size based on the assumption that our predictive models will not be less than 80% sensitive or 80% specific when compared to actual outcomes (mortality and readmission). The incidence of mortality was calculated. Overall 608/4,916 patients died regardless of death in or out of the hospital (12.36%).

Power Calculation Using a Sensitivity Threshold (Table 4) [55]

1^{st} calculation: $TP+FN=z^2 * (SN (1-SN)/w^2$

Using the 1^{st} calculation, we determine sample size based on the lowest possible acceptable sensitivity—here we use 80%. z^2 is set equal to 1.96 for a 95% confidence interval. We set alpha to 0.05 (5% threshold on each side of our sensitivity is allowable).

Where:
TP=true positives
FN=false negatives
Z=1.96 (95% confidence interval (CI))
SN=sensitivity
W=accuracy threshold (alpha) of 0.05

P=population mortality incidence (12.36%)

Then 1.96^2 *(0.80(1-0.80))/0.05^2=246 (sample size without taking into account the population incidence).

2^{nd} calculation : N(sN)=TP + FN/P

Using n=246 obtained in the first calculation and an incidence of 16.06% for deaths for cirrhotic patients obtained from our NSQIP data, *we obtain:* N(sN)=246/0.1236= 1,990 patients. The sample size is sufficient to power the study.

Power Calculation Using a Specificity Threshold [55]

1^{st} calculation: FP+TN=z^2 * (SP (1-SP)/w^2

Using the 1^{st} calculation, we determine sample size based on the lowest possible acceptable specificity—here we use 80%. z^2 is set equal to 1.96 for a 95% confidence interval. We set alpha to 0.05 (5% threshold on each side of our specificity is allowable).

2^{nd} calculation: N(sN)=FP + TN/(1-P)

Where:

TP=true positives
FN=false negatives
Z=1.96 (95% CI)
sN=sensitivity
W=accuracy threshold (alpha) of 0.05
P=population incidence (12.36%)

Then 1.96^2*(0.80(1-0.80))/0.05^2=246 (sample size based on specificity not accounting for mortality incidence in the sample).

Using 246 obtained from the first calculation and using an incidence of 12.36% for deaths in cirrhotic patients obtained from the NSQIP data. *We obtain*: N(sN)=246/(1-0.1236)=280 patients. Our sample size is sufficient to power the study. (Table 4)

Results

Bivariate Analyses (Table 5)

The mortality cohort included 4,916 patients, 608 patients of whom died prior to discharge and 4,308 of whom were discharged alive. In the unadjusted analyses, 15.1% of patients were between 71 and 80 years of age and had a significantly greater incidence of mortality compared to those discharged alive in this age category (19.1%; p=0.004). Patients aged 41–50 had significantly better lower post-surgical mortality (10.2%; p=0.02).

Patients with normal renal function (eGFR>90) had significantly lower mortality. Any degree of kidney dysfunction was associated with significantly greater mortality rates.

43.5% of cases were an emergent surgical procedure. Mortality was significantly greater than those surviving emergency surgical procedures (70.6% versus 39.3%; p<0.001).

Patients having a pre-operative comorbidity of confusion, alcohol use, chronic obstructive pulmonary disease, or chronic heart failure had greater incidence of in-hospital mortality. Underweight patients had a greater mortality incidence (7.1% versus 4.6%; p=0.01).

Functional status, ASA classification, MELD score, Childs-Pugh category, AST to platelet (APRI) score, and total length of stay were all associated with post-surgical mortality in the unadjusted analyses. All pre-operative laboratory values were significantly associated with mortality except for serum sodium. The total length of the hospital stay was associated with mortality with a mean stay of 14 days for those who died versus 9 days for those discharged (p<0.001).

Comparison of Cirrhotic Cohort with and without Laboratory Values (Table 6)

To assess missing pre-operative laboratory values, I compared patients with and without values in an unadjusted analysis. Patients either had all laboratory values necessary to generate the MELD score, mortality model, and readmission model or were missing values that would prohibit generation of any of the three models. Significant differences ($p<0.05$) indicates data may not be missing at random.

In the unadjusted analyses, white patients were more likely to have pre-operative laboratory values ($p=0.03$) while other races (non-white and non-black) were less likely to have pre-operative labs ($p<0.001$). Patients having an emergent procedure were less likely to have pre-operative labs ($p<0.001$).

Patients having a less invasive laparoscopic procedure were less likely to have pre-operative lab values available ($p<0.001$). Hernia repair patients were more likely to have labs ($p=0.005$). If functional status was totally dependent, fewer patients had pre-operative labs ($p<0.001$). Patients with ASA classification of mild disturbance ($p=0.02$) and severe disturbance ($p<0.001$) were more likely to have lab values while those with life threatening ($p=0.01$) or moribund ($p<0.001$) status were less likely.

These laboratory data were not missing completely at random or missing at random. Therefore, the external validity of the analyses may be compromised given the selection bias present in the NSQIP.

In-Hospital Mortality Model (Table 7)

The Poisson mortality model elucidated variables associated with mortality. As expected, advancing age is predictive of in-hospital mortality holding other variables constant. Patients between ages 51 and 60 had a 48.5% risk (IRR 1.485; $p=0.047$) of post-surgical death prior to

discharge compared to those <40 years (referent). As age increased, the risk increased to two times that of patients under 40 for the 71–80-year-olds (IRR 2.186; p<0.001) to >80 years (IRR 2.382; p<0.001).

In general, patients with poorer clinical/health status at baseline were at increased risk of death. Compared to patients that were fully independent, having a functional status of partially dependent carried a 36.5% (p=0.009) relative risk of death while those totally dependent had a relative risk of 38.4% (p=0.002) holding all other variables constant. Those with an ASA classification of life-threatening (IRR 4.030; p<0.001) or moribund (IRR 5.478; p<0.001) had a significantly greater risk of death compared to patients with mild disturbance (referent). The pre-operative eGFR, albumin, platelet count, total bilirubin, and prothrombin INR laboratory values were all predictive of post-surgical mortality. Patients having emergency surgery had a 69.8% (p<0.001) greater risk of death compared to scheduled surgery. Hernia repair surgery, whether laparoscopic or open, carried a 45.8% (p=0.025) reduction in the risk of death compared to other non-hernia open abdominal procedures. No other surgical category showed significant differences in risk compared to open abdominal surgeries.

Final Mortality Model (Table 8)

Mortality model: In-hospital mortality= $\beta 0 + \beta 1$ Age + $\beta 2$ Gender + $\beta 3$ BMI + $\beta 4$ Functional Status + $\beta 5$ ASA classification + $\beta 6$ Race + $\beta 7$ Race + $\beta 8$ eGFR + $\beta 9$ Albumin + $\beta 10$ Platelet count + $\beta 11$ WBC + $\beta 12$ Total bilirubin + $\beta 13$ PT INR + $\beta 14$ Sodium + $\beta 15$ Emergency surgery + $\beta 16$ Type of surgery + $\beta 17$ Comorbidities + $\beta 18$ Surgery category

All variables chosen a priori were kept in the model based on literature support. BMI was the only variable dropped from the model. BMI may be inaccurate because weight may reflect ascites rather than obesity in patients with decompensated cirrhosis.

Although hernia repair was the only surgery type to show a protective effect against mortality when compared to non-hernia open abdominal procedures, other surgery types were left in the model to control for complexity of the procedure and risk attributable to a specific procedure type. For example, cardiothoracic surgery may carry greater risk of mortality than a laparoscopic procedure. Diabetes, chronic heart failure, and COPD were not significant in the Poisson model but included in the final model to control for comorbidities known to carry risk of mortality.

The final mortality model employed a logistic regression reporting ORs. Predictions were obtained for the model and interpreted as the probability of death prior to discharge after the index surgery. The final model included 1,945 patients (Table 8). Advancing age was predictive of mortality with odds increasing from 2.041 ($p=0.029$) in the 51–60 group to 4.385 ($p<0.001$) in patients >80 years of age compared to the <40 referent group, holding other factors constant. Functional status, albumin, platelet count, total bilirubin, prothrombin INR, emergent surgery, and hernia repair were all significant predictors of mortality ($p<0.05$).

Validation of the Mortality Model (Table 9)

To assess calibration, calibration-in-the-large, calibration-in-the-small, and Hosmer-Lemeshow goodness-of fit tests are reported. Calibration-in-the large is a property of the entire mortality cohort sample. The mortality model had a calibration-in-the-large value of 0.0004. By taking the square root of this value, the model shows a 2% difference between the mean predicted probability of death and the actual mean of the binary outcome. Calibration-in-the-

small measures the error within the groups (died vs. discharged alive). The mortality model had a value of 0.0014 that translates to 3.7% difference in the predicted probabilities versus the actual outcome within groups. Finally the Hosmer-Lemeshow goodness-of-fit test indicates if the model is correctly specified. In this case, the p-value for this test was 0.345. This is not significant at a threshold of $p<0.05$ indicating the model is correctly specified.

Discrimination was assessed using AUROC curves, sensitivity, specificity, positive predictive value, negative predictive value, and percent correctly classified at a chosen cut-point. The cut-point that maximized the sensitivity of the model and kept specificity above 60% was 0.2150 (also see the decision curve analysis). This model achieved an AUROC of 0.8282. The sensitivity was 81.1%, specificity 73.6%, positive predictive value 47.6%, and negative predictive value 92.9%. We achieved our goal of at least 0.80 for the AUROC. The sensitivity exceeded our goal of at least 80% but the specificity of the model did not reach the goal of 80%.

We used the k-fold (k=10) and leave-one-out methods to cross-validate the model. Both the k-fold and leave-one-out methods exceeded our AUROC threshold of 0.80 (k-fold: 0.8402; leave-one-out: 0.8420). The Brier score, which considers both the calibration and discriminatory ability of the model, was 0.1269, which is well below the threshold for a non-informative model.

Decision Curve Analyses (Figure 3)

The decision curve analysis indicates predicted probabilities where the model is useful as a decision tool versus using no tool (greatest net benefit). Figure 3 indicates the model may reduce post-surgical deaths (net benefit) if used between 10% and 75% predicted probability of mortality. Above 75%, predictions about whether a patient should have surgery are not useful given the high probability of post-surgical mortality without using the model. For probabilities below 10% the model is not useful given the low probability of death. Below 10% and above

75%, the model contributes no information beyond the surgeon's assessment of risk. The point on the decision curve that maximizes the net benefit of the model to identity true positives is 0.2150 and is where the maximum net benefit is gained.

Comparison of the Model to the MELD Score (Figure 5)

The AUROC curve for the mortality model (0.8422) was significantly ($p<0.001$) better than the MELD (0.7405). Across the range of cut-points, the mortality model is better at predicting post-surgical deaths than the MELD score. This means the NSQIP mortality model is better at discriminating between those who died and those who were discharged alive versus the MELD score. If using the NSQIP mortality model, a clinician would be better able to classify patients by their risk of death post-operatively.

Discussion

The decision to take a patient with decompensated cirrhosis to surgery is difficult for surgeons, given the patient's risk of death. Currently, surgeons often look to the MELD score, even though it was not developed to predict death. We developed and validated the NSQIP Mortality Model, which is significantly better ($p<0.001$) than the MELD score when predicting post-surgical in-hospital mortality in patients with decompensated cirrhosis. This is, perhaps, not surprising because the variables we included were clinically sensible—for example, age, pre-operative laboratory values, etc.

We found that surgery type carries different levels of risk. For example, cardiothoracic surgery carries more risk than a laparoscopic hernia repair. Therefore we kept surgery type in the model to control for varying levels of risk. Not surprisingly, hernia repair was performed most often by a laparoscopic procedure and offered a protective effect. Although any non-hernia/non-cholecystectomy laparoscopic procedure was not significant in the model, the estimate does not

show a protective effect for the less invasive surgery. This may be due to selection bias in the sample. Sicker patients are selected for the less invasive laparoscopic procedures and thus carry a greater risk of mortality.

In terms of pre-operative laboratory values, albumin was strongly associated with mortality. Cirrhotic patients tend to struggle with decreasing albumin levels. The model showed that for every one point increase in albumin prior to surgery, the risk of mortality was reduced by 35% on average. This may suggest that getting patient's albumin levels up prior to surgery could be beneficial in reducing risk of mortality. Increases in total bilirubin and the PT INR were also strongly associated with mortality. For each point increase in bilirubin prior to surgery, the model showed a 22.9% increase in risk of death. For each point increase in PT INR, there was an 87.3% increase in the risk of post-surgical mortality. It is well known that liver dysfunction contributes to increased bilirubin and prothrombin, and our data suggest that close attention should be paid to these pre-operative laboratory values prior to taking patients to surgery.

The model performed significantly better than the MELD score. In the decision curve analysis (Figure 3), the AUROC curve for the NSQIP Mortality Model is greater than for the MELD score. This difference represents the net benefit of using the NSIP Mortality Model to decide to take a patient to surgery across a range of predicted probabilities. Moreover, the NSQIP Mortality Model showed net benefit between across a broader range of predicted probabilities (10%–76%) than the MELD Score (22%–39%). Having a model that can predict mortality risk across the middle-range of threshold probabilities is of greater value to the surgeon. It is in this middle range of risk that decisions to proceed to surgery become difficult and the model offers the most net benefit.

This study has several limitations. First, there is selection bias. Given that pre-operative serum sodium was not predictive of mortality supports selection of healthier patients for surgery. In patients with liver cirrhosis, sodium has been found to be associated with mortality in multiple studies[56, 57]. Surgeons often avoid surgery on patients who have a low or elevated sodium value due to its association with mortality. They assume these patients are at great risk of death post-operatively and choose not to proceed. These patients will not be included in the NSQIP database.

