# Novel Risk Prediction Models for Post-Operative Mortality in Patients with Cirrhosis

**Nadim Mahmud, MD MS MPH MSCE**[1,2,*], **Zachary Fricker, MD**[3,*], **Rebecca A. Hubbard, PhD**[4], **George N. Ioannou, MD MS**[5,6], **James D. Lewis, MD MSCE**[1,2], **Tamar H. Taddei, MD**[7,8], **Kenneth D. Rothstein, MD**[1], **Marina Serper, MD MS**[1,9], **David S. Goldberg, MD MSCE**[10], **David E. Kaplan, MD MSc**[1,9]

[1.]Division of Gastroenterology and Hepatology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

[2.]Leonard David Institute of Health Economics, University of Pennsylvania, Philadelphia, PA

[3.]Division of Gastroenterology, Hepatology, and Nutrition, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA

[4.]Center for Clinical Epidemiology and Biostatistics, Department of Biostatistics, Epidemiology & Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA

[5.]Division of Gastroenterology, Department of Medicine, Veterans Affairs Puget Sound Healthcare System and University of Washington, Seattle, WA

[6.]Health Services Research and Development, Veterans Affairs Puget Sound Healthcare System, Seattle, WA

[7.]Division of Digestive Diseases, Yale University School of Medicine, New Haven, CT

[8.]VA Connecticut Healthcare System, West Haven, CT

[9.]Department of Medicine, Corporal Michael J. Crescenz VA Medical Center, Philadelphia, PA

[10.]Division of Digestive Health and Liver Diseases, Department of Medicine, University of Miami Miller School of Medicine, Miami, FL

## Abstract

**Background & Aims:** Patients with cirrhosis are at increased risk of post-operative mortality. Currently available tools to predict post-operative risk are suboptimally calibrated and do not

**Corresponding Author:** Nadim Mahmud MD MS MPH MSCE, 3400 Civic Center Boulevard, 4th Floor, South Pavilion, Philadelphia, PA 19104, Phone: 215-349-8222, Fax: 215-349-5915, nadim@pennmedicine.upenn.edu.

account for surgery type. Our objective was to use population-level data to derive and internally validate novel cirrhosis surgical risk models.

**Methods:** We conducted a retrospective cohort study using data from the Veterans Outcomes and Costs Associated with Liver Disease (VOCAL) cohort, which contains granular data on patients with cirrhosis from 128 United States medical centers, merged with the Veterans Affairs Surgical Quality Improvement Program (VASQIP) to identify surgical procedures. We categorized surgeries as abdominal wall, vascular, abdominal, cardiac, chest, or orthopedic, and used multivariable logistic regression to model 30, 90, and 180-day post-operative mortality (VOCAL-Penn models). We compared model discrimination and calibration of VOCAL-Penn to the Mayo risk score (MRS), MELD, MELD-Na, and Child-Turcotte-Pugh (CTP) scores.

**Results:** We identified 4,712 surgical procedures in 3,785 patients with cirrhosis. The novel VOCAL-Penn models were derived and internally validated with excellent discrimination (30-day post-operative mortality C-statistic=0.859, 95% confidence interval [CI] 0.809–0.909). Predictors included age, pre-operative albumin, platelet count, bilirubin, surgery category, emergency indication, fatty liver disease, American Society of Anesthesiologists classification, and obesity. Model performance was superior to MELD, MELD-Na, CTP, and MRS at all timepoints (e.g. 30-day post-operative mortality C-statistic for MRS=0.766, 95% CI 0.676–0.855) in terms of discrimination and calibration.

**Conclusion:** The VOCAL-Penn models substantially improve post-operative mortality predictions in patients with cirrhosis. These models may be applied in practice to improve pre-operative risk stratification and optimize patient selection for surgical procedures (www.vocalpennscore.com).

## Introduction

Patients with cirrhosis have increased surgical morbidity and mortality relative to the general population.[1–3] Due to the rising burden of cirrhosis,[4] as well as aging of patients with established liver disease,[5] the number of surgeries in patients with cirrhosis is expected to increase. Gastroenterologists, hepatologists, primary care providers, and surgeons face challenges in estimating surgical risk in patients with cirrhosis. While the National Surgery Quality Improvement Program (NSQIP) calculator is the most commonly used surgical risk prediction tool,[6] the presence of ascites is the only liver-specific parameter included and there is no variable documenting the presence of cirrhosis. Furthermore, this score has never been validated in patients with cirrhosis. The Child-Turcotte-Pugh (CTP) score and the Model for End-Stage Liver Disease (MELD) are frequently utilized,[7] but do not include surgery-specific risks. Currently, the only dedicated surgical risk prediction model validated in patients with cirrhosis is the Mayo risk score (MRS).[8, 9] Predictors in this score include the American Society of Anesthesiologists (ASA) physical status classification, international normalized ratio (INR), total bilirubin, creatinine, age, and etiology of liver disease. While the MRS is widely used in clinical practice, there is evidence that it is inadequately-

calibrated and over-estimates surgical risk.[10] It is also plausible that calibration has declined over time since its derivation in 2007, possibly due to advances in surgical and post-operative care as well as reductions in overall post-operative mortality.

Our group and others have demonstrated that the type of operation constitutes a major predictor of surgical risk in patients with cirrhosis.[3, 11] However, the MRS, which is designed to predict mortality for high-risk cardiovascular, orthopedic, and major abdominal surgeries only, does not stratify risk based upon the type of surgery. The lack of incorporation of the surgical procedure type as an input for MRS, CTP, and MELD-sodium (MELD-Na) could lead to overestimation of risk for minor or minimally-invasive surgeries.

The clinical utility of an updated, accurate risk score for patients with cirrhosis would be of broad interest to providers and patients trying to make informed, personalized decisions about surgery. To address this, we created a novel data merge between large medical and surgical databases from the Veterans Health Administration (VHA). We aimed to (1) evaluate the calibration of the MRS over time, (2) derive accurate prediction scores for cirrhosis surgical risk using advanced modeling methods and incorporating surgery type, (3) internally validate these scores in the VHA dataset, and (4) compare our models to existing clinical prediction models.

## Methods

### Study Design, Data Sources, and Cohort Creation

We performed a retrospective cohort study of patients with cirrhosis using well-phenotyped, longitudinal data from the Veterans Outcomes and Costs Associated with Liver Disease (VOCAL) cohort, which contains data from 128 VHA hospitals. The derivation of the VOCAL cohort has been previously described;[12] in brief, it contains medical data on over 129,000 patients with cirrhosis identified between 2008 and 2016, and has been used for numerous natural history studies of chronic liver disease.[13–19] We merged VOCAL with the Veterans Affairs Surgical Quality Improvement Program (VASQIP) dataset.[20] This dataset contains prospectively adjudicated data on VHA surgical procedures, including pre-operative, intraoperative, and post-operative data with validated reliability.[21] After the VOCAL-VASQIP data merge, we included all patients at least 18 years of age with cirrhosis, defined using a validated algorithm based on International Classification of Diseases (ICD) codes.[22] We excluded patients who did not receive a surgical procedure of interest (hepatic surgeries, minor surgeries, or those with accepted low risk; detailed below), those who received liver transplantation prior to surgery, and those who had a pre-operative ASA classification of 5 due to exceptional circumstances of these cases and associated very high morbidity and mortality. We also excluded patients with insufficient pre-operative laboratory data to compute the MELD-Na and MRS, using a window of 30 days prior to surgery. Finally, surgery categories in which fewer than 50 procedures were performed were excluded given limited statistical power to draw inferences associated with mortality; this resulted in the exclusion of central nervous system (CNS) surgeries.

