Title: Derivation of a simple postoperative delirium incidence and severity prediction model
Abbreviated title: Postoperative delirium incidence and severity model

Authors:
Lindroth, H.,[1,2,3] Bratzke, L.,[2] Twadell, S.,[1,4] Rowley, P.,[1] Kildow, J.,[1,5] Danner, M.,[1] Turner, L.,[1] Hernandez, B.,[1] Chang, W., Brown, R.,[2] Sanders, R.D.[1*]

[1]Department of Anesthesiology, School of Medicine and Public Health, University of Wisconsin-Madison, Madison WI

[2]School of Nursing, University of Wisconsin-Madison, Madison, WI

[3]School of Nursing, School of Medicine-Center for Aging Research, Indiana University, Indianapolis, IN

[4]Charles E. Schmidt College of Medicine, Florida Atlantic University

[5]School of Medicine, Indiana University, Indianapolis, IN

*Corresponding author: Robert D. Sanders, Department of Anesthesiology, University of Wisconsin School of Medicine and Public Health, 600 Highland Avenue, B6/319 CSC Madison, WI 53792-3272 Telephone: 608-263-8100 Fax: 608-263-0575 Madison, USA.
Email: robert.sanders@wisc.edu

Number of Figures: 4

Word Count: 3420

MESH Keywords: aging, delirium, perioperative, prediction, surgical risk

Summary

Background: Delirium is an important postoperative complication, yet a simple and effective delirium prediction model remains elusive. We hypothesized that the combination of the National Surgical Quality Improvement Program (NSQIP) risk calculator for serious complications (NSQIP-SC) or risk of death (NSQIP-D), and cognitive tests of executive function (Trail Making Test A and B [TMTA, TMTB]), could provide a parsimonious model to predict postoperative delirium incidence or severity.

Methods: Data were collected from 100 adults ($\geq$65yo) undergoing major non-cardiac surgery. In addition to NSQIP-SC, NSQIP-D, TMTA and TMTB, we collected participant age, sex, ASA score, tobacco use, type of surgery, depression, Framingham risk score, and preoperative blood pressure. Delirium was diagnosed with the Confusion Assessment Method (CAM), and the Delirium Rating Scale-R-98 (DRS) was used to assess symptom severity. LASSO and Best Subsets logistic and linear regression were employed in line with TRIPOD guidelines.

Results: Three participants were excluded due to intraoperative deaths (2) and alcohol withdrawal (1). Ninety-seven participants with a mean age of 71.68$\pm$4.55, 55% male (31/97 CAM+, 32%) and a mean Peak DRS of 21.5$\pm$6.40 were analyzed. Of the variables included, only NSQIP-SC and TMTB were identified to be predictors of postoperative delirium incidence ($p<0.001$, AUROC 0.81, 95% CI: 0.72, 0.90) and severity ($p<0.001$, Adj. R^2: 0.30).

Conclusions: In this cohort, preoperative NSQIP-SC and TMTB were identified as predictors of postoperative delirium incidence and severity. Future studies should verify whether this two-factor model could be used for accurate delirium prediction.

Introduction

Delirium, an acute brain failure, is a common surgical complication experienced by approximately 50% of patients, incurring an estimated annual U.S. cost of $152 billion.[1-4] As such, delirium is a crucial public health concern, as it is significantly associated with increased mortality[5] and morbidity[6], subsequent cognitive decline,[7, 8] and loss of independence.[5, 9] One in three cases of delirium may be preventable when multicomponent delirium prevention measures are implemented.[10, 11] Implementation of such measures would likely be most efficient in high risk individuals, yet clinicians are encumbered by an arduous list of potential patient and perioperative risk factors to identify at risk individuals. Prediction models facilitate the identification of high-risk individuals with the area under the receiver operator curve (AUROC) statistic, which is used to evaluate predictive ability to differentiate health and disease, with a value of 0.5 being no better than chance and 1.0 being perfect. Our recent systematic review identified moderate predictive ability (AUROC 0.73-0.94) in four postoperative delirium prediction models.[12] While an AUROC of 0.94 is excellent, that model includes postoperative data, which may be inflating model performance, thus invalidating its use as a *preoperative* prediction model. Critically, we did not identify a prediction model for delirium severity in the literature. A severity model may have the most utility, as it would identify patients who are likely to incur the severest delirium and therefore are most likely to benefit from a delirium prevention plan. An ideal model that is clinically applicable would be predictive of delirium incidence and its severity as well as parsimonious, such that it would use a limited number of variables that can be cheaply and easily obtained.