Second, laboratory values were not missing completely at random, which may have introduced bias into our models. One concern is that patients undergoing a laparoscopic procedure had fewer laboratory values necessary to generate the models and the MELD score than those undergoing more invasive (non-lap) procedures. Patients undergoing hernia repair had laboratory values available more often than those having non-hernia repair procedures. Further research in external samples is required to address the potential bias introduced by missing values. Imputation or selection models could have been undertaken but given the wide variation in laboratory values in decompensated cirrhotic patients, they were not chosen as options due to the fear of introducing more bias.

On the one hand, the estimates we obtain may be biased by the inability to control for the two types of selection bias present in the data[58]. On the other hand, the NSQIP is designed specifically to capture pre- and post-surgical risk or mortality and readmission[44]. There is no better database to stratify patients undergoing surgery by risk of mortality and readmission. It is possible to evaluate risk factors prior to, during and after surgery.

Next Steps

The mortality model performed well in this sample. However, more validation is necessary prior to using the model to make informed decisions regarding surgery. Once validation is completed in a prospective, real-world study, a risk score can be provided through the EMR. This score may appear as: "72% risk of mortality predicted: In studies of patients with decompensated liver cirrhosis and predicted probability of mortality of >60%, surgery is not recommended for X-type surgery." This will inform decisions made between the hepatologist, surgeon, and patient.

Future studies will evaluate the mortality model in a real-world setting. I intend to implement the model in an electronic medical record (EMR) system for external validation. Because data in the model are collected in real time, the predicted probabilities can be classified based on patients' outcomes. For example, if the model is allowed to run in an EMR for one year across several hospitals, enough data could be collected to group the predictions into clear go or no-go categories. A clear cut-point may emerge to inform the surgical decision. If a single cut-point does not emerge, risk categories (Low/Medium/High) may be established to help the surgeon and patient decide if surgery is advisable. Validation of predictive models is a long and dynamic process. Until it is clearly proven the model works in multiple settings and across different surgery types, it cannot be implemented into practice.

CHAPTER 4. THE NSQIP READMISSION MODEL

Introduction

Decompensation increases the risk of unplanned readmissions [10-12]. Compared to patients without cirrhosis, those with cirrhosis had among the highest odds of unplanned readmissions[4, 11, 12, 59]. Patients with cirrhosis who undergo surgery are frequently readmitted within 30 days[12]. This is concerning for both the patients and the healthcare system. The ability to predict the risk of 30-day readmission would potentially reduce unplanned readmissions by allocating resources for support patients at the greatest risk of readmission[18]. The urgency to predict readmissions is fueled by financial penalties imposed on hospitals by the Centers for Medicare and Medicaid Services (CMS) for Medicare patients with 30-day readmissions [16, 17]. As the cirrhotic population ages, more will have Medicare as their primary payer and expose hospitals to the risk for significant financial penalties [17]. Other payers will likely adopt CMS's reimbursement strategies for 30-day readmissions. As CMS begins its bundled payments for the care improvement initiative, it will be critical to a hospital's financial health to know the risk of a prolonged stay or complication prior to the service provided (e.g., surgical procedure).

Currently, no validated model exists to predict readmission among patients with cirrhosis who undergo surgery. This study aimed to develop and validate a readmission predictive model and then compare its ability to predict readmission to the MELD score in patients undergoing surgery.

Methods

Data Source and Sample Selection

This study used the American College of Surgeons National Quality Improvement Program data from 2011–2013[44]. NSQIP is a national voluntary program to help hospitals reduce surgical morbidity and mortality. The program offers three levels of participation that vary by labor intensity to allow smaller hospitals with limited resources to participate. To date, more than 300 hospitals nationwide participate in the NSQIP[44].

Data are collected through a combination of electronic medical records and manual chart abstraction at each site. More than 150 variables are collected on surgical cases. Each year, hospitals rotate through 46 eight-day cycles of data collection to capture a random sampling of surgical cases. Included are patients having a major surgery (defined by CPT code) during one of the eight-day collection cycles. Both emergent and non-emergent surgeries are captured. Trauma and transplant surgeries are excluded[44, 45].

Data are collected at the patient level. Pre-, intra-, and post-operative variables are collected for each patient. If a patient died prior to discharge, he or she is captured in the discharge status variable with an "expired" status[44]. Up to five unplanned readmissions are captured within 30 days from the index surgery in readmission outcome variable[44].

Identifying Patients with Liver Cirrhosis

I selected patients with liver disease, defined as having an esophageal hemorrhage or ascites as defined by the ACS NSQIP[44]. The ACS NSQIP definition for these variables requires a history of liver disease on record concurrent with an esophageal varices or ascites diagnosis[44, 45]. For these patients, I will use previously validated laboratory values that define liver cirrhosis: platelet value of $\leq 140{,}000/mm^3$ and albumin <3.5 g/dL and prothrombin

INR>1.5 [45-47]. Finally, any patient with disseminated cancer, an outpatient surgery, or a do not resuscitate (DNR) order on file, as defined by a variable in the NSQIP, was removed. This resulted in a final cohort of patients with liver cirrhosis (n=4,916).

Readmission Model Cohort

Aims 1b, 2b, and 3b will be restricted to the 4,308 patients who were discharged alive after the index surgery (Figure 2). Of those, I determined who was readmitted versus those that were not readmitted within 30 days from the index surgery. The predictive model will be used for patients that are at risk of having an unplanned readmission within 30 days after the index surgery and subsequent discharge. Our final cohort for analysis in aims 1b, 2b and 3b will include patients with laboratory values to generate the MELD score and readmission model (readmitted: n=542) and (n=2,970 discharged alive but not readmitted within 30 days). 153 patients were excluded due to missing laboratory values.

Bivariate Analyses

Patients with and without a readmission were compared using ANOVA, chi-square, and non-parametric test for trends in ordered variables (Table 5). Eligible patients with and without pre-operative laboratory values were compared by ANOVA or chi-square analysis to determine if their baseline characteristics differ significantly ($p<0.05$) (Table 6). Those with and without labs by surgery type were evaluated to further understand any bias that may be introduced into the model. It may be the less invasive surgery types had fewer labs ordered by the surgeon due to the perception of lower risk of mortality. Any significant differences are noted in limitations. It is not possible to control for selection bias in this sample when we do not have physician- or hospital-level factors necessary to generate an appropriate selection model.

Development of the Multivariable Readmission Model

We used a Poisson model with a binary outcome (readmission within 30 days of discharge from the index surgery) to determine predictors of readmission [48, 49]. This model is has two advantages over logistic regression [48, 49]. First, it computes incident risk ratios (IRR) as relative risks, which facilitates interpretation[49, 50]. Second, when cross-sectional data have frequent outcomes of interest, odds ratios (OR) may overestimate the prevalence in the sample[50]. Poisson is appropriate for a binary outcome model when relative risk is desired and events of interest are frequent. The Poisson model requires Huber-White standard error adjustment (robust standard errors) to avoid under dispersion[49, 50]. If a variable was not significant, I tested for joint significance ($p<0.05$). If there was joint significance with other variables in the model, I included these measures in the final model for validation. Ultimately, literature support guided final selection of variables in the model regardless of significance.

Validation of the Readmission Model

After determining the predictors of readmission using a Poisson model, a logistic regression maximum likelihood model with a binary outcome was used to obtain in-range predictions. Predictions for Poisson models (used for development) are often not appropriate for validation methods discussed earlier and shown in Table 1 [51]. I will report ORs for the readmission estimates. Predictions from the model will be compared to those that were readmitted after the index surgery. I am interested in the predictions from the model and how they compare to the actual outcome. I will obtain the predicted probability of readmission for each patient in the sample. The actual outcome is binary with 1 equal to death. Having a binary outcome makes it difficult to compare predictions that are continuous between 0 and 1. It will be necessary to perform a cut-point (i.e., threshold) analysis to determine where along the range of

predicted probabilities the sensitivity and specificity are optimized (e.g., where do you call a "positive" result=readmission versus a negative result=no readmission)[52, 53].

Assessment of the Readmission Model Performance

Overall model performance will be assessed with the Brier score [52, 53] (Table 3). Brier scores measure the distance between the actual outcome and the predicted outcome[53]. Brier scores are similar to the Pearson's R^2 statistic. It is a quadratic scoring equation that uses the squared differences between the actual binary outcome (died/survive) and the predictions (p). [Brier score=$(Y-p)^2$] [52] The lower the Brier score, the better the predictions are calibrated to the actual outcome; for a perfect model with optimal fit, the Brier score equals zero. The Brier score is affected by the incidence of the outcome of interest in the population. When the incidence of the outcome is lower in the sample, the upper limit of the Brier score is lower.

It is not necessary to know the point where we call the prediction positive or negative (cut-point) for computing the Brier score. We will need to know the cut-point when evaluating performance at different areas under the receiving operating characteristic (AUROC) curve discussed later. Calibration (model fit) of a predictive model refers to the process of creating the model and running it on a population that is not used to generate the predictions[53].

Cross-validation was used to predict the performance of the model in an external dataset [52, 54] (Table 3). The algorithm can be assessed for its performance in an external dataset by folding the data repeatedly on a subset of the observations (multiple training sets) and applying it to the remaining observations multiple times (validation sets). K-fold cross-validation has been shown to be more efficient and a good substitute for training set and validation methods when sample size is limited [52, 54]. Cross-validation was chosen over creating a single training and validation set of patients due the rapidly changing surgical methods and procedures in our data.

For example, if we reserve year 2011 for training and apply that to data from 2012–2013, the algorithm may fail due to newer procedures that reduce mortality or shifting populations (aging baby boomers in the cirrhotic cohort). Therefore, it is optimal to train the algorithm on multiple samples selected from different years then apply it to multiple validation samples from the same years (at least 10 k-folds).

The ability of the model to discriminate between those who were readmitted or not will be evaluated using AUROC curves [52] (Table 3). Calibration will be assessed by the calibration-in-the large and calibration-in-the small measures. Both measures are decomposed from the Brier score. Calibration-in-the-large measures the discrepancy between the mean predicted probability and observed fraction of positive outcomes. Taking the square root of this value gives you the percentage difference between the prediction and the actual positive outcome. Calibration-in-the-small measures the difference between the predictive probability and actual outcome within the groups[52].

To determine the cut-point for the predicted probability of readmission that best agrees with the actual outcome, decision curve analysis will be performed[53] (Table 3). This analysis compares the net number of true positives gained by using the model plotted against a range of thresholds (probability of death cut-points). The cut-point that results in the greatest number of true positives will be chosen for the model.

Comparison of the Readmission Model to the MELD Score

For this analysis, we will use only patients that have laboratory values to generate both the MELD score and the readmission model. Those with and without laboratory values will be compared. If significant differences are observed, they will be listed as a limitation and addressed in future validation efforts.

I will use a chi-square statistic distributed with one degree of freedom to compare the AUROC for the model to the MELD score [53]. The MELD score will be calculated for each patient in the cohort using the Mayo Clinic MELD score algorithm:

MELD Score = (0.957 * ln(serum creatinine) + 0.378 * ln(serum bilirubin) + 1.120 * ln(PT INR) + 0.643) * 10 (if hemodialysis, value for creatinine is automatically set to 4.0)

The AUROC for the MELD score will be calculated and compared to the new predictive model [22, 53]. I expect the mortality model to have a greater AUROC than the MELD score when predicting post-surgical in-hospital mortality. I expect this difference to be significant with $p \leq 0.05$.

Sample Size and Power Calculation

To estimate sample size and power, I use methods that are appropriate for diagnostic tests[55] (Table 4). A predictive model is similar to a diagnostic where sensitivity and specificity of the predictions determine the usefulness of the model. We want to optimize true positive and true negative predictions while minimizing false positives and false negatives. Jones et al. propose two separate calculations to determine sample size, one for optimizing sensitivity and one for optimizing specificity. It is desirable to have the sample size to meet both requirements. We estimate power and sample size based on the assumption that our predictive models will not be less than 80% sensitive or 80% specific when compared to actual outcomes (mortality and readmission).

Calculation of the incidence of readmission within 30 days: Overall 566/4,308 (13.14%) had any unplanned readmission. Using Jones et al. equations with a lower limit for sensitivity and specificity of 80%, we can estimate the sample size for the readmission model. We found

readmissions in the cirrhotic cohort were 13.14%. Using the incidence of readmission we find the following:

Sensitivity: N(sN)=246/0.1314=**1,872** total patients in sample
Specificity: N(sN)=246/(1-0.1314)=**283** total patients
The sample size is sufficient to power the study. (Table 4)

Results

Bivariate Analyses (Table 5)

Patients discharged with a 30-day readmission were compared to those without on multiple variables. I found advanced age was not associated with readmission in patients with decompensated cirrhosis. Insulin-dependent diabetes was associated with readmission in ($p=0.001$). Patients classified into the ASA class moribund or life-threatening had significantly more readmissions.

The pre-operative laboratory values for albumin, sodium, total bilirubin, platelets, PT INR, and white blood cell count showed significant differences between those with an unplanned readmission versus those without. Patients with a MELD score of less than or equal to 9 (less severe disease) had significantly more readmissions ($p=0.002$). Those with MELD scores between 10 and 19 also had significantly fewer readmissions ($p=<0.001$).