## Ascertainment of Potential Predictors

We collected data on demographics (age, sex, race/ethnicity), tobacco use, hazardous alcohol use (ascertained using the Alcohol Use Disorders Identification Test [AUDIT-C]), body mass index (BMI), comorbidities (hypertension, diabetes, obesity, atrial fibrillation, coronary artery disease, congestive heart failure), and pre-operative laboratory data (sodium, creatinine, aspartate aminotransferase, alanine aminotransferase, total bilirubin, albumin, platelet count, INR). Pre-operative laboratory values closest to the day of surgery were used, with a maximum window of 30 days prior to. Comorbidities were ascertained using previously described methods.[12, 17] In particular, obesity was defined using a validated VHA algorithm to identify patients with BMI 30.[23] Etiology of liver disease was also ascertained using a validated algorithm,[24] and subsequently categorized as hepatitis C virus (HCV), hepatitis B virus (HBV), alcoholic liver disease (ALD), HCV & ALD, non-alcoholic fatty liver disease (NAFLD), or other. MELD and MELD-Na were computed from pre-operative laboratory parameters, and pre-operative CTP class was computed using a validated VHA algorithm.[12] This included separate classification of ascites (absent, slight, moderate) as well as hepatic encephalopathy (none, grade 1–2, grade 3–4) prior to surgery, which were tested as separate variables in models in addition to the CTP score. History of decompensated liver disease was determined using a validated VHA algorithm.[22] We then used common procedure terminology (CPT) codes to categorize surgical procedure type as follows: abdominal wall, vascular, major orthopedic, chest/cardiac, abdominal – laparoscopic, abdominal – open, or central nervous system (CNS; see Supplemental Table 1 for examples and coding details). These categories were decided based on guidance from the literature, as well as through expert opinion. As noted above, CNS surgeries were ultimately excluded due to low total sample size of these procedures (n=44). ASA classification was determined pre-operatively by anesthesiologist assessment, as adjudicated in VASQIP. An emergency modifier was denoted in cases where delay in surgery would lead to a significant increase in likelihood of death or loss of body part.[25] We evaluated this variable in two ways. In the first, we left ASA as an ordinal variable, where values included 2, 3, or 4 as assigned in VASQIP (no patients had an ASA class of 1). In the second, we reclassified ASA as 4 if the patients had decompensated cirrhosis, and otherwise classified ASA as 3 to represent compensated cirrhosis (i.e. a binary classification). This approach is consistent with the MRS, which dichotomized ASA in this fashion.[9] We also classified surgeries as emergent or non-emergent, and where relevant, as laparoscopic or open. Finally, we classified and confirmed post-operative death at 30, 90, and 180 days using both VASQIP and the VA Vital Status File.[26]

## Descriptive Analysis

Baseline pre-operative characteristics were reported as percentages for categorical variables, and median and interquartile ranges (IQR) for continuous variables. Post-operative mortality at 30, 90, and 180 days was presented as stratified by surgery type and ASA classification (ordinal). Log-rank testing and Kaplan-Meier analysis were used to evaluate the association between ASA classification, surgery category, obesity, and NAFLD and post-operative mortality. As a surrogate for portal hypertension, we evaluated the association between ASA classification and pre-operative platelet count using the Kruskal-Wallis test.

### Computation of the Mayo Score

We referenced the models in the seminal Mayo cirrhosis surgical risk study and the associated website to calculate the MRS,[9, 27] using the binary classification of ASA as detailed above. We confirmed accurate coding of the risk scores through direct comparison with values generated from the MRS website.

### Evaluating the Calibration of the Mayo Score over Time

To test our hypothesis of changing calibration of the MRS over time, we first divided surgeries in the cohort into three-year intervals (2008–2010, 2011–2013, 2014–2016, 2017–2019). We then plotted calibration curves of predicted versus observed post-operative mortality at 30 and 90 days for each time period. We evaluated calibration by visual inspection of the curves in reference to a 45-degree line (which would represent perfect calibration). To formally test for changing calibration over time, we created deciles of predicted risk by the MRS, and computed the difference between the sum of these predictions and the observed mortality proportion for each decile. We plotted the sum of differences over time, and fit a linear regression model to test the hypothesis that the beta coefficient was equal to zero. Noting that the original MRS was derived from patients who received non-laparoscopic abdominal, major cardiovascular (i.e., coronary artery bypass graft or valve replacement), or major orthopedic surgery (e.g., total knee or total hip replacement),[9] we repeated the above analyses in a sub-cohort restricted to these surgeries.

### Derivation of the VOCAL-Penn Cirrhosis Surgical Risk Models

We randomly divided the analytic cohort into a derivation (80%) and validation (20%) cohort using a computerized random number generator. In the derivation cohort, we first performed univariate analyses to evaluate the association between each exposure of interest and binary outcome of post-operative mortality at 30, 90, or 180 days. For categorical variables, we tabulated values against the outcome, and conducted chi-squared tests to identify relevant differences among groups (regarding a p-value of 0.05 to be statistically significant). Based on this analysis, we chose to simplify etiology of liver disease to NAFLD versus non-NAFLD, as NAFLD appeared to be uniquely associated with post-operative mortality relative to other etiologies. For each continuous variable, we plotted locally weighted scatterplot smoothing (LOWESS) curves against the outcome, and for variables observed to have a non-linear association we used restricted cubic splines in logistic regression models (see Supplemental Table 2 for knot specifications; see example in Supplemental Figure 1). We then proceeded with 10-fold cross validation and least absolute shrinkage and selection operator (LASSO) regression to identify a parsimonious base logistic regression model for 30-day post-operative mortality.[28] All variables listed in Table 1 were evaluated as possible predictors, in addition to laparoscopic versus open surgical approach. The base regression model was determined by identifying the value for the LASSO penalty term that minimized the mean squared predicted error (MSPE),[29] which was obtained through the 10-fold cross validation procedure. Multiple modified clinician-driven models were then tested, where variables felt to be clinically meaningful were reintroduced, and minimized Aikake Information Criterion (AIC) and Bayesian Information Criterion (BIC) were used to select a final model.[30] Among similarly-performing models,

parsimony (i.e. fewer model inputs) was also considered. The above process was repeated for 90 and 180 days post-operative mortality, restricted to the subset of patients surviving at least 30 and at least 90 days, respectively. Predicted probabilities from these models conditional on survival up to an earlier time point were then aggregated to obtain marginal probabilities of surviving 30, 90, or 180 days, using the law of total probability. Of note, we chose logistic regression as opposed to a time-to-event analysis given the *a priori* interest in post-operative mortality at specified time points. However, the above approach does represent a discrete time survival model with discrete time periods of 30, 90, and 180 days.[31]

### Evaluation and Internal Validation of the VOCAL-Penn Risk Models

In derivation and validation cohorts, we plotted receiver operating characteristic (ROC) curves for the VOCAL-Penn model at each time point and compared them to the MELD, MELD-Na, and MRS. To evaluate model discrimination, we computed C-statistics (area under the curve) and 95% confidence intervals (CIs) for each model, and compared C-statistics using DeLong's non-parametric method.[32] To evaluate model calibration, we plotted observed versus predicted probability of mortality, consistent with recommended best practices.[33] Finally, we performed two sensitivity analyses. First, given potential concerns regarding model generalizability from a male-predominant VHA dataset, we performed a dedicated analysis evaluating model discrimination only among females. Second, given the possibility that separate surgeries in the same patient may not reflect independent risks, we tested model discrimination only among unique patients, in which only first surgeries were considered.

### Other Considerations

Missingness at the level of each variable in the analytic cohort was <3.5%, and there was no association between data missingness and post-operative mortality at 30, 90, or 180 days (each p>0.05). As such, complete case analysis was performed and imputation was not pursued. Regarding ethical considerations, this study received Institutional Review Board approval at the Corporal Michael J. Crescenz Veterans Affairs Medical Center, Philadelphia, the VA Connecticut Healthcare System, and the Hospital of the University of Pennsylvania. All data management and analyses were performed using a combination of structured query language and Stata 15.1/IC (College Station, TX).

## Results

### Cohort Characteristics

After application of selection criteria (Supplemental Figure 2), the analytic cohort included 3,785 unique patients who underwent 4,712 surgeries (Table 1). The cohort was predominantly male (97.2%) and white (63.3%), with median age 64 (IQR 60–69). The most common surgeries were abdominal wall (27.8%) and major orthopedic (27.5%), and the least common were chest/cardiac (8.8%). Pre-operative laboratory data were obtained from median 0 days prior to surgery (IQR 0–5 days). Most patients were CTP A (88.3%) or CTP B (11.2%), with etiology of liver disease primarily HCV (13.0%), ALD (35.3%), HCV & ALD (29.5%), or NAFLD (12.4%). Post-operative mortality was higher among patients with ASA 4 versus ASA 3 or ASA 2 at 30, 90, and 180 days, with significant variation across

surgery type (each p<0.001; Table 2, Figure 1A/B). For example, open abdominal surgeries had a 15.2% 180-day mortality versus 5.2% for abdominal wall surgeries. Increasing ASA classification was also found to be inversely associated with pre-operative platelet count (p<0.001; Supplemental Figure 3). Finally, non-obese patients (BMI<30) and those with NAFLD as an etiology of liver disease had higher post-operative mortality relative to obese and non-NAFLD patients, respectively (each p<0.05; Figure 1C/D).