The ability to identify high-risk individuals prior to their surgery is critical to delirium prevention. To identify potential candidate predictors, we considered the pathogenesis of delirium and sought to identify both predisposing (age, depression, and medical comorbidities) and precipitating (surgery) factors. The *Cognitive Disintegration Model* posits that an individual with increasing risk, or vulnerability, to delirium will require less of a precipitating stimulus to cross over the "Delirium Threshold" and become delirious.[13] This is illustrated in the cognitive trajectory in **Figure 1**. In contrast, an individual with less predisposing risk factors will require a large stimulus to precipitate delirium. Hence, it is crucial to consider the future precipitating event into delirium prediction models when possible. The National Surgical Quality Improvement Program (NSQIP) online risk calculator[14, 15] combines several comorbidities (that provide information on predisposing risk factors) and the estimated magnitude of the precipitating event, the surgery, to calculate risk of events. The NSQIP risk score has been widely validated and applied to predict outcomes in various surgical populations, but has not been applied in delirium prediction.[12, 16-19] We hypothesized that NSQIP risk of serious complications (NSQIP-SC) would be a stronger predictor of delirium over NSQIP risk of death (NSQIP-D) as the causal relationship between serious complications and delirium is likely stronger than the association between delirium and the risk of death.[20] As delirium is a cognitive disorder, we further hypothesized that cognitive data (that is not included in NSQIP scores) could enhance the prediction of the surgical risk scores. Indeed our recent systematic review identified that current delirium prediction models do not evaluate specific cognitive domains, such as executive function.[12] Executive function facilitates attention and problem-solving.[21] Significant associations between preoperative executive function and postoperative delirium incidence have been reported.[22-24] In sum, our aim was to examine the predictive ability of the

NSQIP risk scores with a measure of executive function among other potential delirium risk factors to develop a parsimonious model to predict postoperative delirium incidence and severity.

Methods

This analysis is a sub-study drawn from an ongoing prospective perioperative cohort study that is approved by the University of Wisconsin-Madison, Health Sciences Institutional Review Board (#2015-0374) and registered with ClinicalTrials.gov (ref: NCT03124303, NCT01980511). Between August 2015 and May 2018, 1,049 potential participants were screened from vascular, urology, general and spine surgical clinics (HL, ST, LT, PR, JK, MD, BH, RDS). As shown in **Figure 2**, 100 subjects were recruited and 97 were included in the final analysis.

Preoperative Predictors and Assessment
Preoperatively, participants underwent an interview and completed assessments of executive function, functional status, depression, and delirium using the Confusion Assessment Method (CAM).[25] Executive function was assessed through two well- validated and widely used measures, Trail Making Test A (TMTA) and Trail Making Test B (TMTB).[26] These tests require participants to connect a series of circles in ascending order. TMTA is composed of 25 encircled numbers; TMTB alternates between 23 encircled numbers and letters. Scoring is based on time to completion; a longer completion time indicates worse executive function. Raw scores were used in the analysis. Functional ability was assessed using the Instrumental Activities of Daily Living (iADL).[27] Depression was assessed with the Geriatric Depression Scale (GDS), which is composed of 15 yes/no questions and is widely validated in the older adult population.[28] Demographic information included age, sex, years of education, and tobacco use history. Data extracted from the electronic health record included preoperative blood pressure values, height, weight, past medical history including comorbidities, current outpatient medication use, and the American Society of Anesthesiologists Physical Status (ASA) classification for the planned surgical procedure. ASA classification ranges from 1-6 and is determined by the attending anesthesiologists prior to surgical intervention.[29, 30] Each level has a specific definition with higher levels indicating increasing levels of disease, comorbidities and risk. As vascular surgery is often associated with delirium,[31] it was selected *a priori* to be included as a covariate to examine whether surgical type was sufficient to predict delirium or if the additional information provided by a surgical risk score was necessary for a prediction model. Preoperative comorbidities as predisposing risk factors were assessed through the ASA score and Framingham Cardiovascular Disease 10-year Risk Calculator (Framingham CVD).[32, 33] Framingham CVD was calculated through the online, interactive risk score calculator using data extracted from electronic health records and participant interview. This calculator includes sex, age, systolic blood pressure (SBP), treatment for hypertension, current smoker, diabetes and BMI. The most recent SBP value prior to surgery was used. BMI was calculated from collected height/weight values. The ACS NSQIP online surgical risk calculator[34], (http://riskcalculator.facs.org) was used to obtain the risk scores for serious complications (NSQIP-SC) and death (NSQIP-D). Risk is calculated from both predisposing factors and precipitating factors (**Figure 1**). This calculator employs twenty patient predictors and pairs these with the Current Procedural Terminology (CPT) code, providing a risk score specific to each procedure. The inputted variables are as follows: age, sex, functional status (independent, partially dependent, dependent), emergency