Comparison of Cirrhotic Cohort with and without Laboratory Values (Table 6)

To assess missing pre-operative laboratory values, I compared patients with and without values in an unadjusted analysis. Patients either had all laboratory values necessary to generate the MELD score, mortality model, and readmission model or were missing values that would prohibit generation of any of the three models. Significant differences ($p<0.05$) indicates data may not be missing at random.

In the unadjusted analyses, white patients were more likely to have pre-operative laboratory values ($p=0.03$) while other races (non-white and non-black) were less likely to have pre-operative labs ($p<0.001$). Patients having an emergent procedure were less likely to have pre-operative labs ($p<0.001$).

Patients having a less invasive laparoscopic procedure were less likely to have pre-operative lab values available ($p<0.001$). Hernia repair patients were more likely to have labs ($p=0.005$). If functional status was totally dependent, fewer patients had pre-operative labs ($p<0.001$). Patients with ASA classification of mild disturbance ($p=0.02$) and severe disturbance ($p<0.001$) were more likely to have lab values while those with life-threatening ($p=0.01$) or moribund ($p<0.001$) status were less likely.

These laboratory data were not missing completely at random or missing at random. Therefore, the external validity of the analyses may be compromised given the selection bias present in the NSQIP.

NSQIP Unplanned 30-Day Readmission Model (Table 10)

Age, gender, body mass index, functional status, race, surgery type, discharge destination, and emergent surgery were not significant predictors of readmission in the model. However, using a p-value significance threshold of less than 0.05, gender and discharge to a rehabilitation facility were borderline significant with p-values of 0.05.

Having an ASA classification of life-threatening increased the risk of readmission by 62.8% ($p=0.009$) compared to those classified as having a mild disturbance while holding other variables constant. For every 1 point increase in albumin, there is a 14.7% ($p=0.049$) reduction in the risk of readmission. For every increase in the PT INR, there is a 48.0% ($p=0.003$) increase in

the risk of readmission. Patients who are insulin dependent have a 61.7% (p=0.001) increase in the risk of being readmitted compared to non-diabetics.

Certain post-operative complications were found to increase the risk of readmission. Sepsis dramatically increased the risk of readmission by 96.6% (p<0.001). Acute or progressive renal failure increased readmission risk by 57.0% (p=0.032) while a urinary tract infection increased risk more than two-fold (RR 2.25; p=0.001). If a patient had an unplanned intubation, the risk of readmission increased by 51.8% (p=0.038).

Time on the operating table contributes to the risk of readmission. For every additional hour spent on the table, the risk of readmission was increased by 12.0% (p=0.002).

Final NSQIP Readmission Model (Table 11)

Readmission model: Readmission (within 30 days of index surgery)= β0 + β1 Age + β2 Gender + β3 BMI + β4 Functional Status + β5 ASA classification + β6 Race + β7 Race + β8 eGFR + β9 Albumin + β10 Platelet count + β11 WBC + β12 Total bilirubin + β13 PT INR + β14 Sodium + β15 Emergency surgery + β16 Type of surgery + β17 Comorbidities + β18 Surgery category + β19 Discharge destination + β20 Pneumonia + β21 Sepsis + β22 Time on table + β23 Time under anesthesia + β24 Acute/Progressive Renal Failure + β25 Urinary Tract Infection + β26 Ventilator Wean Time >48 hours + β27 Total Length of hospital stay + constant

All variables, regardless of significance in the Poisson model were kept in the final logit readmission model. Readmissions are notoriously difficult to predict. The Poisson model was used to find significant predictors of readmission. However, when using only significant predictors in the logit model, important factors that need to be controlled for (in theory) are lost

and may cause bias in predictors left in the model (joint significance). It is best practice to keep all variables in the readmission model, regardless of significance, because theory and literature dictate they contribute to readmission risk. For example, age is not significant in the Poisson model or logit model but there is support in the literature that advancing age increases the risk of readmission [17]. Omission of important variables may lead to omitted variable bias. However, including too many variables may result in over-fitting of the model. Subsequent tests were conducted to guard against over-fitting (see validation section). The final decision was to keep all variables from the Poisson model in the final readmission model. This was driven by theory and literature support.

Being male was found to decrease the odds of readmission compared to females (OR 0.705; p=0.041). Several laboratory values were predictive of readmission. As albumin increases by one point, the odds of readmission also decreased (OR 0.793; p=0.037). A one-point increase in PT INR increases the odds of readmission more than two-fold (OR 1.734; p=0.009). Patients who are insulin dependent have a two-fold increase in odds of readmission (OR 1.980; p=0.001).

Certain post-operative complications remained significant in the final model when predicting readmission. Post-operative sepsis increased the odds of readmission two-fold as did acute renal failure (OR 2.641; p<0.001 and OR 2.028; p=0.033). Urinary tract infection remained a strong predictor of readmission (OR 3.248; p=0.001). Patients that were discharged to a rehabilitation facility had lower odds of readmission compared to those discharged to a facility that was not home, acute care, or rehabilitation facilities (OR 0.452; p=0.041).

Validation of the Readmission Model (Table 12)

For the readmission model, calibration and discrimination were assessed. The calibration-in-the-large was zero for this model. This indicates that for the entire readmission cohort, there is

no difference between the mean of all predictions and actual outcome. However, calibration-in-the-small was 0.007 or 2.65% difference in the predicted mean of those readmitted versus the actual mean of those readmitted (within-group calibration). This is a well-calibrated model. The Hosmer-Lemeshow goodness-of-fit test was not significant (p=0.310), indicating the model is correctly specified.

The ability of the model to discriminate between those readmitted and those not readmitted was assessed. The AUROC was 0.7541 with a sensitivity of 76.6%, specificity of 60.9%, positive predictive value of 27.7%, and negative predictive value of 93.0% at a cut-point of 0.1425. Neither the sensitivity nor specificity of the readmission model reached our goal of 80%.

Although the model did not reach our 80% sensitivity and specificity goals, the out-of-sample performance indicates a larger sample may improve performance. Cross-validation using the 10-fold method gave an AUROC of 0.7602. This is greater than the AUROC from running the full model and an indicator that performance in another sample or larger sample could improve performance. The leave-one-out method had an AUROC equal to that of the full model.

The Brier score was 0.1186. The maximum Brier score, accounting for 13.1% incidence of readmission, was 0.7540 for a non-informative model. Our Brier score was quite low, indicating good discrimination and calibration. Overall the model was able to classify 63.5% of the observations correctly.

Decision Curve Analysis (Figure 4)

The decision curve analysis showed the readmission model had a positive net benefit between threshold probabilities of 19% and 70%. This means that for those at very low risk of readmission (<19%) or those at very high risk (>70%) using the model to prevent readmissions is

of no benefit. A model that can inform decisions in the middle range for risk of readmission is most useful because this will be where it is difficult for a hospital to decide how to tier discharge planning to prevent readmission.

The point on the decision curve that maximizes true positives for readmission is 0.1425. This is where the maximum net benefit is gained by using the model.

Comparison of the Readmission Model to the MELD Score (Figure 6)

The readmission model was compared to the MELD score. Although the MELD score is not an optimal comparator and is not intended to predict readmissions, no other readmission model exists specifically for decompensated patients with cirrhosis undergoing surgery. The readmission model AUROC (0.7541) exceeded the MELD score AUROC (0.5663) when predicting 30-day readmission ($p<0.001$).

Discussion

Readmissions have been notoriously difficult to predict [60]. The combination of in-hospital factors and patient-level factors after discharge are required to capture accurate risk of readmission. In our dataset, we had in-hospital factors. It is evident in the final readmission model there are few predictors that are significant. Prediction of in-hospital mortality relies on pre-, intra-, and post-operative factors that are captured in our dataset. When a patient is discharged the ability to capture important factors such as medication compliance and wound care is lost. Therefore, many readmission models, including this model, suffer from omitted variable bias. However, this model is a good start to finding predictors of readmission in decompensated cirrhotic patients. Further research will target data sources that capture out-of-hospital factors.

The current model had some surprising results. First, being male had a protective effect on readmission compared to females. There was a 29% decrease in the risk of readmission if male holding other factors constant (Table 10). This may be an artifact given over 50% of the patients in our sample are male (56.8%) or it could be that women are traditionally caretakers and have fewer resources for taking care of themselves. When they return home from the hospital, they may have less time to recover and address their own medical needs or may not have the support of a caretaker. Future studies would be warranted to investigate this hypothesis by capturing information on the home environment after discharge.

It was not surprising that having a life-threatening ASA classification was associated with readmission compared to those with mild disturbance. These patients are sicker and have greater baseline risk of complications after surgery. Patients with a moribund ASA classification were not associated with readmission. This may be due to the fact they are more likely to die after surgery given their grave health state.

The type of surgery, emergent surgery, and length of hospital stay were not predictive of readmission. This is surprising given that different surgery types carry variable levels of risk. It seems readmission may be more likely if undergoing a cardiothorasic procedure versus a hernia repair. Length of hospital stay has been associated with readmission in previous studies but our model did not show a significant association. It could be patients are hospitalized long enough to stabilize their health prior to discharge thus reducing the risk of readmission. In fact, discharge to a rehabilitation facility was associated with a 54.8% reduction in the risk of readmission compared to discharge to another type of facility that was not home or acute care (Table 10). Stabilizing patients with decompensated cirrhosis may be key to mitigating readmission. This

may include fluid management to prevent encephalopathy, protein management to stabilize albumin levels, and control of bleeding to avoid variceal hemorrhage.

Patients with insulin-dependent diabetes are almost two times as likely to be readmitted compared to those without diabetes. It is well established that liver dysfunction can make diabetes management difficult due to glycogen storage in the liver being impaired[61]. Readmission may be due to complications of slow wound healing or infection given both are frequent complications in diabetics.

There were no surprises in the post-operative predictors of readmission. Patients are at greater risk of infections even without surgery. Post-surgical urinary tract infections (UTI) increased the risk of readmission three-fold compared to those without a UTI. Sepsis increased readmission risk almost three-fold. Acute renal failure or progressive renal failure after surgery increased risk two-fold. Both sepsis and acute renal failure are common in decompensated cirrhotic patients even without surgery. Unplanned intubation after surgery was associated with readmission and may be a proxy for capturing post-surgical complications not available in the NSQIP database such as metabolic acidosis, encephalopathy, or fluid management problems.

Surprisingly, several pre-operative laboratory values were associated with readmission. Albumin, platelet count, and PT INR were all associated with readmission. As seen in the mortality model, decreasing albumin in decompensated cirrhotic patients is a clear predictor of health. As albumin falls, fluid management becomes more difficult, possibly leading to readmission. Platelet count and PT INR measures are associated with bleeding. Patients with cirrhosis have a difficult time with bleeding and surgery further complicates this delicate balance, possibly resulting in readmission.

This study has several limitations evident in the results of the analyses. First, there is clear selection bias in the sample. Finding that pre-operative serum sodium was not predictive of mortality supports selection of healthier patients for surgery. In patients with liver cirrhosis, sodium has been found to be associated with mortality in multiple studies[56, 57].

Second, for the readmission model, key factors may not be captured that are associated with readmission. This may introduce omitted variable bias in our model and decrease accuracy of the predictions (sensitivity/specificity/AUROC). Three such factors are socioeconomic status, home lifestyle factors, and insurance status of patients in our sample. Further studies are necessary to determine what factors are predictive of readmission and determine interventions to prevent readmission.

Finally, laboratory values were not missing completely at random or missing at random. This may have introduced bias into our models given we used complete case analysis. Further research in external samples is required to address the potential bias introduced by missing values. Imputation or selection models could have been undertaken but given the wide variation in laboratory values and missing information on hospitals or physicians in the dataset, they were not viable options.

Overall the readmission model performed reasonably well but did not reach the AUROC, sensitivity, and specificity goals of this study. Post-discharge patient-level data is required to accurately predict readmissions. Future studies will focus on collection of these factors to improve model performance.

Next Steps

This model will require substantial modification to improve performance moving forward. In fact, more development is required using post-discharge variables. Once the model is

fully developed and validated, it can be used to identify patients at greater risk of readmission and allocate at-home resources to these patients. One example may include scheduling more intensive home health visits for patients at greater risk of readmission. Prospective studies can assist in design of discharge plans tailored to the level of risk that successfully prevent unplanned readmissions.

CHAPTER 5. DISCUSSION AND FUTURE DIRECTIONS

Discussion

With the implementation of financial penalties for in-hospital mortality and 30-day readmissions by CMS, it is increasingly important to maximize value by reducing these costly events. To deliver high-quality care that reduces these penalties, it is necessary to predict which patients require more intensive in-hospital and post-discharge management. Predictive models are being used in health care to risk-stratify patients so that allocation of costly resources can be delivered to the right patients at the right time, resulting in better outcomes. This study develops and validates two predictive models that can help inform decisions for a high-risk population—patients with decompensated liver cirrhosis undergoing surgical procedures. This study is the first to deliver such models for decompensated cirrhotic patients.

Patients with decompensated cirrhosis are at greater risk of mortality due to sepsis, hepatic encephalopathy, and variceal hemorrhage[62]. Although sepsis is not unique to this population, hepatic encephalopathy and variceal hemorrhage are primarily seen in patients with cirrhosis. These comorbid conditions not only increase the risk of mortality but also the risk of readmission. Thus, specific models to predict mortality and readmissions in this population are necessary to capture these unique risk factors.