### Changing Calibration of the Mayo Surgical Risk Score

When considering all surgery types, we found declining calibration of the MRS over time by visual inspection, for both 30- and 90-day post-operative mortality predictions (Figure 2A/B). There was a significant rise in the sum of differences between predicted and observed mortality over time (p<0.001, Figure 2C/D). When restricting the cohort to only major cardiovascular, major orthopedic, and major digestive surgeries, the calibration of the MRS was somewhat improved by visual inspection from 2008–2010, however this again declined substantially over time (Supplemental Figure 4A/B), with corresponding significant rise in sum of differences over time (p<0.001, Supplemental Figure 4C/D). In all cases, declining calibration resulted from overprediction of post-operative mortality relative to observed mortality.

### Post-operative Mortality Prediction Models

The derivation (80%) and validation (20%) cohorts were similar across the full range of demographic, laboratory, and comorbidity variables (Table 1). In the final adjusted regression models, ASA classification (ordinal) was an important predictor of post-operative mortality at 30, 90, and 180 days (Table 3). Additional predictors associated with post-operative mortality included emergency surgery indication (OR 2.53 for 30-day mortality, 95% CI 1.54–4.16, p<0.001). NAFLD cirrhosis (OR 2.29 for 30-day mortality, 95% CI 1.37–3.82, p=0.002), hyperbilirubinemia, hypoalbuminemia, and older age. Obesity (BMI 30) was protective against post-operative mortality (30-day OR 0.47, 95% CI 0.30–0.74, p=0.001). Finally, surgery type was also significantly associated with post-operative mortality (30-day joint p=0.012). For example, at 30 days, chest/cardiac surgeries had a 6.64-fold increased odds of death (95% CI 1.20–36.64) relative to laparoscopic abdominal surgeries. Importantly, we found that ASA classification treated as an ordinal variable yielded superior model performance as compared to ASA treated as a binary variable (see examples in Supplemental Table 3). Additionally, we found ascites (absent, slight, moderate) and platelet count to perform near interchangeably in multivariable models, yielding similar AIC, BIC, and discrimination estimates when exchanged. Furthermore, the addition of ascites to models already containing platelet count did not improve model performance (Supplemental Table 3). Given the subjectivity associated with ascites assessment, we opted for final models containing platelet count rather than ascites, where applicable. Finally, multiple models substituting cardiovascular risk factors in place of NAFLD did not improve model performance (Supplemental Table 3).

At all timepoints in both derivation and validation cohorts, the VOCAL-Penn models had improved discrimination relative to the MELD, MELD-Na, and CTP scores, and at 30 and 90 days had improved discrimination relative to the MRS (Figures 3A/B/D/E, 4A/B [p-

values noted in figures]; Table 4). For example, in the validation cohort at 30 days, the VOCAL-Penn c-statistic was 0.859 versus 0.766 for the MRS, and 0.752 for MELD-Na (p=0.003). The calibration of the VOCAL-Penn models was also excellent at each timepoint (Figures 3C/F, 4C), whereas the MRS overestimated post-operative mortality across the entire spectrum of risk at both 30 and 90 days. In a restricted analysis where the VOCAL-Penn models were applied only to females, the model discrimination for post-operative mortality remained excellent (each C-statistic >0.89) at all evaluable timepoints (Supplemental Table 4). Similar results were also found when restricting the cohort to first surgeries among unique patients (each C-statistic >0.81; Supplemental Table 5)

### Summary of the VOCAL-Penn Prediction Tool

Based on the models detailed above, the final prediction tool consists of nine variables to be inputted by the user: age, ASA classification (2 = mild systemic disease, 3 = severe systemic disease, 4 = severe systemic disease that is a constant threat to life), surgery category (abdominal – laparoscopic, abdominal – open, abdominal wall, vascular, major orthopedic, and chest/cardiac), emergency surgery indication (yes/no), NAFLD as etiology of cirrhosis (yes/no), albumin, platelet count, total bilirubin, and obesity (BMI 30; yes/no). The output of the prediction tool is post-operative predicted mortality at 30, 90, and 180 days. Derivation of model formulas and predicted probabilities is detailed in the Supplement Appendix.

## Discussion

In this study of 4,712 diverse surgeries performed on patients with cirrhosis, we demonstrate the need for improved pre-operative risk stratification, and subsequently report the creation of a highly accurate post-operative mortality prediction models. The first major finding in this study is the significant decline in the calibration of the MRS over time. In particular, poor calibration manifested as significant *overprediction* of mortality risk, often for patients with very low observed post-operative mortality. This pattern persisted even when isolating the cohort to the surgical procedures from which the MRS was originally derived.[9] Importantly, this finding is consistent with the only prior study to externally validate the MRS. Among 160 patients with cirrhosis who underwent surgery with general anesthesia, Kim et al found that while the discrimination of the score was excellent (c-statistic=0.803 at 90 days), the MRS consistently overpredicted mortality risk (median predicted 1-year mortality 22.6% versus 8.9% observed).[10] The reasons for declining MRS calibration over time are likely complex, however it is important to note that the MRS predates numerous innovations in the surgical field, such as the introduction of advanced endovascular techniques and transcatheter-based valve replacements.[34–36] These developments have fundamentally changed the risk profile of patients considered for open procedures. Changes in the demographics of patients with cirrhosis,[37] the ability to cure hepatitis C,[38] and improved peri-operative cirrhosis-related care have likely altered the dynamics of surgical selection and post-operative outcomes as well. Finally, as the MRS was derived from single-center data, it is possible that local practices impacted model calibration.

The declining calibration of the MRS firmly establishes the imperative to develop improved risk models such as those presented here. We found that surgery type is a critical predictor, where major chest and major cardiac surgeries confer particularly high risk, and abdominal wall surgeries (e.g. hernia repairs) are much lower risk. Markers of worsening intrinsic liver function, including thrombocytopenia, hypoalbuminemia, and hyperbilirubinemia, were all significant predictors of mortality in our models. In the course of modeling, it is important to note that platelet count was near interchangeable with ascites category, implying that the utility of platelet count in models is as a surrogate for portal hypertension, a known predictor of cirrhosis surgical risk.[39, 40] Indeed, a recent study by Reverter et al demonstrated that pre-operative hepatic venous pressure gradient was a significant independent predictor of post-operative mortality in a cohort comprised of various extrahepatic surgeries.[41] Our study suggests that portal hypertension may also be partially captured in ASA classification, which we found to be associated with pre-operative platelet count. Interestingly, although NAFLD cirrhosis was positively associated with post-operative mortality, obesity (BMI 30) was protective. This may be on the basis of NAFLD serving as a surrogate for increased cardiovascular risk,[42, 43] whereas the protective effect of obesity is a reflection of the "obesity paradox," whereby patients with cirrhosis who have increased nutritional reserve fare better after an acute stressor.[44] There may be independent effects of NAFLD separate from cardiovascular risk, however, as we found that NAFLD provided more predictive information in models as compared to classic cardiovascular risk factors.

There are several important strengths to this work. First, given the volume and breadth of surgeries under study, this multi-center cohort allowed for creation of broadly applicable cirrhosis surgical risk models, thus minimizing the influence of center-specific idiosyncrasies. Second, the VHA cohort represents diverse etiologies of liver disease, similar to those encountered in clinical practice. Third, use of modern inferential methods such as LASSO and restricted cubic splines allowed for development of parsimonious models with high-fidelity calibration across the spectrum of risk, a major improvement over previous efforts. Finally, our models rely on readily available laboratory predictors and standard comorbid pre-operative parameters, and thus predicted probabilities can easily be obtained. To make this readily accessible for clinical practice, an online risk calculator using the VOCAL-Penn models will be made available for public use.

The clinical implications of this work are significant. Historically, over-prediction of cirrhosis surgical risk has led to undue denial of surgery for patients with an otherwise acceptable risk profile. In the case of elective procedures such as hernia repairs, this may result in avoidance of surgery until the indication becomes emergent (i.e. incarceration), when the surgical risk is uniformly higher.[45–47] This risk avoidance is a documented phenomenon with perverse outcomes,[48] which may be ameliorated with improved risk prediction. Indeed, careful pre-operative risk assessments are routine for any patient with cirrhosis, and the VOCAL-Penn models represent a modern, updated tool available for discrete post-operative mortality predictions.