case, ASA classification, steroid use for chronic condition, ascites within 30-day prior to surgery, systemic sepsis within 48-hours prior to surgery, ventilator dependency, disseminated cancer, diabetes, hypertension with medications, congestive heart failure (within 30-days prior to surgery), dyspnea, current smoker (within 1-year), history of severe COPD, dialysis, acute renal failure, height and weight as well as surgical procedure. There are 1,557 distinct CPT codes, ranging from minor surgeries such as a cholecystectomy to major surgeries such as thoracoabdominal aortic aneurysm repair.

Delirium Assessment

Pre- and postoperatively, participants were formally assessed for delirium and symptoms using the widely validated Confusion Assessment Method (CAM),[25] 3D-CAM,[35] and Delirium Rating Scale-R-98 (DRS),[36] twice daily, between the hours of 0500-1000 and 1600-2200 on postoperative days 1-4 (regardless of day of the week). The CAM and 3D-CAM were administered concurrently to provide both a global and comprehensive view of delirium symptoms while providing a structured interview format. If the participant was CAM positive at the postoperative day 4 afternoon assessment, the participant was followed until delirium resolved to collect data on delirium duration. If participants were ventilated in the intensive care unit (ICU), the CAM-ICU[37, 38] was administered. The DRS-R-98 (DRS) is a 16-item assessment tool that measures delirium symptoms and severity. The maximum score is 44-points, an increasing score indicates worse delirium.[36] The research team (RDS, HL, ST, JK, LT, PR, MD, BH) met at least weekly to discuss delirium assessment findings and DRS ratings.

Research Team

Each research team member (RDS, HL, ST, JK, LT, PR, MD, BH) underwent intensive training on delirium interview completion including CAM, 3D-CAM, CAM-ICU, and DRS. CAM training was completed as part of the NeuroVISION cohort study[39] and the first author (HL) was officially trained on the CAM at the 2016 CEDARTREE Delirium Bootcamp and the CAM-ICU at the 2016 American Delirium Society pre-conference. Team members viewed 6 videos on CAM and 3D-CAM administration from the Hospital Elder Life Program[40] website followed by an interactive team discussion on CAM/3D- CAM completion and observations. Team members shadowed the first author for six in-person CAM/3D-CAM assessments followed by in-depth discussion on observations, symptom rating (DRS), and bedside manner. The first author shadowed the team members for six CAM/3D- CAM/CAM-ICU/DRS assessments to ensure competency.

Sample Size

The sample size was determined based on the need for a parsimonious delirium prediction model. We estimated three to four risk factors would form a reasonable clinical model if an AUROC >0.80 was obtained. Sample size was based on logistic regression and determined using the rule of 8-10 outcome events (delirium) per variable.[41, 42] The decision to analyze was made after 100 participants were recruited with a delirium incidence rate of 32%.

Statistical Analysis

Patient characteristics were described using means $\pm$ standard deviations for continuous variables and frequency counts with percentages for categorical variables. Dependent on the distribution of the data, continuous variables were compared using Student's t-test or Mann-Whitney U-test.

Categorical variables were compared using X^2. The outcome variables were Delirium Yes/No (DELYN) for logistic regression and delirium severity using the Peak DRS Total Score (DRS) for linear regression. Missing data was identified in the following variables (#missing): TMTB (1), GDS15 (1), TMTA (6), and Tobacco Pack Years (10). Little's test of missing completely at random (MCAR) was not significant indicating that the missing data were missing completely at random and did not influence the analysis. Therefore, as outlined by Little (1988) the listwise deletion of participants with missing data was appropriate.[43] Significance was notated with a p-value ≤ 0.05. NCSS v12.0, Stata/IC v15.0 and R v1.1453 were used for statistical analysis. These statistical packages were used to verify stable results across statistical software packages. HL, WC, and RB conducted the statistical analysis.