The National Surgical Quality Improvement Program (NSQIP) mortality model performance met the study goals and shows great promise for implementation in a real-world setting. Using secondary data to develop and internally validate models is a practical first step prior to its external validation and widespread implementation to make clinical decisions. In

addition to using the model to predict mortality, independent factors in the model shed light on potential opportunities to improve patient care. For example, the mortality model found hernia repair to have a protective effect. This may suggest that hernia repair is critical to reduce the decompensated patient's risk of premature death. Further study is warranted on the benefits of hernia repair in this population. Additionally, the model found pre-operative albumin, total bilirubin, and Prothrombin International Normalized Ratio (PT INR) levels were strongly associated with mortality. For every one-point increase in the patient's pre-operative albumin, mortality is reduced by 34% on average. Getting the albumin levels up in the cirrhotic patient prior to surgery may be critical in reducing the risk of post-operative mortality. The predictions from the mortality model were found to be informative in patients with a predicted probability of death between 10% and 76%. This is exactly where you want the model to be most useful. It is in patients that are in the middle range, not those with high or low risk of mortality, that the model is most needed to guide surgical decisions. When it is difficult for the surgeon, patient, and heptaologist to make the decision to proceed to surgery, the model is most informative and may better inform those decisions. Given the current paradigm is to use the MELD score to drive decisions in cirrhotic patients, this model is a much better option because it was significantly better than the MELD score at predicting post-surgical mortality.

Although the NSQIP readmission model was reasonable, it did not meet the study performance goals. The inability of the readmission model to meet the study goals may be due inadequate sample size and/or the lack of information on patients once they leave the hospital. More often, consumer data is used to better understand patient behavior after discharge from a hospital. This model may benefit from addition of these factors in future development and validation efforts. In the current model, there were some useful findings. For example, insulin-

dependent diabetes was a risk factor for readmission in decompensated cirrhotic patients. Better management of diabetes after surgery may be required for these patients. Implementing more intensive home health services for insulin-dependent diabetics may be one strategy for reducing readmission. If serum pre-operative albumin is increased by one point prior to surgery, there is a 21% decrease in the risk of post-surgical readmission. The albumin levels can be managed pre-operatively to reduce the risk of readmission in these patients. Predictions from the full model can guide the intensity of discharge planning. The model did outperform the Model for End Stage Liver Disease (MELD) score and may be a better way to assess risk of readmission in this population.

Both models are a strong starting point for improving the quality of care patients with cirrhosis receive. By implementing predictive models such as these in the electronic medical record (EMR) clinicians will be better able to judge risk of mortality or readmission. The current paradigm is for a hepatologist and surgeon to discuss a patient's risk of mortality using the MELD score, clinical comorbidities, and many other factors (patient compliance, insurance status). MELD scores are inadequate for predicting risk of death or readmission across a broad range of patients with cirrhosis. These NSQIP models give a concrete risk score to the clinician that can better inform the decision-making process.

Hospital discharge teams would benefit from knowing the risk of readmission in this special population of decompensated cirrhotic patients. Fluid management and keeping albumin levels in the up is key in patients with decompensated cirrhosis, especially after surgery. This model identifies that for those at greater risk of readmission and home health services, remote weight monitoring, albumin supplementation, and giving a 30-day supply of medication at discharge may all be strategies that can be implemented based on the readmission risk score. To

implement these interventions for everyone with decompensated cirrhosis, without knowing the risk of readmission, would be too costly.

Overall, both the mortality and readmission models move clinical care for patients with cirrhosis forward. First, prior to this study, no model existed to predict mortality or readmissions in decompensated cirrhotic patients. Second, the MELD score is overused for decision making when caring for patients with cirrhosis. It has multiple limitations that these models overcome. Last, targeted interventions can be investigated using the risk score to stratify patients into specific clinical care plans.

After further validation of each model using real-world clinical data, it is possible to implement the models in a widely used electronic medical record system such as Epic™. Epic™ Web is an Epic™ data warehouse for user-developed programs and registries that can be implemented by sites. It is possible to post models to the Epic™ Web after further validation. This is beyond the scope of this dissertation.

The models may be implemented as smartphone applications in addition to Epic™ programs. This would allow non-Epic™ users to access the programs easily in the clinical environment regardless of the electronic medical record system in use at their facility.

Both mortality and readmission models will require further study prior to using in the real-world setting. My first priority is to work with the Carolina Data Warehouse at the University of North Carolina to implement the mortality model for further validation. Being able to get a risk score for patients undergoing any type of surgery with both compensated and decompensated liver cirrhosis will broaden the scope of the model. Following patients outcomes will be easy considering the goal is to prevent in-hospital mortality after the surgical procedure.

My second priority is to improve the readmission model by obtaining information on patients once discharged. There are several initiatives underway at UNC Healthcare to prevent readmissions and better manage population health. One such initiative is to use consumer data on patients to gain better insight into their at-home habits to improve readmission prediction. I am very interested in novel data sources that may work to improve the model in this particular population of cirrhotic patients. I plan to collaborate with the UNC Healthcare population management data scientists to build a better model.

Two papers will be produced from the dissertation, one each for the NSQIP mortality and NSQIP readmission models. I plan to submit both papers to clinical journals such as *Gastroenterology* and *Clinical Gastroenterology and Hepatology*. Additionally, I plan to submit abstracts to the American College of Gastroenterology (ACG) and American Association for the Study of Liver Diseases (AASLD) meetings in 2015.

APPENDIX. TABLES AND FIGURES

Table 1. Model Measures and Justification

Measures	Description	Theory Behind Selection	Specific Aims
Age	Patient's age at time of surgery	Control for declining kidney and liver function due to age[7]	All aims
Gender	Male/Female	Control for differences in creatinine between genders[40, 41]	All aims
Pre-operative Body Mass Index (BMI)	Using height and weight from the data: BMI=[mass(lbs)/height (in)^2] x 703	Risk of mortality is increased in moribund patients[63-65]	All aims
Pre-operative Functional Status	Functional status assessment prior to surgery: 1-independent 2-partially dependent 3-Totally dependent	Controls for poor health status prior to surgery. Those with lower functional status may carry greater risk of mortality pre-operatively[66]	All aims
Ethnicity/Race	African American/White/Asian/Other	Control for differences across ethnicities	All aims
Pre-operative estimated Glomerular Filtration Rate (eGFR)	Use of the 2005 modification of diet and renal disease (MDRD*) formula: eGFR= 186 x $SCr^{-1.154}$ x $Age^{-0.203}$ x [1.210 if Black] x [0.743 if Female]	Captures early acute kidney injury often due to fluid overload from decompensated cirrhosis[40]	All aims
Pre-operative Serum Albumin	Measured in g/dL	Made by the liver and captures biosynthetic liver function-declines as liver function declines in cirrhosis[33]	All aims
Pre-operative Platelet Count	Measured in mmol/L	Hyponatremia (i.e., low sodium) associated with mortality across multiple conditions[35]	All aims
Pre-operative White Blood Cell Count (WBC)	May be indicative of pre-operative infection or inflammation	Has been shown counts >10,000 are associated with a high mortality rate of 54% versus 19% in intra-abdominal surgery in patients with cirrhosis of any etiology[67]	All aims

Pre-operative Serum Total Bilirubin	Measured in mg/dL	Captures biosynthetic liver function. Increases with cellular damage in the liver[67]	All aims
Pre-operative Prothrombin INR	INR(international normalized ratio): standard measure of the clotting capacity of blood	May detect risk of post-surgical bleeding as well as capture liver function[68]	All aims
Emergency Surgery	Surgery was due to emergent condition	Controls for a greater probability of mortality due to an emergent condition[67]	All aims
Type of Surgery	Surgical types include: 1) Cardiothoracic 2) General 3) Other-not classified elsewhere 4) Laparoscopic procedures 5) Cholecystectomy (lap or open) 6) Open abdominal 7) Hernia Repair (lap or open)	Controls for type of surgery[9]	All aims
Comorbidities	Includes: 1) Diabetes (Oral medication, Non-insulin or injectable insulin dependent)	Control for comorbidities that may increase the risk of post-surgical mortality[69-72]	All aims
American Society of Anesthesiologists Physical Status Classification (ASA class)	• Healthy patient • Mild systemic disease-no functional limitation • Severe systemic disease-definite functional limitation • Severe systemic disease that is a constant threat to life • Moribund-patient is not expected to survive without surgery	ASA classification is associated with post-operative moratlity when used pre-operatively to stratify risk in patients[73]	All aims
Duration of Anesthesia and Duration of Operation	Total duration of surgery from anesthesia start to anesthesia stop (minutes)	Greater time in surgery increases risk of readmission due to complications[74]	2a, 2b, 2c
Post-operative	Patient developed pneumonia	Risk of readmission is	2a, 2b, 2c

Pneumonia	based on radiology or laboratory data	greater if post-operative pneumonia developed[75]	
Post-operative Sepsis or Septic Shock	Patient developed sepsis or septic shock.	Patients with cirrhosis are at greater risk of sepsis[12, 75]	2a, 2b, 2c
Post-operative Acute or Progressive Renal Failure	Patient developed either acute or progressive renal failure.	Patients with decompensated cirrhosis are at greater risk of acute renal failure[76, 77]	2a, 2b, 2c
Post-operative Urinary Tract Infection	Patient developed urinary tract infection.	Urinary tract infection increases the risk of readmission[75]	2a, 2b, 2c
Post-operative Unplanned Intubation	Patient was removed from ventilator and re-intubated.	Complication resulted in an unplanned reintubation may increase risk of readmission[78]	2a, 2b, 2c
Post-operative Failure to Wean from Ventilator	On ventilator >48 hours	Extended time on the ventilator is associated with greater risk of mortality, length of stay and pneumonia[79]	2a, 2b, 2c
Discharge Destination	1-Skilled care, not home 2- Unskilled facility, not home. 3-Same pre-operative facility 4-Home 5-Separate acute care 6- Rehabilitation facility	Discharge destination has been associated with the odds of readmission[4]	2a, 2b, 2c
Total Length of Hospital stay	Days	Risk of secondary nosocomial infection may increase as length of hospital stay increases[80]	2a, 2b, 2c

Abbreviations: *MDRD=Modification of Diet in Renal Disease eGFR equation

Table 2. Surgical Categories

Surgery Type	Description
Cardiothoracic	*Surgeries requiring cutting into the thoracic cavity.*
General Surgery	*Surgeries not classified as cardiothoracic or abdominal.*
Open Abdominal	*Any surgery requiring cutting into the abdominal wall excluding hernia repair and cholecystectomy*
Any Laparoscopic Procedure	*Any laparoscopic procedure excluding lap cholecystectomy or lap hernia repair.*
Cholecystectomy	*Any cholecystectomy (lap or open)*
Hernia Repair	*Any hernia repair (lap or open)*
Other-not classified elsewhere	*Surgeries not classified elsewhere*

Table 3. Summary of Model Performance Measures

Measure	Statistic	Interpretation	Expected Value
Overall model performance	Brier Score	**Range: 0-0.25** Lower scores indicate better calibration and discrimination.	<=0.25 (given 50% chance of death)
Discrimination	AUROC, Sensitivity, Specificity, Positive Predictive Value, Negative Predictive Value, % correctly classified	AUROC: Plots sensitivity (SN)* vs. 1-specificity (SP)**. Range 0-1 with scores closer to 1 indicating better discrimination.	AUROC≥0.80
	Note: all measures are at a specified cut-point	Sensitivity: Ability of the model to identify true positives correctly at a specific cut-point	Sensitivity≥80% Specificity≥80%
		Specificity: Ability of the model to identify true negatives at a specific cut-point	
Calibration	Calibration-in-the Large	Taking the square root of the calibration-in-the-large gives the % difference between the predictive fraction of positive results versus the predicted fraction of negative results.	<5% difference
	Calibration-in-the Small	The square root of the calibration-in-the small gives the % difference between the predictive positives and negatives versus the actual fraction amongst the groups.	<5% difference
	Hosmer-Lemeshow Goodness-of-fit Test	The Hosmer-Lemeshow goodness-of-fit test is an indicator of how well my model fits the data in my sample. It indicates if the model is correctly specified. To simplify, is the actual number of patients that died or were readmitted match the predicted numbers. Hosmer-Lemeshow breaks the groups into 10 different groups of fitted values. If p>0.05 then we can assume our model's predictions are similar to the actual outcome across 10 different areas of the range of probabilities.	p>0.05 (accept model is correctly specified)

Cross-Validation	AUROC for both the **k-fold** analysis and the **leave-one-out** analysis	Predicts the ability of the model to maintain performance in an external dataset. k-fold: fold data into 10 smaller training/validation sets. Leave-one-Out: use entire set to train and apply to the observation left out. Continue until you have predicted all observations.	AUROC≥0.80

*SN=true positives/(true positives(TP) + false negatives(FN))
**SP= true negatives/(true negatives(TN) + false positives(FP))

Table 4. Sample Size Requirements for the Mortality and Readmission Cohorts

Cohorts	Incidence	Power	Alpha	Sample Size* Required (Sensitivity)	Sample Size Required* (Specificity)
Mortality Cohort	12.36%	80%	0.05	1,989	281
Readmissions Cohort	13.14%	80%	0.05	1,871	283