There are several important limitations that we acknowledge in this study. As with any large dataset, there is the possibility of misclassification of exposures and outcomes. However, VASQIP is manually-adjudicated, and validated algorithms were used wherever possible to

minimize misclassification.[21] Second, the VHA dataset may not generalize well to all populations. The VHA cohort is largely male, and reflects a higher burden of psychosocial comorbidities relative to the general population.[49, 50] However, this cohort does reflect diverse etiologies of liver disease, and it is not clear that unique differences in this cohort would significantly impact surgical risk. For example, no prior literature has demonstrated differences in cirrhosis surgical risk on the basis of sex. Furthermore, in a stratified analysis limited to females, we found that the VOCAL-Penn models retained excellent discrimination at all timepoints. Third, in this study we do not address potentially modifiable factors such as pre-operative transjugular intrahepatic portosystemic shunt or medication use such as non-selective beta blockade. These factors may impact cirrhosis surgical risk, and warrant future study in this context. Fourth, the nature of the data available obviously does not include patients with cirrhosis who had an indication for surgery but did not receive it, potentially due to perception of perioperative risk. This helps to explain the skew towards lower MELD and CTP A patients in the cohort, as presumably patients with severe derangements in MELD parameters, for example, rarely proceeded to surgery. However, there may be additional variables such as sarcopenia and general disability that influence surgical decision making for otherwise apparently "low risk" patients that are not captured in the VOCAL-Penn models. Thus it is critical to highlight that the VOCAL-Penn models are not intended to substitute for clinical judgment, but rather to be used as adjunctive tools in risk prognostication for discussions between patients and clinicians. In particular, the models should not be applied until after a standard pre-operative clinical assessment, which historically includes an assessment of liver disease severity through MELD and decompensation status. Fifth, related to the previous point, it is possible that overestimation of risk from the MRS has impacted surgical decisions such that higher-risk surgeries were not performed. We unfortunately do not have specific information as to how patients were selected for surgery, however this was likely based on clinical grounds and use of existing tools such as MELD-Na and the MRS. It is therefore unclear if the VOCAL-Penn models will remain well-calibrated as an expanded range of surgeries are attempted, and it is likely that periodic recalibration will be required. Finally, although this study presents compelling multicenter internal validation data, the VOCAL-Penn risk models have not been externally validated in an independent cohort. This is an important area of future study.

In conclusion, the VOCAL-Penn risk models improve significantly over existing clinical standards, such as the MRS, to predict 30, 90, and 180 day mortality for patients with cirrhosis who undergo surgery. These models will be made publicly available in the form of an online calculator for clinical use (www.vocalpennscore.com). Further research will define the external validity of the models and the boundaries of their applicability.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

Author Manuscript

## Abbreviations:

| | |
|---|---|
| **NSQIP** | National Surgery Quality Improvement Program |
| **CTP** | Child-Turcotte-Pugh |
| **MELD** | model for end-stage liver disease |
| **MRS** | Mayo risk score |
| **ASA** | American Society of Anesthesiologists |
| **INR** | international normalized ratio |
| **MELD-Na** | model for end-stage liver disease-sodium |
| **VHA** | Veterans Health Administration |
| **VOCAL** | Veterans Outcomes and Costs Associated with Liver Disease |
| **VASQIP** | Veterans Affairs Surgical Quality Improvement Program |
| **ICD** | International Classification of Diseases |
| **CNS** | central nervous system |
| **BMI** | body mass index |
| **HCV** | hepatitis C virus |
| **HBV** | hepatitis B virus |
| **ALD** | alcoholic liver disease |
| **NAFLD** | non-alcoholic fatty liver disease |
| **CPT** | current procedure terminology |
| **IQR** | interquartile range |
| **LOWESS** | locally weighted scatterplot smoothing |

| **LASSO** | least absolute shrinkage and selection operator |
| **MSPE** | mean squared predicted error |
| **AIC** | Aikake Information Criterion |
| **BIC** | Bayesian Information Criterion |
| **ROC** | receiver operating characteristic |
| **CI** | confidence interval |

## References

1. Friedman LS. The risk of surgery in patients with liver disease. Hepatology 1999;29:1617–1623. [PubMed: 10347099]

2. Goel NJ, Agarwal P, Mallela AN, et al. Liver disease is an independent predictor of poor 30-day outcomes following surgery for degenerative disease of the cervical spine. The Spine Journal 2019;19:448–460. [PubMed: 30053522]

3. Johnson KM, Newman KL, Green PK, et al. Incidence and Risk Factors of Postoperative Mortality and Morbidity After Elective Versus Emergent Abdominal Surgery in a National Sample of 8193 Patients With Cirrhosis. Annals of surgery 2019.

4. Parikh ND, Marrero WJ, Wang J, et al. Projected increase in obesity and non-alcoholic-steatohepatitis–related liver transplantation waitlist additions in the United States. Hepatology 2019;70:487–495. [PubMed: 28833326]

5. Peery AF, Crockett SD, Murphy CC, et al. Burden and cost of gastrointestinal, liver, and pancreatic diseases in the United States: update 2018. Gastroenterology 2018.

6. Bilimoria KY, Liu Y, Paruch JL, et al. Development and evaluation of the universal ACS NSQIP surgical risk calculator: a decision aid and informed consent tool for patients and surgeons. Journal of the American College of Surgeons 2013;217:833–842. e3. [PubMed: 24055383]

7. Hoteit MA, Ghazale AH, Bain AJ, et al. Model for end-stage liver disease score versus Child score in predicting the outcome of surgical procedures in patients with cirrhosis. World journal of gastroenterology: WJG 2008;14:1774. [PubMed: 18350609]

8. Northup PG, Friedman LS, Kamath PS. AGA clinical practice update on surgical risk assessment and perioperative management in cirrhosis: expert review. Clinical Gastroenterology and Hepatology 2019;17:595–606. [PubMed: 30273751]

9. Teh SH, Nagorney DM, Stevens SR, et al. Risk factors for mortality after surgery in patients with cirrhosis. Gastroenterology 2007;132:1261–1269. [PubMed: 17408652]

10. Kim SY, Yim HJ, Park SM, et al. Validation of a Mayo post-operative mortality risk prediction model in Korean cirrhotic patients. Liver International 2011;31:222–228. [PubMed: 21134111]

11. Mahmud N, Fricker Z, Serper M, et al. In-Hospital Mortality Varies by Procedure Type among Cirrhosis Surgery Admissions. Liver International 2019.

12. Kaplan DE, Dai F, Aytaman A, et al. Development and performance of an algorithm to estimate the Child-Turcotte-Pugh score from a national electronic healthcare database. Clinical Gastroenterology and Hepatology 2015;13:2333–2341. e6. [PubMed: 26188137]

13. Mahmud N, Hubbard RA, Kaplan DE, et al. Declining Cirrhosis Hospitalizations in the Wake of the COVID-19 Pandemic: A National Cohort Study. Gastroenterology 2020.

14. Serper M, Taddei TH, Mehta R, et al. Association of provider specialty and multidisciplinary care with hepatocellular carcinoma treatment and mortality. Gastroenterology 2017;152:1954–1964. [PubMed: 28283421]

15. Goldberg DS, Taddei TH, Serper M, et al. Identifying barriers to hepatocellular carcinoma surveillance in a national sample of patients with cirrhosis. Hepatology 2017;65:864–874. [PubMed: 27531119]

16. Goldberg DS, French B, Forde KA, et al. Association of distance from a transplant center with access to waitlist placement, receipt of liver transplantation, and survival among US veterans. Jama 2014;311:1234–1243. [PubMed: 24668105]

17. Mahmud N, Kaplan DE, Taddei TH, et al. Incidence and Mortality of Acute-on-Chronic Liver Failure Using Two Definitions in Patients with Compensated Cirrhosis. Hepatology 2019;69:2150–2163. [PubMed: 30615211]

18. Kanneganti M, Mahmud N, Kaplan DE, et al. Survival Benefit of Liver Transplantation for Hepatocellular Carcinoma. Transplantation 2019.

19. Xiao KY, Hubbard RA, Kaplan DE, et al. Models for acute on chronic liver failure development and mortality in a veterans affairs cohort. Hepatology International 2020.