First, to evaluate the predictive ability of NSQIP-SC over NSQIP-D (as composites of the predisposing and precipitating factors), logistic (DELYN) regression models were completed and compared also to ASA classification and Framingham risk (as measures of predisposing factors only). Second, a delirium prediction model was developed. We did not employ univariate statistics to select candidate predictors as this may lead to poor performing predictors and overfitting.[44] To counter the effects of small sample sizes and reduce bias within data, we employed a statistical shrinkage regression technique, using Least Absolute Shrinkage and Selection Operator (LASSO).[45-47] This technique reduces the noise within the data, allowing true signals to be detected and avoids common problems such as model overfitting. Candidate variables demonstrating the smallest Mallow's Cp value,[48] indicating precise predictors, were then applied in Best Subsets regression. Best Subsets regression is an automated regression approach that evaluates all possible combinations of candidate predictors.[49, 50] The output provides a set models with model fit statistics. Model selection was based on assessment of model fit using Akaike information criteria (AIC), Bayesian information criteria (BIC), and McKelvey and Zavoina's Pseudo-R2.[51] The area under the receiver operating characteristic curve (AUROC) with 95% CI was calculated. Calibration was assessed through goodness-of-fit tests calculated by the Hosmer-Lemeshow statistic. Sensitivity, specificity, positive predictive and negative predictive values were calculated and reported.

The peak delirium severity (DRS) score was transformed using the Box-Cox Method[52] with the optimal Lambda value due to the positive skew, please refer to **Figure 4-A and B for raw and transformed plots**. The regression modeling procedures outlined in the paragraph above for logistic regressions were repeated for the linear regression model. Model selection was based on assessment of model fit using Akaike information criteria (AIC), Bayesian information criteria (BIC), and adjusted R^2.[51]

Results

Thirty-one participants (32%) experienced postoperative delirium with a mean peak DRS severity total score of 21.48 ($\pm$SD 6.40). Forty-two percent experienced hypoactive delirium. The median delirium duration was one day (24 hours). Participant characteristics are summarized in **Table 1**. Delirious patients had higher preoperative NSQIP risk scores of serious complications (NSQIP-SC) and death (NSQIP-D), worse executive function tests, were more likely to have had a vascular surgery, and higher ASA status (univariate, $p<0.05$). Significant pairwise correlations were demonstrated between the DRS and NSQIP-SC, NSQIP-D, TMTA and TMTB (univariate, $p<0.05$). No significant differences were identified between the non-delirious and delirious groups in terms of age, sex, education level, past/present tobacco use, blood pressure metrics, functional status, GDS15, and Framingham CVD.

Surgical Risk Scores and Delirium Incidence
We confirmed preoperative NSQIP-SC as a robust predictor of postoperative delirium incidence using single factor logistic regression models for NSQIP-SC and NSQIP-D. Both demonstrated moderate to fair predictive ability with an AUROC of 0.76 (95% CI: 0.66, 0.87) and 0.73 (95% CI: 0.62-0.84), respectively. Although similar in their projected ability to predict individuals at a higher risk for delirium development, support for the NSQIP-SC model, over the NSQIP-D model, is provided by optimal AIC (1.086) and the BIC (-333.254) metrics. NSQIP-SC also performed better than the ASA classification (AUROC 0.63, 95% CI: 0.53-0.73) and the Framingham risk score (0.53, 95% CI: 0.40-0.65). The reported odds ratios, coefficients, AUROC, sensitivity and specificity, and model fit statistics including AIC and BIC are in **Supplementary Table 1**.

Derivation of a Delirium Incidence Prediction Model
In order to derive a potential delirium incidence prediction model we followed TRIPOD guidelines to apply a statistical shrinkage technique (LASSO) followed by Best Subsets regression using age, sex, NSQIP-SC, NSQIP-D, tobacco pack years, vascular surgery, ASA classification, Framingham CVD, GDS15, TMTA, TMTB, and blood pressure metrics. A two-factor logistic regression model containing preoperative NSQIP-SC and TMTB to predict postoperative delirium incidence was selected as this model demonstrated optimal model fit statistics. Hosmer-Lemeshow goodness-of-fit test was not significant ($p=0.37$), indicating accurate model calibration. The model demonstrated moderate predictive ability (AUROC 0.81, 95% CI: 0.72-0.90), a 5-point increase in the NSIQP-SC score increased the probability of delirium incidence by 10% (**Figure 3**). The logistic regression model including classification metrics is displayed in **Table 2**.