**The sample sizes are sufficient for all aims.*

Table 5. Bivariate Analyses

	Mortality Cohort				Unplanned Readmission Cohort			
Variable	**Overall**	**Discharged Alive**	**Died In-house**		**Overall**	**Unplanned Readmission**	**No Readmission**	
	n=4,535	**n=3,927**	**n=608**	**p-value**	**n=3,475**	**n=542**	**n=2,933**	**p-value**
Age (%)								
<40	6.9	7.2	5.1	0.06	7.5	7.3	8.1	0.52
41-50	13.2	13.7	10.2	0.02*	14.4	14.2	15.7	0.37
51-60	28.9	29.4	26.2	0.1	30.9	30.7	32.3	0.46
61-70	26.9	26.5	28.9	0.21	27	27.2	25.6	0.44
71-80	15.1	14.5	19.1	0.004*	13.4	13.5	12.7	0.64
>80	8.9	8.7	10.5	0.14	6.8	7.1	5.5	0.2
Race (%)								
White	75	75.5	71.9	0.06	76.7	76.7	76.8	0.96
Black	12	12.2	11.4	0.59	11.5	11.1	13.5	0.11
Other	12.9	12.4	16.7	0.003	11.7	12.1	9.6	0.1
Gender (%)								
Female	42.4	42.3	43.1	0.71	43.2	43.3	42.4	0.7
Male	57.6	57.7	56.9	0.71	56.8	56.7	57.6	0.7
Chronic Kidney Disease Stage (%)								
I. Normal	30.1	32.2	16.5	<0.001**	34.7	34.8	33.9	0.68
IIa. Some kidney damage	26.6	28	17.4	<0.001**	28.8	29.1	26.7	0.26

IIb. Moderate kidney damage	23.6	23	27.5	0.01*		22.4	22.4	22.3	0.99
III. Severe kidney damage	11.3	9.2	24.9	<0.001**		7.5	7.2	8.7	0.24
Kidney Failure	8.5	7.7	13.7	<0.001**		6.7	6.4	8.3	0.1
Emergent surgery (%)	43.5	39.3	70.6	<0.001**		36.5	36.7	35.2	0.51
Pre-operative Comorbidities & Procedures (%)									
Confusion (proxy for encephalopathy)	9.2	5.5	30	<0.001**		5.8	5.8	5.9	0.96
Alcohol use	8.2	7.2	13.7	0.003*		7.4	7.1	8.9	0.42
Chronic obstructive pulmonary disease	10	9.3	14.5	<0.001**		8.8	9.1	7.2	0.14
Peripheral Vascular Disease	4.6	4.2	6.3	0.21		4.4	4.5	3.6	0.57
Chronic Heart Failure	8	7	14.5	<0.001**		6.4	6.3	6.6	0.77
Smoker	27.6	27.4	28.8	0.47		28.9	28.9	28.6	0.88
Diabetic Status (%)									
Not diabetic	74.6	74.8	72.9	0.3		75.6	76.8	69.0	<0.001**
Insulin dependent	15.8	15.5	17.8	0.15		14.7	13.8	19.6	<0.001**
Non-Insulin dependent	9.7	9.7	9.4	0.78		9.7	9.3	11.4	0.13
Pre-operative transfusion required (%)	15.8	13.7	29.3	<0.001**		11	10.9	11.4	0.74
Type of Surgery (%)									
Open Abdominal	9.2	10.2	3.3	<0.001**		10.1	10.3	9.2	0.45
Cardiothorasic	7.1	7.2	6.6	0.59		5.1	5.1	5.2	0.96

General surgery	46.0	43.4	63.0	<0.001**	43.4	43	45.6	0.27
Other surgery type-not specifi	2.2	2.3	2.1	0.84	2.2	2.1	3	0.21
Any lapararoscopic procedure	11.4	10.4	17.8	<0.001**	10.6	10.8	9.2	0.27
Cholecystectomy	5.2	5.6	2.5	0.001*	5.2	5.4	4.4	0.37
Hernia repair	18.8	21.0	4.8	<0.001**	23.4	23.4	23.4	0.97
BMI Category (%)								
Normal	34.3	34.6	32	0.22	35.8	35.1	39.5	0.05
Underweight	4.9	4.6	7.1	0.01*	4.6	4.3	5.8	0.13
Overweight	29.8	30.7	23.7	<0.001**	31	31.5	28.6	0.19
Obese	31	30.1	37.3	<0.001**	28.6	29.1	26	0.14
Functional Status (%)								
Independent	85	87.7	66.8	<0.001**	88.3	88.5	87.2	0.41
Partially Dependent	9.8	8.7	16.8	<0.001**	8.4	8.2	9.3	0.43
Totally Dependent	5.3	3.6	16.3	<0.001**	3.3	3.3	3.5	0.8
ASA Class								
Mild Disturb	8	9.1	1.3	<0.001**	9.9	10.4	7.6	0.05
Severe Disturb	46.5	50.7	19	<0.001**	53.5	53.7	52.8	0.7
Life Threat	40	36.5	62.6	<0.001**	34.6	33.8	38.9	0.02*
Moribund	5.5	3.7	17	<0.001**	1.9	2.1	0.7	0.03*

Pre-Operative Laboratory Values [Median (25th, 75th percentile]								
Albumin	2.8 (2.2,3.4)	2.9 (2.3,3.4)	2.4 (1.9,2.9)	<0.001**	2.9 (2.4,3.5)	2.9 (2.1,3.5)	2.8 (2.3,3.4)	0.001*
Sodium	137 (134,140)	137 (134,140)	137 (133,141)	0.77	137 (134,140)	137 (134,140)	136 (133,139)	0.001*
Total Bilirubin	1.0 (0.5,1.9)	0.9 (0.5,1.7)	1.6 (0.8,3.4)	<0.001**	0.9 (0.5,1.8)	0.9 (0.5,1.7)	1.0 (0.5,2.1)	0.003*
Platelets	145 (91,254)	148 (94,259)	134 (75,218)	<0.001**	179 (103,280)	182 (106,283)	155 (91,263)	<0.001**
PT INR	1.3 (1.1,1.6)	1.3 (1.1,1.6)	1.5 (1.2,1.9)	<0.001**	1.2 (1.1,1.4)	1.2 (1.1,1.4)	1.3 (1.1,1.5)	0.005*
White Blood Cell count	7.9 (5.2,12.4)	7.6 (5.0,11.5)	11.8 (7.2,18.7)	<0.001**	7.6 (5.0,11.3)	7.6 (5.1, 11.4)	7.2 (4.8,10.5)	0.03*
estimated glomerular filtration rate (eGFR) [Median (25th, 75th percentile]	67.2 (36.3, 97.0)	71.1 (41.2,98.9)	40.0 (22.2,72.4)	<0.001**	73.8 (45.5,101.7)	74.1 (46.1,101.8)	71.5 (41.6,101.3)	0.17
MELD Category (%)								
MELD <=9	23.7	26.0	8.0	<0.001**	30.1	31.1	24.5	0.002**
MELD 10-19	51.9	53.3	41.7	<0.001**	53.6	53.4	54.4	0.66
MELD 20-29	19.1	16.4	37.9	<0.001**	13.6	12.8	18.3	<0.001**
MELD 30-39	4.9	4.0	11.1	<0.001**	2.6	2.6	2.6	>0.99
MELD >40	0.4	0.3	1.2	<0.001**	0.1	0.1	0.2	0.79
Childs-Pugh Category (%)								
CTP A	6.5	7.1	3.0	<0.001**	4.6	4.7	4.1	0.49

CTP B	29.3	31.7	14.0	<0.001**		23.9	24.6	20.1	0.02*
CTP C	64.1	61.2	83.1	<0.001**		71.5	70.7	75.8	0.01*
				.					
APRI score [Median (25th, 75th percentile]	0.6 (0.3,1.6)	0.6 (0.3,1.4)	1.1 (0.5,3.0)	<0.001**		0.5 (0.2,1.3)	0.5 (0.2,1.2)	0.6 (0.3,1.7)	<0.001**
Total Length of Stay in Days [Median (25th, 75th percentile]	10 (4,20)	9 (4,19)	14 (6,26)	<0.001**		9 (4,19)	9 (4,20)	10 (5,16)	0.17

*P<0.05

**P<0.001

Mortality and Readmission cohorts include only patients with laboratory values available to generate both the MELD score and the Moratlity/Readmission models.

P-values by non-parametric test for trend

Table 6. Comparison of the Cirrhotic Cohort with and without Laboratory Values

	Cirrhotic Cohort			
Variable	Overall	Missing Laboratory Values	All laboratory values available for models	
	n=4,916	n=273	n=4,643	p-value
Year of Operation				
2011	31.4	31.1	31.4	0.93
2012	28.1	24.5	28.3	0.18
2013	40.5	44.3	40.3	0.19
Age				
<40	6.9	7.7	6.8	0.57
41-50	13.1	14.3	13.0	0.54
51-60	28.6	24.9	28.9	0.16
61-70	26.8	26.7	26.8	0.98
71-80	15.4	17.2	15.3	0.39
>80	9.2	9.2	9.2	0.96
Race				
White	75.8	69.6	76.2	0.03*
Black	11.3	8.3	11.5	0.15
Other	12.9	22.1	12.4	<0.001**
Gender				
Female	42.2	43.6	42.2	0.64
Male	57.8	56.4	57.8	0.64
Emergent surgery	43.9	56.4	43.1	<0.001**
Pre-operative Comorbidities & Procedures				
Confusion (proxy for encephalopathy)	9.5	19.4	9.0	0.003**
Alcohol use	8.1	6.9	8.2	0.7
Chronic obstructive pulmonary disease	10.0	11.0	10.0	0.59
Peripheral Vascular Disease	4.6	0.0	4.8	0.06
Chronic Heart Failure	8.1	11.7	7.8	0.02*
Smoker	27.8	28.9	27.7	0.66

Diabetic Status (%)				
Not diabetic	74.5	76.2	74.4	0.5
Insulin dependent	15.9	15.0	15.9	0.69
Non-Insulin dependent	9.7	8.8	9.7	0.62
Pre-operative transfusion required (%)	15.7	23.4	15.2	<0.001**
Type of Surgery (%)				
Open Abdominal	9.3	10.6	9.2	0.44
Cardiothorasic	7.3	8.4	7.2	0.45
General surgery	45.8	41.0	46.1	0.1
Other surgery type-not specifi	2.4	2.9	2.3	0.52
Any lapararoscopic procedure	11.5	21.6	10.9	<0.001**
Cholecystectomy	5.2	3.3	5.3	0.15
Hernia repair	18.6	12.1	19.0	0.005*
BMI Category (%)				
Normal	34.5	29.9	34.8	0.12
Underweight	4.9	5.3	4.9	0.76
Overweight	29.6	29.1	29.7	0.85
Obese	30.9	35.7	30.7	0.1
Functional Status (%)				
Independent	84.7	82.3	84.8	0.27
Partially Dependent	9.9	6.8	10.1	0.08
Totally Dependent	5.5	10.9	5.1	<0.001**
ASA Class				
Mild Disturb	7.7	4.1	7.9	0.02*
Severe Disturb	45.8	30.4	46.7	<0.001**
Life Threat	40.7	47.8	40.2	0.01*
Moribund	5.8	17.8	5.1	<0.001**
Total Length of Stay in Days [Median (25th, 75th percentile]	10 (4,20)	11 (4,19)	10 (4,20)	>0.99

*P<0.05

**P<0.001

P-values by non-parametric test for trend

Table 7. Poisson Model for In-hospital Mortality (N=1,834)

					95% CI	
Variable	**IRR**	**Robust SE**	**Z**	**p-value**	**Low**	**High**
Age						
41-50	1.437	0.318	1.640	0.101	0.932	2.216
51-60	1.485	0.295	1.990	0.047*	1.005	2.192
61-70	2.047	0.407	3.600	<0.001**	1.386	3.022
71-80	2.186	0.458	3.730	<0.001**	1.450	3.295
>80	2.382	0.540	3.830	<0.001**	1.528	3.713
Male (ref. female)	1.006	0.085	0.070	0.948	0.852	1.186
BMI (ref. Normal)						
Underweight	1.310	0.214	1.660	0.097	0.952	1.803
Overweight	0.887	0.101	-1.050	0.294	0.709	1.110
Obese	1.065	0.104	0.650	0.517	0.880	1.290
Functional Status (ref: Independent)						
Partially Dependent	1.365	0.161	2.630	0.009*	1.083	1.720
Totally Dependent	1.384	0.148	3.040	0.002*	1.122	1.707
ASA Class (ref: Mild Disturbance)						
Life Threat	4.030	1.541	3.650	<0.001**	1.905	8.525
Moribund	5.478	2.159	4.310	<0.001**	2.530	11.862
Race (ref: White)						
Black	0.994	0.126	-0.050	0.963	0.776	1.274
Other	0.974	0.115	-0.220	0.824	0.773	1.228
estimated glomerular filtration rate (eGFR)	0.997	0.001	-2.030	0.043*	0.995	1.000
Albumin	0.801	0.044	-4.050	<0.001**	0.719	0.892
Sodium	1.004	0.007	0.540	0.590	0.990	1.017
Platelet count	0.999	0.000	-2.810	0.005*	0.998	1.000
White blood cell count	1.007	0.004	1.660	0.097	0.999	1.016
Total bilirubin	1.100	0.012	8.800	<0.001**	1.077	1.124
PT INR	1.274	0.059	5.220	<0.001**	1.164	1.396