20. Best WR, Khuri SF, Phelan M, et al. Identifying patient preoperative risk factors and postoperative adverse events in administrative databases: results from the Department of Veterans Affairs National Surgical Quality Improvement Program. Journal of the American College of Surgeons 2002;194:257–266. [PubMed: 11893128]

21. Davis CL, Pierce JR, Henderson W, et al. Assessment of the reliability of data collected for the Department of Veterans Affairs national surgical quality improvement program. Journal of the American College of Surgeons 2007;204:550–560. [PubMed: 17382213]

22. Re III VL, Lim JK, Goetz MB, et al. Validity of diagnostic codes and liver-related laboratory abnormalities to identify hepatic decompensation events in the Veterans Aging Cohort Study. Pharmacoepidemiology and drug safety 2011;20:689–699. [PubMed: 21626605]

23. Husain N, Blais P, Kramer J, et al. Nonalcoholic fatty liver disease (NAFLD) in the Veterans Administration population: development and validation of an algorithm for NAFLD using automated data. Alimentary pharmacology & therapeutics 2014;40:949–954. [PubMed: 25155259]

24. Beste LA, Leipertz SL, Green PK, et al. Trends in burden of cirrhosis and hepatocellular carcinoma by underlying liver disease in US veterans, 2001–2013. Gastroenterology 2015;149:1471–1482. e5. [PubMed: 26255044]

25. ASo A. ASA physical status classification system. ASA House of Delegates 2014.

26. Sohn M-W, Arnold N, Maynard C, et al. Accuracy and completeness of mortality data in the Department of Veterans Affairs. Population health metrics 2006;4:2. [PubMed: 16606453]

27. Post-operative Mortality Risk in Patients with Cirrhosis. Volume 2019: Mayo Clinic.

28. McNeish DM. Using lasso for predictor selection and to assuage overfitting: A method long overlooked in behavioral sciences. Multivariate Behavioral Research 2015;50:471–484. [PubMed: 26610247]

29. Hastie T, Tibshirani R, Friedman J, et al. The elements of statistical learning: data mining, inference and prediction. The Mathematical Intelligencer 2005;27:83–85.

30. Schwarz G Estimating the dimension of a model. The annals of statistics 1978;6:461–464.

31. Cox D, Oakes D. Analysis of Survival Data. Boca Raton, LA: CRC Press LCC/Chapman & Hall, 1984.

32. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988:837–845. [PubMed: 3203132]

33. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. Epidemiology (Cambridge, Mass.) 2010;21:128.

34. Murashita T, Matsuda H, Domae K, et al. Less invasive surgical treatment for aortic arch aneurysms in high-risk patients: a comparative study of hybrid thoracic endovascular aortic repair and conventional total arch replacement. The Journal of thoracic and cardiovascular surgery 2012;143:1007–1013. [PubMed: 21783209]

35. Canaud L, Hireche K, Berthet J-P, et al. Endovascular repair of aortic arch lesions in high-risk patients or after previous aortic surgery: midterm results. The Journal of thoracic and cardiovascular surgery 2010;140:52–58. [PubMed: 19919865]

36. Mack MJ, Leon MB, Smith CR, et al. 5-year outcomes of transcatheter aortic valve replacement or surgical aortic valve replacement for high surgical risk patients with aortic stenosis (PARTNER 1): a randomised controlled trial. The Lancet 2015;385:2477–2484.

37. Moon AM, Singal AG, Tapper EB. Contemporary Epidemiology of Chronic Liver Disease and Cirrhosis. Clinical Gastroenterology and Hepatology 2019.

38. Kim D, Li AA, Perumpail BJ, et al. Changing Trends in Etiology-Based and Ethnicity-Based Annual Mortality Rates of Cirrhosis and Hepatocellular Carcinoma in the United States. Hepatology 2019;69:1064–1074. [PubMed: 30014489]

39. Berzigotti A, Reig M, Abraldes JG, et al. Portal hypertension and the outcome of surgery for hepatocellular carcinoma in compensated cirrhosis: a systematic review and meta-analysis. Hepatology 2015;61:526–536. [PubMed: 25212123]

40. Maithel SK, Kneuertz PJ, Kooby DA, et al. Importance of low preoperative platelet count in selecting patients for resection of hepatocellular carcinoma: a multi-institutional analysis. Journal of the American College of Surgeons 2011;212:638–648. [PubMed: 21463803]

41. Reverter E, Cirera I, Albillos A, et al. The prognostic role of hepatic venous pressure gradient in cirrhotic patients undergoing elective extrahepatic surgery. Journal of hepatology 2019;71:942–950. [PubMed: 31330170]

42. Bonora E, Targher G. Increased risk of cardiovascular disease and chronic kidney disease in NAFLD. Nature reviews Gastroenterology & hepatology 2012;9:372. [PubMed: 22565095]

43. Misra VL, Khashab M, Chalasani N. Nonalcoholic fatty liver disease and cardiovascular risk. Current gastroenterology reports 2009;11:50. [PubMed: 19166659]

44. Karagozian R, Bhardwaj G, Wakefield DB, et al. Obesity paradox in advanced liver disease: obesity is associated with lower mortality in hospitalized patients with cirrhosis. Liver international 2016;36:1450–1456. [PubMed: 27037497]

45. Gray SH, Vick CC, Graham LA, et al. Umbilical herniorrhapy in cirrhosis: improved outcomes with elective repair. Journal of Gastrointestinal Surgery 2008;12:675–681. [PubMed: 18270782]

46. O'Hara ET, Oliai A, Patek AJ Jr, et al. Management of umbilical hernias associated with hepatic cirrhosis and ascites. Annals of surgery 1975;181:85. [PubMed: 1119872]

47. Wong NZ, Mahmud N. The Imperative for an Updated Cirrhosis Surgical Risk Score. Annals of Hepatology 2020.

48. Choi SB, Hong KD, Lee JS, et al. Management of umbilical hernia complicated with liver cirrhosis: an advocate of early and elective herniorrhaphy. Digestive and Liver Disease 2011;43:991–995. [PubMed: 21872542]

49. Fortney JC, Curran GM, Hunt JB, et al. Prevalence of probable mental disorders and help-seeking behaviors among veteran and non-veteran community college students. General hospital psychiatry 2016;38:99–104. [PubMed: 26598288]

50. Kazis LE, Miller DR, Clark J, et al. Health-related quality of life in patients served by the Department of Veterans Affairs: results from the Veterans Health Study. Archives of internal medicine 1998;158:626–632. [PubMed: 9521227]
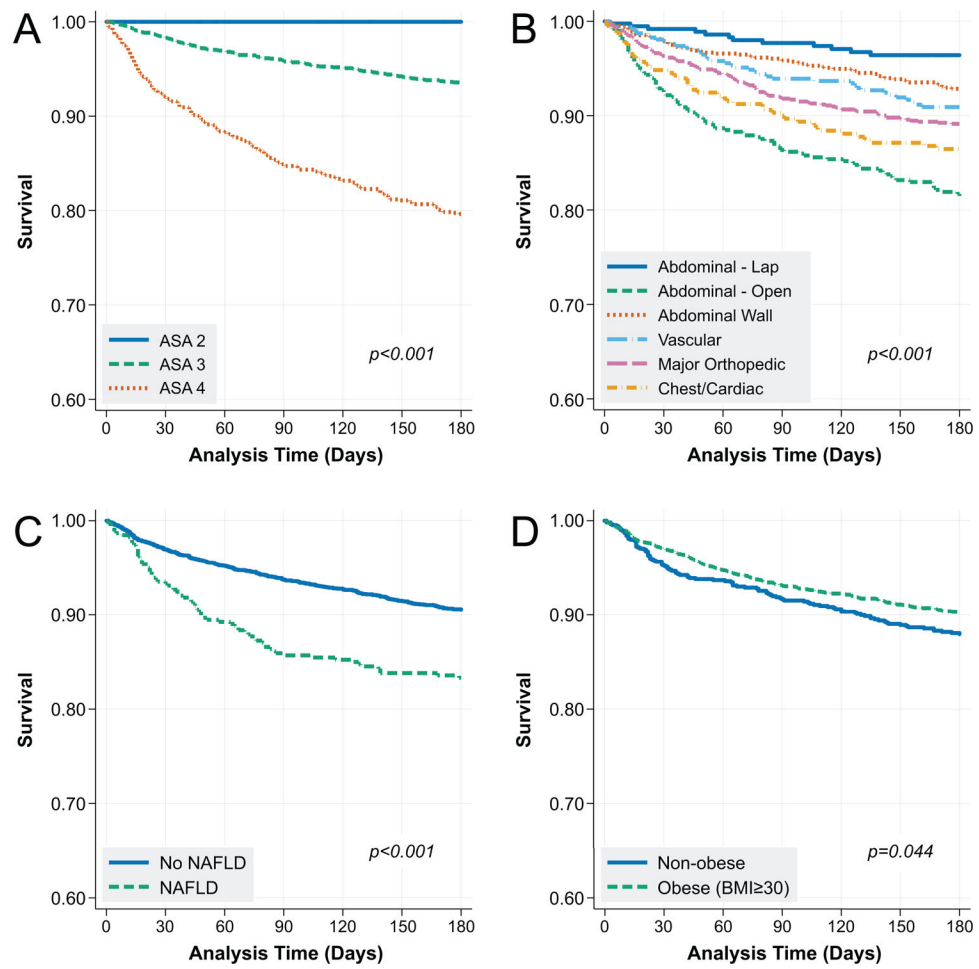
**Figure 1:**