Derivation of a Delirium Symptom Prediction Model
Preoperative NSQIP-SC was confirmed as a predictor of postoperative peak DRS using simple linear regression models for NSQIP-SC ($p<0.0001$, AdjR^2: 0.184). Similar to the logistic regression results above NSQIP-D was also significantly associated with DRS ($p=0.04$, AdjR^2: 0.03). However, the NSQIP-SC model demonstrated higher adjusted R^2, and lower AIC (1.525) and BIC (-290.718) metrics, providing support for that predictor over NSQIP-D. Similar to logistic regression for postoperative delirium incidence, ASA ($p=0.04$, AdjR^2: 0.03) and

Framingham risk score (p=0.69, AdjR^2: 0.01) did not demonstrate a strong predictive relationship with DRS.

The model demonstrating optimal fit statistics for postoperative peak delirium severity was a two-factor linear regression model containing preoperative NSQIP-SC and TMTB using LASSO and Best Subsets regression. This two-factor model reports an adjusted R^2 of 0.30 (p<0.001), thus explaining 30% of the variability in observed delirium symptoms. For every 1-standard deviation (SD) increase in the NSQIP-SC score, the DRS SD will increase by 0.42 points. Further model details are displayed in **Table 2** and **Figure 4**. Age, sex, NSQIP-D, ASA, tobacco pack years, vascular surgery, Framingham CVD, GDS15, TMTA, and blood pressure metrics were not identified as significant predictors.

Discussion

Our analysis of a prospective perioperative cohort found NSQIP-SC to be a robust predictor of postoperative delirium incidence and severity. The preoperative NSQIP-SC score combines data on both predisposing risk factors and the estimated magnitude of the surgery, i.e. the precipitating event and is well positioned to contribute information on the risk of delirium. The addition of the executive function measure, TMT-B, improved performance of both prediction models. Previous research has identified executive function to be significantly associated with delirium incidence.[22-24] However, as identified by our recent systematic review of delirium prediction models in older adults, an executive function measure has not been applied previously.[12] Given the breakdown in executive function during delirium and prior data on the predisposition to delirium by impaired cognition, incorporating a cognitive variable appears biologically important; our data show that it is statistically important too. This study expands current knowledge by examining the utility of NSQIP risk scores and executive function in predicting delirium incidence and severity.

The NSQIP-D score was significantly associated with both postoperative delirium incidence and severity, however, it was not selected as a predictor in our variable selection. Delirium often results from a complicated perioperative course, particularly following a major surgery, hence a relationship with NSQIP-SC is plausible. Furthermore, a recent systematic review questioned the strength of the association between postoperative delirium and mortality, hence *a priori* we hypothesized that NSQIP-SC would perform better than NSQIP-D for predicting postoperative delirium incidence and severity.[20] This was supported by several statistical measures in our dataset. The modeling procedures, LASSO and Best Subsets regression, did not identify vascular burden, comorbidities, smoking history, and depression as important predictors although these variables have been identified as significant risk factors for postoperative delirium incidence in prior research. This may be due to a number of factors. First, their prevalence in this population of study may not be sufficient for prediction. In order for a risk factor to also be an accurate predictor, it must be sufficiently prevalent in the at-risk population.[53, 54] Secondly, late-life depression, vascular burden, and tobacco use often co-occur leading to overlapping data capture. Variables that capture similar information fail to contribute important information during modeling procedures.[12] Lastly, these variables contribute information on predisposing risk, but do not include valuable information on the future precipitating event, the planned surgical

procedure. The inclusion of information about the future precipitating event, as done in NSQIP-SC, contributes crucial information to the prediction of postoperative delirium.

Strengths

The strengths of this study include its prospective perioperative design, statistical methods chosen, and rigorous delirium assessments including outcomes based on incidence and severity of delirium. This novel application of a readily available, simple tool has potential for broad application in delirium-focused clinical care. The NSQIP-SC score combines several potential preoperative risk factors for delirium (age, functional status, current tobacco use, vascular burden) with the precipitating event, the planned surgery, and provides a single risk score that is easy to interpret, *i.e.,* a 5 point increase in NSQIP-SC results in a 10% increase in the probability of a patient experiencing delirium. Given that the patient becomes delirious postoperatively, quantifying the potential impact of this precipitating event is clearly a key feature of a delirium prediction model.