Emercenty surgery	1.698	0.187	4.810	<0.001**	1.368	2.107
Diabetic Status (ref: Non-diabetic)						
Insulin dependent	0.854	0.093	-1.460	0.145	0.690	1.056
Non-insulin dependent	0.933	0.130	-0.500	0.618	0.710	1.226
Chronic obstructive pulmonary disease	1.001	0.114	0.010	0.995	0.800	1.251
Chronic Heart Failure	1.179	0.125	1.560	0.119	0.958	1.451
Surgery Type (ref: Open Abdominal)						
Cardiothorasic	1.484	0.459	1.280	0.202	0.809	2.721
General surgery	1.303	0.345	1.000	0.317	0.776	2.188
Other surgery type-not specified	1.032	0.453	0.070	0.943	0.437	2.438
Any lapararoscopic procedure	1.450	0.401	1.350	0.178	0.844	2.492
Cholecystectomy	1.118	0.398	0.310	0.754	0.556	2.247
Hernia repair	0.468	0.159	-2.230	0.025*	0.241	0.911
_cons	0.014	0.015	-3.850	0.000	0.002	0.121

*P<0.05

**P<0.001

Table 8. Logit Model for In-Hospital Mortality (N=1,945)

					95% CI	
Variable	**Odds Ratio**	**Robust SE**	**Z**	**p-value**	**Low**	**High**
Age (ref: <=40)						
41-50	1.799	0.639	1.650	0.098	0.897	3.609
51-60	2.041	0.666	2.190	0.029*	1.077	3.868
61-70	3.519	1.165	3.800	<0.001**	1.839	6.733
71-80	3.688	1.297	3.710	<0.001**	1.851	7.349
>80	4.385	1.666	3.890	<0.001**	2.083	9.232
Race (ref: White)						
Black	1.011	0.206	0.050	0.958	0.678	1.507
Other	1.071	0.202	0.370	0.714	0.741	1.549
Functional Status (ref: Independent)						
Partially	1.731	0.323	2.940	0.003*	1.201	2.496
Totally	2.171	0.426	3.950	<0.001**	1.478	3.189
ASA Class (ref: Mild Disturbance)						
Life Threat	3.563	1.443	3.140	0.002*	1.611	7.881
Moribund	8.802	3.915	4.890	<0.001**	3.681	21.047
Comorbidities						
Chronic Heart Failure	1.185	0.229	0.880	0.379	0.812	1.731
Chronic Obstructive Pulmonary Disease	1.060	0.191	0.320	0.747	0.744	1.510
Diabetic status (ref: Non-diabetic)						
Insulin	0.727	0.125	-1.860	0.063	0.519	1.017
Non-Insulin	0.882	0.192	-0.580	0.564	0.575	1.352
Pre-Operative Laboratory Values						
estimated glomerular filtration rate (eGFR)	0.997	0.002	-1.730	0.084	0.994	1.000
Albumin	0.658	0.057	-4.840	<0.001**	0.555	0.779
Platelet Count	0.999	0.001	-2.160	0.031*	0.997	1.000
Total Bilirubin	1.229	0.036	6.980	<0.001**	1.160	1.302
PT INR	1.873	0.246	4.790	<0.001**	1.449	2.422

Emergency surgery (ref: Scheduled)	2.239	0.338	5.350	<0.001**	1.666	3.010
Surgery Type (ref: Open Abdominal)						
Cardiothorasic	1.950	0.831	1.570	0.117	0.845	4.497
General surgery	1.467	0.536	1.050	0.294	0.717	3.004
Other surgery type-not specified	1.015	0.586	0.030	0.980	0.327	3.147
Any lapararoscopic procedure	1.932	0.752	1.690	0.091	0.901	4.142
Cholecystectomy	1.009	0.529	0.020	0.986	0.362	2.817
Hernia repair	0.349	0.151	-2.430	0.015*	0.149	0.815
_cons	0.015	0.011	-5.950	0.000	0.004	0.060

*P<0.01
**P<0.001

Table 9. Mortality Model Validation Summary

Calibration		
Calibration -in-the-Large	0.0004	
Calibration-in-the-Small	0.0014	
Hosmer-Lemeshow Goodness-of-fit (chi^2 test)	8.97	p=0.345
Discrimination		
AUROC	0.8282	
Sensitivity (cut-point=0.2150)	81.10%	
Specificity (cut-point=0.2150)	73.60%	
Positive Predictive Value (cut-point=0.2150)	47.60%	
Negative Predictive Value (cut-point=0.2150)	92.90%	
Correctly Classified	75.30%	

Out-of-Sample performance	
	AUROC
K-fold (k=10) method	0.8402
Leave-one-out method	0.8420

Overall Model Performance	
Brier Score	0.1269

Cut-Point Performance

Cutpoint	Sensitivity	Specificity	Classified	LR+	LR-
0.149	88.96%	60.83%	67.25%	2.271	0.1814
0.150	88.29%	61.09%	67.30%	2.2692	0.1917
0.160	86.71%	62.82%	68.28%	2.3325	0.2115
0.170	85.14%	65.09%	69.67%	2.4387	0.2284
0.180	84.01%	67.42%	71.21%	2.5787	0.2372
0.190	83.78%	68.62%	72.08%	2.6701	0.2363
0.200	82.66%	70.95%	73.62%	2.8456	0.2444
0.215	81.08%	73.55%	75.27%	3.0656	0.2572
0.300	67.12%	82.74%	79.18%	3.8897	0.3974

Table 10. Poisson Model for Readmission within 30-days of the Index Surgery (N=1,495)

					95% CI	
Variable	**IRR**	**Robust SE**	**Z**	**p-value**	**Low**	**High**
Age (ref: <=40)						
41-50	1.061	0.255	0.250	0.805	0.663	1.698
51-60	0.851	0.188	-0.730	0.466	0.552	1.313
61-70	0.790	0.177	-1.050	0.293	0.509	1.226
71-80	0.775	0.208	-0.950	0.343	0.458	1.312
>80	0.818	0.257	-0.640	0.522	0.442	1.513
Male (ref. female)	0.789	0.097	-1.930	0.054	0.619	1.004
BMI (ref. Normal)						
Underweight	0.813	0.226	-0.740	0.457	0.472	1.402
Overweight	0.806	0.112	-1.550	0.121	0.614	1.058
Obese	0.789	0.117	-1.600	0.109	0.591	1.054
Functional Status (ref: Independent)						
Partially Dependent	0.737	0.170	-1.320	0.187	0.469	1.159
Totally Dependent	1.374	0.339	1.290	0.198	0.847	2.229
ASA Class (ref: Mild Disturbance)						
Life Threat	1.628	0.302	2.630	0.009*	1.132	2.342
Moribund	0.721	0.377	-0.630	0.531	0.259	2.006
Race (ref: White)						
Black	1.157	0.186	0.910	0.364	0.844	1.586
Other	0.936	0.190	-0.330	0.743	0.629	1.393
Pre-operative Laboratory Values						
estimated glomerular filtration rate (eGFR)	0.999	0.001	-0.910	0.365	0.996	1.001
Albumin	0.853	0.069	-1.970	0.049*	0.728	0.999
Sodium	0.989	0.013	-0.860	0.388	0.964	1.014
Platelet count	0.999	0.001	-1.930	0.054	0.998	1.000
White blood cell count	1.008	0.010	0.750	0.452	0.988	1.028
Total bilirubin	0.990	0.032	-0.300	0.767	0.929	1.056
PT INR	1.480	0.193	3.010	0.003*	1.146	1.911

Emercency surgery	0.930	0.134	-0.500	0.615	0.702	1.233
Diabetic Status (ref: Non-diabetic)						
Insulin dependent	1.617	0.232	3.360	0.001*	1.221	2.141
Non-insulin dependent	1.048	0.220	0.230	0.822	0.695	1.581
Surgery Type (ref: Open Abdominal)						
Cardiothorasic	0.986	0.282	-0.050	0.959	0.563	1.726
General surgery	1.119	0.259	0.480	0.628	0.711	1.760
Other surgery type-not specified	1.285	0.408	0.790	0.429	0.690	2.393
Any lapararoscopic procedure	1.258	0.346	0.840	0.403	0.734	2.157
Cholecystectomy	0.581	0.243	-1.300	0.194	0.256	1.318
Hernia repair	0.780	0.202	-0.960	0.336	0.469	1.295
Post-Operative Variables						
Pneumonia	1.032	0.397	0.080	0.935	0.485	2.195
Sepsis	1.966	0.301	4.410	<0.001**	1.456	2.656
Time on Table	1.002	0.001	3.040	0.002*	1.001	1.003
Time from operation to discharge	0.932	0.011	-5.760	<0.001**	0.910	0.955
Acute or Progressive Renal Failure	1.570	0.330	2.150	0.032*	1.041	2.369
Urinary Tract Infection	2.225	0.514	3.460	0.001*	1.415	3.499
On Ventilator >48 hrs	0.760	0.154	-1.350	0.175	0.511	1.130
Unplanned intubation	1.518	0.305	2.080	0.038*	1.024	2.250
Discharge Destination						
Home	1.210	0.203	1.140	0.256	0.871	1.682
Rehab	0.550	0.172	-1.910	0.056	0.298	1.016
Separate Acute Care	0.492	0.223	-1.560	0.118	0.202	1.196
Total Hospital Length of Stay	1.006	0.009	0.730	0.466	0.990	1.023
_cons	0.993	1.866	0.000	0.997	0.025	39.506

*P<0.01, **P<0.001

Table 11. Logit Model for Readmission within 30-days of the Index Surgery (N=1,495)

					95% CI	
Variable	**OR**	**Robust SE**	**Z**	**p-value**	**Low**	**High**
Age (ref: <=40)						
41-50	1.039	0.344	0.110	0.909	0.543	1.988
51-60	0.788	0.240	-0.780	0.435	0.433	1.433
61-70	0.699	0.215	-1.170	0.244	0.383	1.276
71-80	0.661	0.241	-1.140	0.256	0.324	1.349
>80	0.693	0.287	-0.880	0.377	0.308	1.562
Male (ref. female)	0.705	0.120	-2.050	0.041*	0.504	0.985
BMI (ref. Normal)						
Underweight	0.763	0.275	-0.750	0.453	0.376	1.548
Overweight	0.769	0.143	-1.410	0.159	0.534	1.108
Obese	0.729	0.146	-1.590	0.113	0.493	1.078
Functional Status (ref: Independent)						
Partially Dependent	0.657	0.200	-1.380	0.167	0.362	1.192
Totally Dependent	1.666	0.596	1.430	0.154	0.826	3.360
ASA Class (ref: Mild Disturbance)						
Life Threat	1.828	0.432	2.560	0.011*	1.151	2.905
Moribund	0.689	0.460	-0.560	0.577	0.186	2.550
Race (ref: White)						
Black	1.185	0.264	0.760	0.446	0.766	1.834
Other	0.896	0.238	-0.410	0.679	0.533	1.507
Pre-operative Laboratory Values						
estimated glomerular filtration rate (eGFR)	0.998	0.002	-0.960	0.339	0.995	1.002
Albumin	0.793	0.088	-2.090	0.037*	0.638	0.986
Sodium	0.986	0.017	-0.840	0.401	0.953	1.019
Platelet count	0.999	0.001	-1.990	0.047*	0.997	1.000
White blood cell count	1.008	0.013	0.630	0.528	0.983	1.034
Total bilirubin	0.993	0.048	-0.140	0.891	0.904	1.091
PT INR	1.734	0.367	2.600	0.009*	1.145	2.624

Emercency surgery	0.884	0.168	-0.650	0.518	0.609	1.284
Diabetic Status (ref: Non-diabetic)						
Insulin dependent	1.980	0.403	3.360	0.001*	1.329	2.951
Non-insulin dependent	1.094	0.298	0.330	0.741	0.642	1.866
Surgery Type (ref: Open Abdominal)						
Cardiothorasic	0.992	0.375	-0.020	0.983	0.473	2.082
General surgery	1.210	0.358	0.650	0.519	0.678	2.161
Other surgery type-not specified	1.449	0.670	0.800	0.423	0.585	3.586
Any lapararoscopic procedure	1.397	0.494	0.940	0.345	0.698	2.794
Cholecystectomy	0.500	0.254	-1.360	0.172	0.185	1.353
Hernia repair	0.732	0.246	-0.930	0.353	0.379	1.414
Post-Operative Variables						
Pneumonia	0.977	0.492	-0.050	0.963	0.364	2.624
Sepsis	2.641	0.618	4.150	<0.001**	1.670	4.177
Time on Table	1.002	0.001	2.800	0.005*	1.001	1.004
Time from operation to discharge	0.910	0.016	-5.460	<0.001**	0.880	0.941
Acute or Progressive Renal Failure	2.028	0.673	2.130	0.033*	1.059	3.885
Urinary Tract Infection	3.248	1.154	3.320	0.001*	1.619	6.516
On Ventilator >48 hrs	0.692	0.195	-1.310	0.191	0.399	1.202
Unplanned intubation	1.837	0.559	2.000	0.046*	1.012	3.333
Discharge Destination						
Home	1.251	0.286	0.980	0.327	0.799	1.960
Rehab	0.452	0.176	-2.040	0.041*	0.211	0.968
Separate Acute Care	0.382	0.230	-1.600	0.110	0.117	1.244
Total Hospital Length of Stay	1.008	0.012	0.680	0.497	0.985	1.031
_cons	2.340	5.784	0.340	0.731	0.018	297.365