Kaplan-Meier Analysis of the Association between (A) ASA Classification, (B) Surgery Category, (C) Non-alcoholic Fatty Liver Disease, and (D) Obesity on Post-operative Mortality

Abbreviations: ASA = American Society of Anesthesiologists; NAFLD = non-alcoholic fatty liver disease; BMI = body mass index
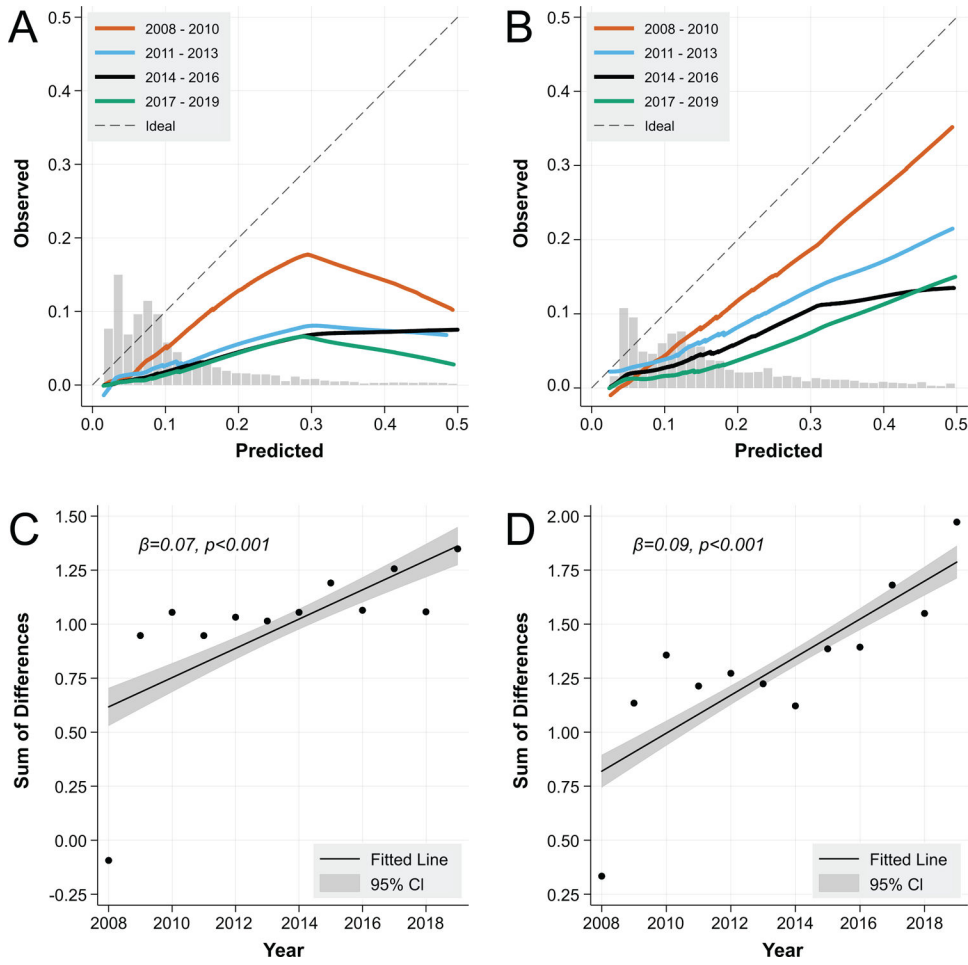
**Figure 2:**
Post-operative Mortality Calibration of the Mayo Score over Time at 30 Days (A) and 90 Days (B), and Sum of Differences between Predicted and Observed Mortality over Time at 30 Days (C) and 90 Days (D)

Caption: * Overlaid histograms indicate the distribution of contributing data points
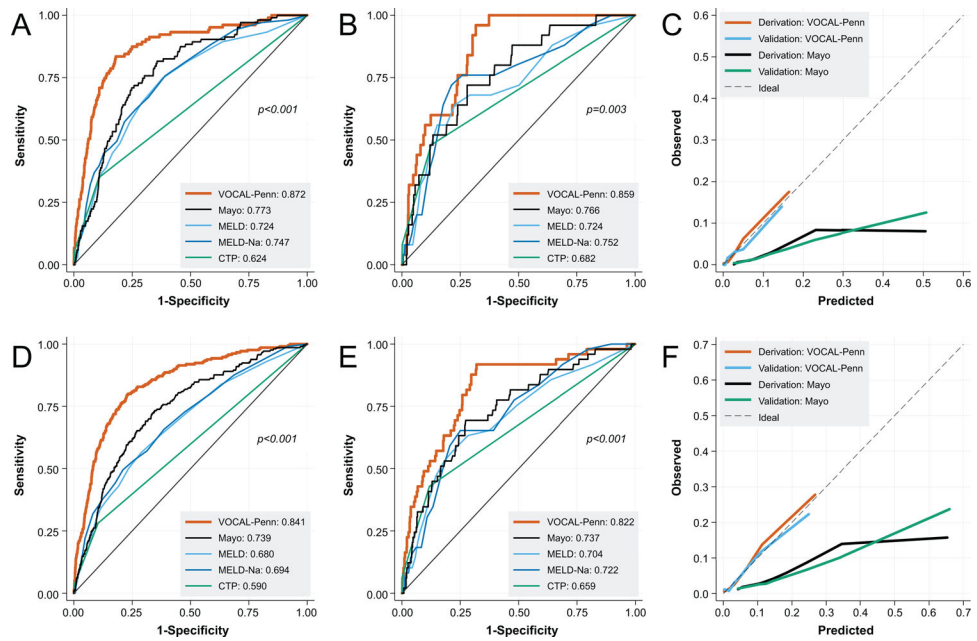
**Figure 3:**
Receiver Operating Characteristic Curves for 30-Day Post-Operative Mortality in Derivation (A) and Validation Cohorts (B), with Associated Calibration Curves (C), and for 90-Day Post-Operative Mortality in Derivation (D) and Validation Cohorts (E), with Associated Calibration Curves (F)

**Figure 4:**
Receiver Operating Characteristic Curves for 180-Day Post-Operative Mortality in
Derivation (A) and Validation Cohorts (B), with Associated Calibration Curves (C)

**Table 1:**

Cohort Characteristics

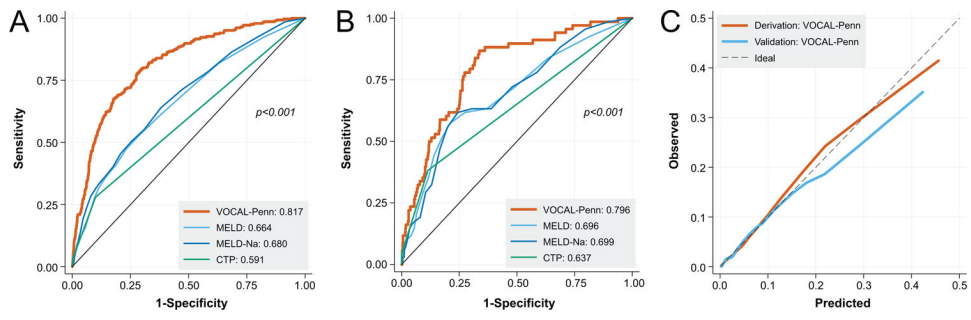| Variable | Overall (N = 4,712) | Derivation (N = 3,770) | Validation (N = 942) |
|---|---|---|---|
| **Age, median (IQR)** | 64 (60, 69) | 64 (60, 69) | 64 (60, 69) |
| **Male Sex** | 4582 (97.2%) | 3667 (97.3%) | 915 (97.1%) |
| **Race** | | | |
| White | 2981 (63.3%) | 2365 (62.7%) | 616 (65.4%) |
| Black | 741 (15.7%) | 604 (16.0%) | 137 (14.5%) |
| Hispanic | 323 (6.9%) | 250 (6.6%) | 73 (7.7%) |
| Asian | 51 (1.1%) | 43 (1.1%) | 8 (0.8%) |
| Other | 616 (13.1%) | 508 (13.5%) | 108 (11.5%) |
| **Smoking History** | | | |
| Never smoker | 846 (18.8%) | 676 (18.8%) | 170 (19.0%) |
| Former smoker | 979 (21.8%) | 768 (21.3%) | 211 (23.5%) |
| Current smoker | 2676 (59.5%) | 2160 (59.9%) | 516 (57.5%) |
| **Hazardous Alcohol Use** | 821 (18.0%) | 658 (18.1%( | 163 (18.0%) |
| **Surgery Category** | | | |
| Abdominal – Laparoscopic | 476 (10.1%) | 387 (10.3%) | 89 (9.4%) |
| Abdominal – Open | 665 (14.1%) | 531 (14.1%) | 134 (14.2%) |
| Abdominal Wall | 1308 (27.8%) | 1054 (28.0%) | 254 (27.0%) |
| Vascular | 550 (11.7%) | 442 (11.7%) | 108 (11.5%) |
| Major Orthopedic | 1298 (27.5%) | 1032 (27.4%) | 266 (28.2%) |
| Chest/Cardiac | 415 (8.8%) | 324 (8.6%) | 91 (9.7%) |
| **Emergency Surgery** | 476 (10.1%) | 388 (10.3%) | 88 (9.3%) |
| **ASA Classification (Ordinal)** | | | |
| 2 | 200 (4.2%) | 161 (4.3%) | 39 (4.1%) |
| 3 | 3196 (67.8%) | 2578 (68.4%) | 618 (65.6%) |
| 4 | 1316 (27.9%) | 1031 (27.3%) | 285 (30.3%) |
| **ASA Reclassified (Binary)** [*] | | | |
| 3 | 2147 (45.6%) | 1725 (45.8%) | 422 (44.8%) |
| 4 | 2565 (54.4%) | 2045 (54.2%) | 520 (55.2%) |
| **Sodium, median (IQR)** | 138 (136, 140) | 138 (136, 140) | 138 (136, 140) |
| **Creatinine, median (IQR)** | 1.0 (0.8, 1.2) | .96 (.8, 1.2) | 1 (.8, 1.2) |
| **AST, median (IQR)** | 32 (22.5, 50) | 32 (23, 50) | 32 (22.5, 48) |
| **ALT, median (IQR)** | 27 (18, 43) | 27 (18, 43) | 27 (18, 42) |
| **Total Bilirubin, median (IQR)** | 0.8 (0.5, 1.1) | .75 (.5, 1.1) | .75 (.5, 1.2) |
| **Albumin, median (IQR)** | 3.7 (3.2, 4.1) | 3.7 (3.2, 4.1) | 3.7 (3.15, 4.1) |
| **Platelet Count, median (IQR)** | 152 (107, 207) | 151 (106, 205.5) | 156 (112, 213) |
| **INR, median (IQR)** | 1.1 (1.0, 1.2) | 1.1 (1.02, 1.245) | 1.1 (1.02, 1.24) |

| Variable | Overall (N = 4,712) | Derivation (N = 3,770) | Validation (N = 942) |
|---|---|---|---|
| **MELD, median (IQR)** | 8 (7, 11) | 8 (7, 11) | 9 (7, 11) |
| **MELD-Na, median (IQR)** | 10 (8, 14) | 10 (8, 14) | 11 (8, 14) |
| **Child-Turcotte-Pugh Class** | | | |
| A | 4159 (88.3%) | 3344 (88.7%) | 815 (86.5%) |
| B | 530 (11.2%) | 406 (10.8%) | 124 (13.2%) |
| C | 23 (0.5%) | 20 (0.5%) | 3 (0.3%) |
| **Ascites Category** | | | |
| None | 4092 (86.8%) | 3272 (86.8%) | 820 (87.0%) |
| Slight | 469 (10.0%) | 389 (10.3%) | 80 (8.5%) |
| Moderate | 151 (3.2%) | 109 (2.9%) | 42 (4.5%) |
| **Hepatic Encephalopathy** | | | |
| No encephalopathy | 4577 (97.1%) | 3660 (97.1%) | 917 (97.3%) |
| Grade 1–2 | 122 (2.6%) | 98 (2.6%) | 24 (2.5%) |
| Grade 3–4 | 13 (0.3%) | 12 (0.3%) | 1 (0.1%) |
| **Etiology of Liver Disease** | | | |
| Hepatitis C | 612 (13.0%) | 497 (13.2%) | 115 (12.2%) |
| Hepatitis B | 73 (1.5%) | 58 (1.5%) | 15 (1.6%) |
| Alcohol | 1662 (35.3%) | 1318 (35.0%) | 344 (36.5%) |
| Hepatitis C + Alcohol | 1388 (29.5%) | 1110 (29.4%) | 278 (29.5%) |
| Fatty Liver Disease | 585 (12.4%) | 480 (12.7%) | 105 (11.1%) |
| Other | 392 (8.3%) | 307 (8.1%) | 85 (9.0%) |
| **History of Prior Decompensation** | 2066 (43.8%) | 1653 (43.8%) | 413 (43.8%) |
| **Hypertension** | 3904 (85.4%) | 3120 (85.2%) | 784 (86.3%) |
| **Diabetes** | 2355 (51.7%) | 1882 (51.6%) | 473 (52.2%) |
| **Obesity (body mass index ≥ 30)** | 3359 (73.8%) | 2704 (74.2%) | 655 (72.3%) |
| **Atrial Fibrillation** | 803 (17.0%) | 651 (17.3%) | 152 (16.1%) |
| **Coronary Artery Disease** | 1622 (34.4%) | 1291 (34.2%) | 331 (35.1%) |
| **Congestive Heart Failure** | 1239 (26.3%) | 985 (26.1%) | 254 (27.0%) |

Abbreviations: IQR = interquartile range; ASA = American Society of Anesthesiologists

*ASA reclassified refers to patients with prior decompensated cirrhosis being recategorized as ASA 4, and all others as ASA 3, consistent with the definition used in the Mayo Risk Score

**Table 2:**

Post-operative Mortality by Surgery Category and ASA (Ordinal) Classification

| Surgery Category and ASA Class | 30-day Mortality | 90-day Mortality | 180-day Mortality |
|---|---|---|---|
| **Abdominal – Laparoscopic (N = 476)** | | | |
| ASA 2 (n = 26) | 0 (0.00%) | 0 (0.00%) | 0 (0.00%) |
| ASA 3 (n = 389) | 2 (0.51%) | 7 (1.80%) | 9 (2.31%) |
| ASA 4 (n = 61) | 1 (1.64%) | 1 (1.64%) | 3 (4.92%) |
| *Overall* | *3 (0.63%)* | *8 (1.68%)* | *12 (2.52%)* |
| **Abdominal – Open (N = 665)** | | | |
| ASA 2 (n = 17) | 0 (0.00%) | 0 (0.00%) | 0 (0.00%) |
| ASA 3 (n = 424) | 15 (3.54%) | 28 (6.60%) | 39 (9.20%) |
| ASA 4 (n = 224) | 29 (12.95%) | 49 (21.88%) | 62 (27.68%) |
| *Overall* | *44 (6.62%)* | *77 (11.58%)* | *101 (15.19%)* |
| **Abdominal Wall (N = 1308)** | | | |
| ASA 2 (n = 103) | 0 (0.00%) | 0 (0.00%) | 0 (0.00%) |
| ASA 3 (n = 958) | 6 (0.63%) | 16 (1.67%) | 32 (3.34%) |
| ASA 4 (n = 247) | 15 (6.07%) | 25 (10.12%) | 36 (14.57%) |
| *Overall* | *21 (1.61%)* | *41 (3.13%)* | *68 (5.20%)* |
| **Vascular (N = 550)** | | | |
| ASA 2 (n = 1) | 0 (0.00%) | 0 (0.00%) | 0 (0.00%) |
| ASA 3 (n = 344) | 4 (1.16%) | 14 (4.07%) | 18 (5.23%) |
| ASA 4 (n = 205) | 6 (2.93%) | 13 (6.34%) | 21 (10.24%) |
| *Overall* | *10 (1.82%)* | *27 (4.91%)* | *39 (7.09%)* |
| **Major Orthopedic (N = 1298)** | | | |
| ASA 2 (n = 49) | 0 (0.00%) | 0 (0.00%) | 0 (0.00%) |
| ASA 3 (n = 920) | 14 (1.52%) | 34 (3.70%) | 46 (5.00%) |
| ASA 4 (n = 329) | 27 (8.21%) | 53 (16.11%) | 66 (20.06%) |
| *Overall* | *41 (3.16%)* | *87 (6.70%)* | *112 (8.63%)* |
| **Chest/Cardiac (N = 415)** | | | |
| ASA 2 (n = 4) | 0 (0.00%) | 0 (0.00%) | 0 (0.00%) |
| ASA 3 (n = 161) | 3 (1.86%) | 5 (3.11%) | 9 (5.59%) |
| ASA 4 (n = 250) | 17 (6.80%) | 31 (12.40%) | 38 (15.20%) |
| *Overall* | *20 (4.82%)* | *36 (8.67%)* | *47 (11.33%)* |