*P<0.01, **P<0.001

Table 12. Readmission Model Validation Summary

Calibration					
Calibration -in-the-Large	0.0000			0	
Calibration-in-the-Small	0.0007			2.65%	
Hosmer-Lemeshow Goodness-of-fit (chi^2 test)	9.40	p=0.310			
Discrimination					
AUROC	0.7541				
Sensitivity (cut-point=0.1425)	76.64%				
Specificity (cut-point=0.1425)	60.91%				
Positive Predictive Value (cut-point=0.1425)	27.66%				
Negative Predictive Value (cut-point=0.1425)	93.04%				
Correctly Classified	63.48%				
Out-of-Sample performance					
	AUROC				
K-fold (k=10) method	0.7602				
Leave-one-out method	0.7541				
Overall Model Performance					
Brier Score	0.1186				
Cut-Point Performance					
Cutpoint	**Sensitivity**	**Specificity**	**Classified**	**LR+**	**LR-**
0.143	76.64%	60.91%	63.48%	1.9607	0.3835
0.150	75.00%	63.47%	65.35%	2.0531	0.3939
0.160	72.95%	65.79%	66.96%	2.1323	0.4112
0.170	70.49%	68.82%	69.10%	2.2612	0.4287
0.180	66.39%	71.46%	70.64%	2.3266	0.4703
0.190	62.70%	73.86%	72.04%	2.3989	0.5049
0.200	61.48%	76.02%	73.65%	2.5635	0.5068

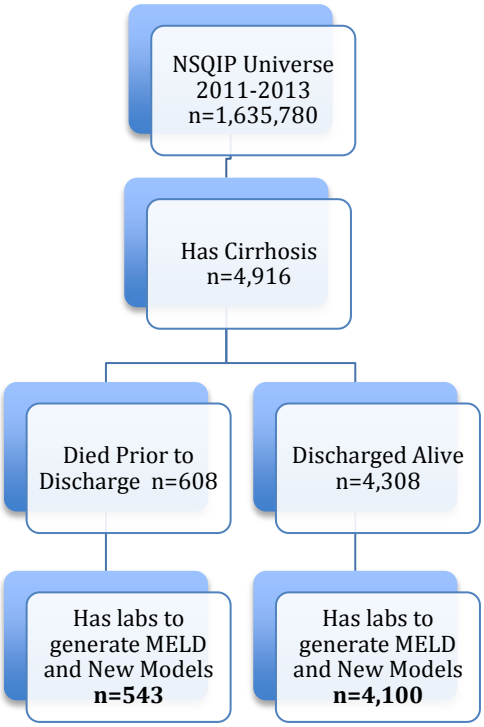

Figure 1. Mortality cohort.

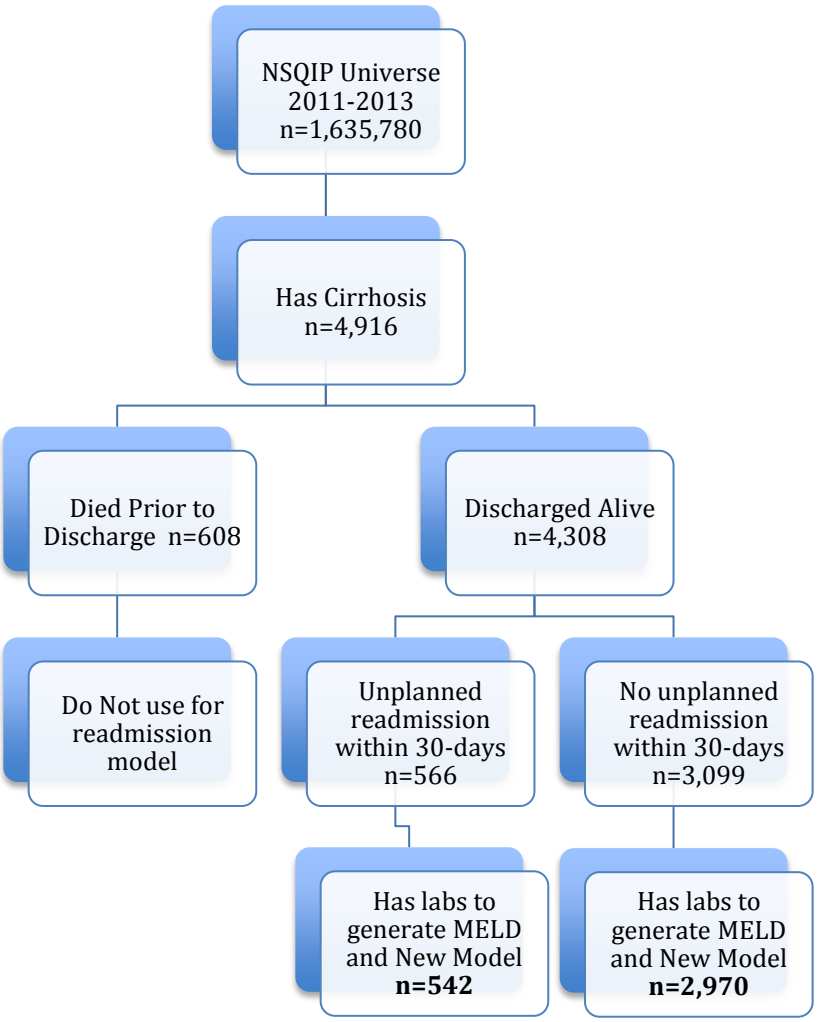

Figure 2. Readmission cohort.

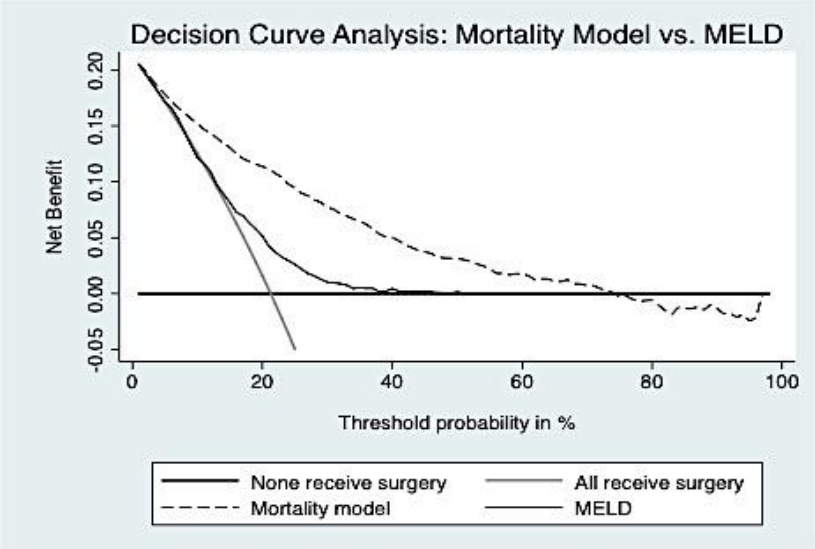

Figure 3. Decision curve analysis (mortality model).

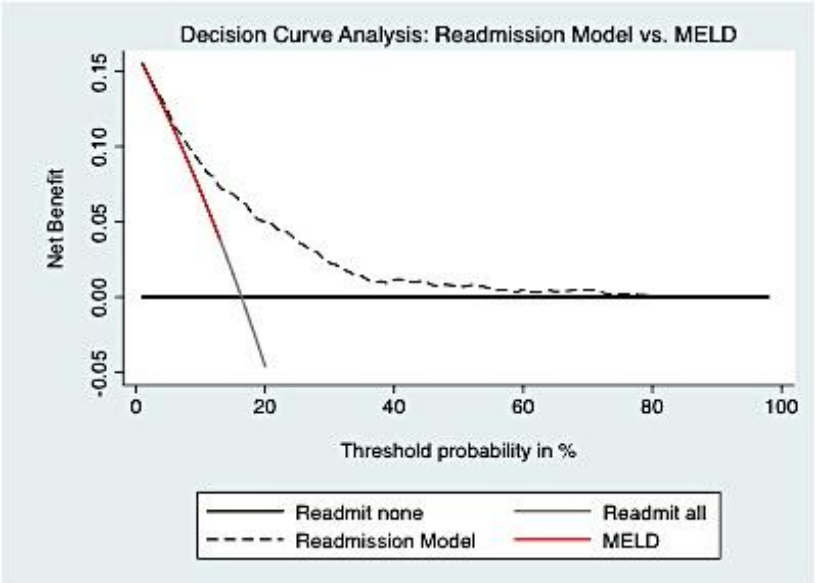

Figure 4. Decision curve analysis (readmission model).

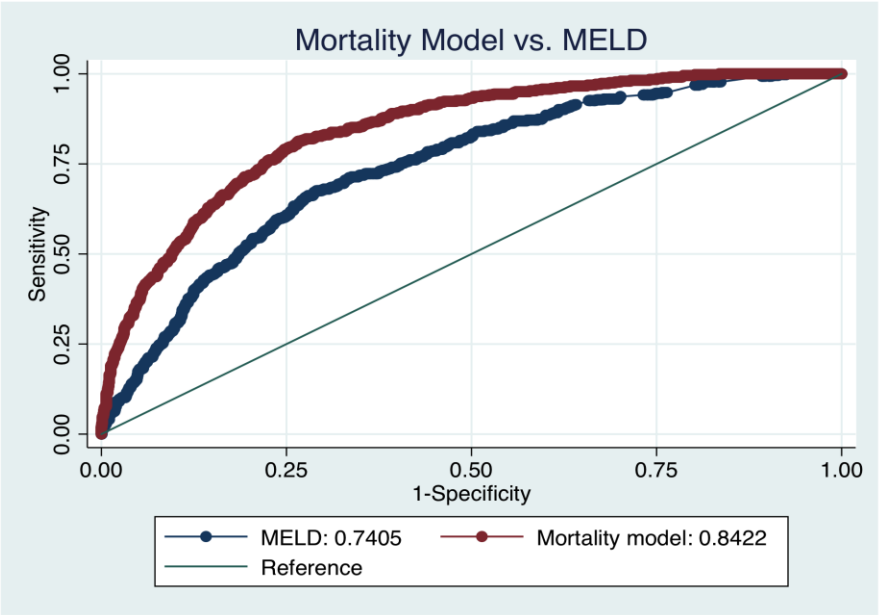

Figure 5. Comparison of the mortality model to the MELD score (AUROC).

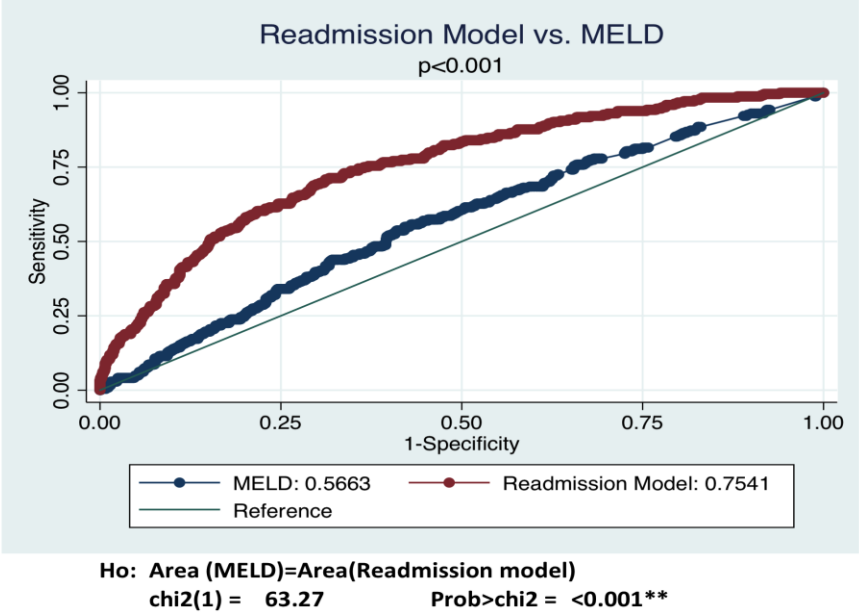

Figure 6. Comparison of the readmission model to the MELD score (AUROC).

REFERENCES

1. Runyon, B.A. and A.P.G. Committee, *Management of adult patients with ascites due to cirrhosis: an update.* Hepatology, 2009. **49**(6): p. 2087-107.

2. Davis, G.L., et al., *Aging of hepatitis C virus (HCV)-infected persons in the United States: a multiple cohort model of HCV prevalence and disease progression.* Gastroenterology, 2010. **138**(2): p. 513-21, 521 e1-6.

3. Talwalkar, J.A., *Determining the extent of quality health care for hospitalized patients with cirrhosis.* Hepatology, 2005. **42**(2): p. 492-4.

4. Volk, M.L., et al., *Hospital readmissions among patients with decompensated cirrhosis.* Am J Gastroenterol, 2012. **107**(2): p. 247-52.

5. Stepanova, M., et al., *In-hospital mortality and economic burden associated with hepatic encephalopathy in the United States from 2005 to 2009.* Clin Gastroenterol Hepatol, 2012. **10**(9): p. 1034-41 e1.

6. Nicoll, A., *Surgical risk in patients with cirrhosis.* J Gastroenterol Hepatol, 2012. **27**(10): p. 1569-75.

7. Teh, S.H., et al., *Risk factors for mortality after surgery in patients with cirrhosis.* Gastroenterology, 2007. **132**(4): p. 1261-9.

8. Csikesz, N.G., et al., *Nationwide volume and mortality after elective surgery in cirrhotic patients.* J Am Coll Surg, 2009. **208**(1): p. 96-103.

9. Kim, T.H., et al., *The risk of perioperative adverse events in patients with chronic liver disease.* Liver Int, 2014.

10. Scott, R.A., et al., *Acute kidney injury is independently associated with death in patients with cirrhosis.* Frontline Gastroenterol, 2013. **4**(3): p. 191-197.

11. Deleuran, T., et al., *Cirrhosis patients have increased risk of complications after hip or knee arthroplasty.* Acta Orthop, 2014: p. 1-6.