Abbreviations: ASA = American Society of Anesthesiologists

**Table 3:**

Multivariable Regression Models for 30, 90, and 180-day Post-Operative Mortality

| Variable | Odds Ratio | 95% Confidence Interval | P-value |
|---|---|---|---|
| *30-day Post-Operative Mortality Model* | | | |
| **ASA Class (Ordinal; 2, 3, or 4)** | 2.89 | (1.84 – 4.54) | <0.001 * |
| **Emergency Procedure** | 2.53 | (1.54 – 4.16) | <0.001 * |
| **Surgery Category** | | | 0.012 * (joint) |
| Abdominal – Laparoscopic | 1.0 (Ref) | | |
| Abdominal – Open | 4.76 | (0.89 – 25.54) | 0.069 |
| Abdominal Wall | 2.10 | (0.39 – 11.23) | 0.386 |
| Vascular | 2.50 | (0.42 – 14.95) | 0.315 |
| Major Orthopedic | 4.32 | (0.82 – 22.69) | 0.083 |
| Chest/Cardiac | 6.64 | (1.20 – 36.64) | 0.030 * |
| **Albumin (per 1 g/dL)**[‡] | | | <0.001 * |
| Spline 1 | 0.60 | (0.36 – 0.98) | |
| Spline 2 | 0.37 | (0.13 – 1.08) | |
| **Platelet Count (per 1,000/μL)**[‡] | | | 0.002 * |
| Spline 1 | 0.99 | (0.99 – 1.00) | |
| Spline 2 | 1.00 | (0.99 – 1.01) | |
| **Total Bilirubin (per 1 mg/dL)** | 1.16 | (1.06 – 1.26) | 0.001 * |
| **Etiology of Cirrhosis** | | | |
| **Non-NAFLD** | 1.0 (Ref) | | |
| **NAFLD** | 2.29 | (1.37 – 3.82) | 0.002 * |
| **Obesity (body mass index 30)** | 0.47 | (0.30 – 0.74) | 0.001 * |
| *90-day Post-Operative Mortality Model* | | | |
| **ASA Class (Ordinal; 2, 3, or 4)** | 1.99 | (1.33 – 2.99) | 0.001 * |
| **Emergency Procedure** | 1.87 | (1.12 – 3.12) | 0.017 * |
| **Surgery Category** | | | 0.015 * (joint) |
| Abdominal – Laparoscopic | 1.0 (Ref) | | |
| Abdominal – Open | 3.86 | (1.12 – 13.25) | 0.032 * |
| Abdominal Wall | 1.28 | (0.36 – 4.56) | 0.702 |
| Vascular | 2.87 | (0.81 – 10.17) | 0.102 |
| Major Orthopedic | 2.91 | (0.88 – 9.66) | 0.081 |
| Chest/Cardiac | 3.54 | (0.99 – 12.75) | 0.053 |
| **Age (per year)** | 3.86 | (1.12 – 13.25) | 0.008 * |
| **Albumin (per 1 g/dL)**[‡] | | | <0.001 * |
| Spline 1 | 0.62 | (0.38 – 1.00) | |

| Variable | Odds Ratio | 95% Confidence Interval | P-value |
|---|---|---|---|
| Spline 2 | 0.37 | (0.15 – 0.94) | |
| **Etiology of Cirrhosis** | | | |
| **Non-NAFLD** | 1.0 (Ref) | | |
| **NAFLD** | 2.29 | (1.44 – 3.63) | <0.001[*] |
| *180-day Post-Operative Mortality Model* | | | |
| **ASA Class (Ordinal; 2, 3, or 4)** | 2.34 | (1.46 – 3.74) | <0.001[*] |
| **Surgery Category** | | | 0.270[†] (joint) |
| Abdominal – Laparoscopic | 1.0 (Ref) | | |
| Abdominal – Open | 2.42 | (0.79 – 7.38) | 0.120 |
| Abdominal Wall | 1.22 | (0.40 – 3.68) | 0.726 |
| Vascular | 1.59 | (0.49 – 5.14) | 0.440 |
| Major Orthopedic | 1.13 | (0.37 – 3.44) | 0.829 |
| Chest/Cardiac | 1.54 | (0.45 – 5.31) | 0.494 |
| **Age (per year)** | 1.03 | (1.00 – 1.06) | 0.073[†] |
| **Platelet Count (per 1,000/µL)**[‡] | | | 0.057[†] |
| Spline 1 | 1.00 | (0.99 – 1.00) | |
| Spline 2 | 1.00 | (0.99 – 1.01) | |
| **Albumin (per 1 g/dL)**[‡] | | | 0.001[*] |
| **Spline 1** | 0.63 | (0.35 – 1.15) | |
| **Spline 2** | 0.66 | (0.26 – 1.69) | |
| **Obesity (body mass index 30)** | 0.60 | (0.36 – 0.98) | 0.040[*] |

Abbreviations: ASA = American Society of Anesthesiologists; NAFLD = non-alcoholic fatty liver disease

[*] Significant at the p<0.05 level

[†] This variable was retained on the basis of minimized Aikake Information Criterion and Bayesian Information Criterion values

[‡] Joint hypothesis tests were performed to test the statistical significance of variables modeled using restricted cubic splines

**Table 4:**

Discrimination (C-statistics) of Post-Operative Mortality Models at 30, 90, and 180 Days

|  | Risk Score | 30-Day | 90-Day | 180-Day |
|---|---|---|---|---|
| **Derivation** | **VOCAL-Penn** | **0.872 (0.836 – 0.908)** | **0.841 (0.813 – 0.868)** | **0.817 (0.792 – 0.842)** |
|  | **Mayo Score** | 0.773 (0.731 – 0.815) | 0.739 (0.705 – 0.773) | - |
|  | **MELD** | 0.724 (0.674 – 0.774) | 0.680 (0.642 – 0.718) | 0.664 (0.630 – 0.697) |
|  | **MELD-Sodium** | 0.747 (0.701 – 0.794) | 0.694 (0.656 – 0.732) | 0.680 (0.647 – 0.713) |
|  | **CTP** | 0.624 (0.576 – 0.671) | 0.590 (0.559 – 0.621) | 0.591 (0.564 – 0.618) |
| **Validation** | **VOCAL-Penn** | **0.859 (0.809 – 0.909)** | **0.822 (0.760 – 0.883)** | **0.796 (0.743 – 0.849)** |
|  | **Mayo Score** | 0.766 (0.676 – 0.855) | 0.737 (0.663 – 0.810) | - |
|  | **MELD** | 0.724 (0.617 – 0.832) | 0.704 (0.623 – 0.784) | 0.696 (0.626 – 0.765) |
|  | **MELD-Sodium** | 0.752 (0.651 – 0.854) | 0.722 (0.652 – 0.792) | 0.699 (0.632 – 0.764) |
|  | **CTP** | 0.682 (0.580 – 0.785) | 0.659 (0.587 – 0.731) | 0.637 (0.577 – 0.697) |

Abbreviations: MELD = model for end-stage liver disease; CTP = Child-Turcotte-Pugh