12. Ganesh, S., et al., *Risk factors for frequent readmissions and barriers to transplantation in patients with cirrhosis.* PLoS One, 2013. **8**(1): p. e55140.

13. Davis, C., et al., *The Revised Cardiac Risk Index in the new millennium: a single-centre prospective cohort re-evaluation of the original variables in 9,519 consecutive elective surgical patients.* Can J Anaesth, 2013. **60**(9): p. 855-63.

14. Bozic, K.J., et al., *Bundled payments in total joint arthroplasty: targeting opportunities for quality improvement and cost reduction.* Clin Orthop Relat Res, 2014. **472**(1): p. 188-93.

15. Miller, A., et al., *Implementing goal-directed protocols reduces length of stay after cardiac surgery.* J Cardiothorac Vasc Anesth, 2014. **28**(3): p. 441-7.

16. Shih, T., et al., *Medicare's Hospital Readmission Reduction Program in Surgery May Disproportionately Affect Minority-Serving Hospitals.* Ann Surg, 2014.

17. Berkowitz, S.A. and G.F. Anderson, *Medicare beneficiaries most likely to be readmitted.* J Hosp Med, 2013. **8**(11): p. 639-41.

18. Bowles, K.H., et al., *Successful Electronic Implementation of Discharge Referral Decision Support Has a Positive Impact on 30- and 60-day Readmissions.* Res Nurs Health, 2015.

19. Volk, M.L., M. Roney, and R.M. Merion, *Systematic bias in surgeons' predictions of the donor-specific risk of liver transplant graft failure.* Liver Transpl, 2013. **19**(9): p. 987-90.

20. Weeks, W.B., et al., *Geographic variation in admissions for knee replacement, hip replacement, and hip fracture in France: evidence of supplier-induced demand in for-profit and not-for-profit hospitals.* Med Care, 2014. **52**(10): p. 909-17.

21. Kahn, J.M., *Predicting outcome in critical care: past, present and future.* Curr Opin Crit Care, 2014. **20**(5): p. 542-3.

22. Ji, R., et al., *A novel risk score to predict 1-year functional outcome after intracerebral hemorrhage and comparison with existing scores.* Crit Care, 2013. **17**(6): p. R275.

23. Donovan, M.J., et al., *Postoperative systems models more accurately predict risk of significant disease progression than standard risk groups and a 10-year postoperative nomogram: potential impact on the receipt of adjuvant therapy after surgery.* BJU Int, 2012. **109**(1): p. 40-5.

24. Legare, F. and H.O. Witteman, *Shared decision making: examining key elements and barriers to adoption into routine clinical practice.* Health Aff (Millwood), 2013. **32**(2): p. 276-84.

25. Blatnik, J.A., et al., *Thirty-day readmission after ventral hernia repair: predictable or preventable?* Surg Endosc, 2011. **25**(5): p. 1446-51.

26. Al Sibae, M.R. and M.S. Cappell, *Accuracy of MELD scores in predicting mortality in decompensated cirrhosis from variceal bleeding, hepatorenal syndrome, alcoholic hepatitis, or acute liver failure as well as mortality after non-transplant surgery or TIPS.* Dig Dis Sci, 2011. **56**(4): p. 977-87.

27. Kamath, P.S., et al., *A model to predict survival in patients with end-stage liver disease.* Hepatology, 2001. **33**(2): p. 464-70.

28. Bernardi, M., S. Gitto, and M. Biselli, *The MELD score in patients awaiting liver transplant: strengths and weaknesses.* J Hepatol, 2011. **54**(6): p. 1297-306.

29. Dorwart, W.V., *Bilirubin interference in kinetic creatinine determination.* Clin Chem, 1979. **25**(1): p. 196-7.

30. Cholongitas, E., et al., *Different methods of creatinine measurement significantly affect MELD scores.* Liver Transpl, 2007. **13**(4): p. 523-9.

31. Kuster, N., et al., *Limitations of compensated Jaffe creatinine assays in cirrhotic patients.* Clin Biochem, 2012. **45**(4-5): p. 320-5.

32. Gluhovschi, C., et al., *Is There any Difference Between the Glomerular Filtration Rate of Patients With Chronic Hepatitis B and C and Patients With Cirrhosis?* Hepat Mon, 2013. **13**(4): p. e6789.

33. Myers, R.P., et al., *Revision of MELD to include serum albumin improves prediction of mortality on the liver transplant waiting list.* PLoS One, 2013. **8**(1): p. e51926.

34. Hsu, C.Y., et al., *Comparison of the model for end-stage liver disease (MELD), MELD-Na and MELDNa for outcome prediction in patients with acute decompensated hepatitis.* Dig Liver Dis, 2010. **42**(2): p. 137-42.

35. Biggins, S.W., et al., *Evidence-based incorporation of serum sodium concentration into MELD.* Gastroenterology, 2006. **130**(6): p. 1652-60.

36. Luca, A., et al., *An integrated MELD model including serum sodium and age improves the prediction of early mortality in patients with cirrhosis.* Liver Transpl, 2007. **13**(8): p. 1174-80.

37. Kao, H.K., et al., *The roles of albumin levels in head and neck cancer patients with liver cirrhosis undergoing tumor ablation and microsurgical free tissue transfer.* PLoS One, 2012. **7**(12): p. e52678.

38. Arif, R., et al., *Predictive risk factors for patients with cirrhosis undergoing heart surgery.* Ann Thorac Surg, 2012. **94**(6): p. 1947-52.

39. Lopez-Delgado, J.C., et al., *Short-term independent mortality risk factors in patients with cirrhosis undergoing cardiac surgery.* Interact Cardiovasc Thorac Surg, 2013. **16**(3): p. 332-8.

40. Myers, R.P., et al., *Gender, renal function, and outcomes on the liver transplant waiting list: assessment of revised MELD including estimated glomerular filtration rate.* J Hepatol, 2011. **54**(3): p. 462-70.

41. Lai, J.C., et al., *Height contributes to the gender difference in wait-list mortality under the MELD-based liver allocation system.* Am J Transplant, 2010. **10**(12): p. 2658-64.

42. Seamon, M.J., et al., *Do chronic liver disease scoring systems predict outcomes in trauma patients with liver disease? A comparison of MELD and CTP.* J Trauma, 2010. **69**(3): p. 568-73.

43. Suman, A., et al., *Predicting outcome after cardiac surgery in patients with cirrhosis: a comparison of Child-Pugh and MELD scores.* Clin Gastroenterol Hepatol, 2004. **2**(8): p. 719-23.

44. Surgeons, A.C.o., *National Surgical Quality Improvement Program*, A.C.o. Surgeons, Editor. 2002, ACS: Chicago, IL.

45. Prohic, D., et al., *Prognostic markers in patients with decompensated cirrhosis.* Med Glas (Zenica), 2014. **11**(1): p. 99-104.

46. Lu, S.N., et al., *Thrombocytopenia as a surrogate for cirrhosis and a marker for the identification of patients at high-risk for hepatocellular carcinoma.* Cancer, 2006. **107**(9): p. 2212-22.

47. Farnsworth, N., et al., *Child-Turcotte-Pugh versus MELD score as a predictor of outcome after elective and emergent surgery in cirrhotic patients.* Am J Surg, 2004. **188**(5): p. 580-3.

48. Zou, G., *A modified poisson regression approach to prospective studies with binary data.* Am J Epidemiol, 2004. **159**(7): p. 702-6.

49. Zou, G.Y. and A. Donner, *Extension of the modified Poisson regression model to prospective studies with correlated binary data.* Stat Methods Med Res, 2013. **22**(6): p. 661-70.

50. Barros, A.J. and V.N. Hirakata, *Alternatives for logistic regression in cross-sectional studies: an empirical comparison of models that directly estimate the prevalence ratio.* BMC Med Res Methodol, 2003. **3**: p. 21.

51. Dalton, J.E., *Flexible recalibration of binary clinical prediction models.* Stat Med, 2013. **32**(2): p. 282-9.

52. Steyerberg, E.W., et al., *Internal validation of predictive models: efficiency of some procedures for logistic regression analysis.* J Clin Epidemiol, 2001. **54**(8): p. 774-81.

53. Steyerberg, E.W., et al., *Assessing the performance of prediction models: a framework for traditional and novel measures.* Epidemiology, 2010. **21**(1): p. 128-38.

54. Meijer, R.J. and J.J. Goeman, *Efficient approximate k-fold and leave-one-out cross-validation for ridge regression.* Biom J, 2013. **55**(2): p. 141-55.

55. Jones, S.R., S. Carley, and M. Harrison, *An introduction to power and sample size estimation.* Emergency medicine journal : EMJ, 2003. **20**(5): p. 453-8.

56. Umemura, T., et al., *Serum sodium concentration is associated with increased risk of mortality in patients with compensated liver cirrhosis.* Hepatol Res, 2015. **45**(7): p. 739-44.

57. John, S. and P.J. Thuluvath, *Hyponatremia in cirrhosis: pathophysiology and management.* World J Gastroenterol, 2015. **21**(11): p. 3197-205.

58. Sheikh, K., *Investigation of selection bias using inverse probability weighting.* Eur J Epidemiol, 2007. **22**(5): p. 349-50.

59. Singal, A.G., et al., *An automated model using electronic medical record data identifies patients with cirrhosis at high risk for readmission.* Clin Gastroenterol Hepatol, 2013. **11**(10): p. 1335-1341 e1.

60. Walsh, C. and G. Hripcsak, *The effects of data sources, cohort selection, and outcome definition on a predictive model of risk of thirty-day hospital readmissions.* J Biomed Inform, 2014. **52**: p. 418-26.

61. Hoehn, R.S., et al., *Effect of pretransplant diabetes on short-term outcomes after liver transplantation: a national cohort study.* Liver Int, 2015. **35**(7): p. 1902-9.

62. *Issue 4 (August 2010) Management of Ascites, Spontaneous Bacterial Peritonitis, and Hepatorenal Syndrome in Cirrhosis*. EASL Clinical Practice Guidelines 2010 October 2014]; Available from: http://www.easl.eu/_clinical-practice-guideline.

63. Schumann, R., et al., *Association of metabolic syndrome and surgical factors with pulmonary adverse events, and longitudinal mortality in bariatric surgerydagger.* Br J Anaesth, 2014.

64. Conzen, K.D., et al., *Morbid obesity in liver transplant recipients adversely affects longterm graft and patient survival in a single-institution analysis.* HPB (Oxford), 2014.

65. Yap, C.H., et al., *Effect of obesity on early morbidity and mortality following cardiac surgery.* Heart Lung Circ, 2007. **16**(1): p. 31-6.

66. Dayama, A., et al., *Early outcomes and perioperative risk assessment in elective open thoracoabdominal aortic aneurysm repair: An analysis of national data over a five-year period.* Vascular, 2015.

67. Garrison, R.N., et al., *Clarification of risk factors for abdominal operations in patients with hepatic cirrhosis.* Ann Surg, 1984. **199**(6): p. 648-55.

68. Kim, Y.K., et al., *Factors associated with changes in coagulation profiles after living donor hepatectomy.* Transplant Proc, 2010. **42**(7): p. 2430-5.

69. Korol, E., et al., *A systematic review of risk factors associated with surgical site infections among surgical patients.* PLoS One, 2013. **8**(12): p. e83743.

70. Cingoz, F., et al., *Is chronic obstructive pulmonary disease a risk factor for epistaxis after coronary artery bypass graft surgery?* Cardiovasc J Afr, 2014. **25**: p. 1-3.

71. Garcia, S. and E.O. McFalls, *Perioperative clinical variables and long-term survival following vascular surgery.* World J Cardiol, 2014. **6**(10): p. 1100-7.

72. Blessberger, H., et al., *Perioperative beta-blockers for preventing surgery-related mortality and morbidity.* Cochrane Database Syst Rev, 2014. **9**: p. CD004476.

73. Wolters, U., et al., *ASA classification and perioperative variables as predictors of postoperative outcome.* Br J Anaesth, 1996. **77**(2): p. 217-22.

74. Kim, B.D., et al., *Predictors of unplanned readmission in patients undergoing lumbar decompression: multi-institutional analysis of 7016 patients.* J Neurosurg Spine, 2014. **20**(6): p. 606-16.

75. Vogel, T.R., V.Y. Dombrovskiy, and S.F. Lowry, *Impact of infectious complications after elective surgery on hospital readmission and late deaths in the U.S. Medicare population.* Surg Infect (Larchmt), 2012. **13**(5): p. 307-11.

76. Garcia-Tsao, G., C.R. Parikh, and A. Viola, *Acute kidney injury in cirrhosis.* Hepatology, 2008. **48**(6): p. 2064-77.

77. Hampel, H., et al., *Risk factors for the development of renal dysfunction in hospitalized patients with cirrhosis.* Am J Gastroenterol, 2001. **96**(7): p. 2206-10.

78. Nandyala, S.V., et al., *Incidence, risk factors, and outcomes of postoperative airway management after cervical spine surgery.* Spine (Phila Pa 1976), 2014. **39**(9): p. E557-63.

79. MacIntyre, N.R., *The ventilator discontinuation process: an expanding evidence base.* Respir Care, 2013. **58**(6): p. 1074-86.

80. Bajaj, J.S., et al., *Second infections independently increase mortality in hospitalized patients with cirrhosis: the North American consortium for the study of end-stage liver disease (NACSELD) experience.* Hepatology, 2012. **56**(6): p. 2328-35.