Delirium represents a severe breakdown in an individual's executive function. The *Cognitive Disintegration Model* posits that there is a critical threshold (i.e. the "Delirium Threshold") in cognition (or network connectivity) that must be crossed for delirium to result. Individuals with more vulnerability (predisposing factors) going into surgery, i.e. older age, multiple comorbidities and worsening executive function, are closer to that "Delirium Threshold" (**Figure 1**) and may need a smaller precipitating event to push them into delirium. In this context, inclusion of executive function, which is not captured in NSQIP-SC, improved model performance. Furthermore, this simple two-factor model captures information on both the predisposing and precipitating factors for delirium. Its parsimonious nature is a clinical strength.

Limitations

This study has several limitations to consider. While these models were built using statistical methods optimized for modeling a limited number of events, the small sample size may still have had an effect on the results. The recommendations by the TRIPOD guidelines and the CHARMS checklist were followed to use statistical shrinkage procedures to minimize model overfitting. Nonetheless, it is worth emphasizing the study sample is small with only 31 delirious participants. The population is largely homogenous in terms of years of education and ethnicity. In larger and more diverse populations, additional factors may enhance model performance. Future studies should focus on the broad external validation of these models following the statistical methodology outlines by the TRIPOD guidelines. This requires a new perioperative cohort study that will recruit patients from diverse populations and will be the target of future grant applications.

Conclusion

This analysis of a prospective perioperative cohort study identified NSQIP-SC and TMT-B as predictors for delirium incidence and severity. A preliminary delirium prediction model was created for both delirium incidence and severity. These models should be validated in future studies.

Declaration of Interests

None declared

Details of Authors Contributions
Patient recruitment: HL, ST, PR, JK, MD, LT, BH, RDS
Clinical execution: HL, ST, PR, JK, MD, LT, BH, RDS
Guidance on cognitive tests: LB
Data analysis and interpretation: HL, WC, RB, RDS
Manuscript writing: HL, RDS
Manuscript revision: LB, ST, PR, JK, MD, LT, BH, WC, RB
Study Conception, protocol: RDS
Study idea: HL, RDS

References

1 Inouye SK, Westendorp RG, Saczynski JS. Delirium in elderly people. *Lancet* 2014; **383**: 911-22

2 Leslie DL, Inouye SK. The importance of delirium: economic and societal costs. *Journal of the American Geriatrics Society* 2011; **59 Suppl 2**: S241-3

3 Leslie DL, Marcantonio ER, Zhang Y, Leo-Summers L, Inouye SK. One-year health care costs associated with delirium in the elderly population. *Archives of internal medicine* 2008; **168**: 27-32

4 Diagnostic and Statistical Manual of Mental Disorders-5. *Diagnostic and Statistical Manual of Mental Disorders-5*: American Psychiatric Association, 2013

5 Ha A, Krasnow RE, Mossanen M, et al. A contemporary population-based analysis of the incidence, cost, and outcomes of postoperative delirium following major urologic cancer surgeries. *Urologic oncology* 2018

6 Gleason LJ, Schmitt EM, Kosar CM, et al. Effect of Delirium and Other Major Complications on Outcomes After Elective Surgery in Older Adults. *JAMA Surgery* 2015; **150**: 1134-40 7p

7 Inouye SK, Marcantonio ER, Kosar CM, et al. The short-term and long-term relationship between delirium and cognitive trajectory in older surgical patients. *Alzheimers Dement* 2016

8 Pandharipande PP, Girard TD, Jackson JC, et al. Long-Term Cognitive Impairment after Critical Illness. *The New England Journal Of Medicine* 2013; **369**: 1306-16

9 Oh ES, Fong TG, Hshieh TT, Inouye SK. Delirium in older persons: Advances in diagnosis and treatment. *Jama* 2017; **318**: 1161-74

10 Hshieh TT, Yue J, Oh E, et al. Effectiveness of Multicomponent Nonpharmacological Delirium Interventions: A Meta-analysis. *JAMA internal medicine*, 2015

11 Siddiqi N, Harrison JK, Clegg A, et al. Interventions for preventing delirium in hospitalised non-ICU patients. *Cochrane Database Syst Rev* 2016; **3**: CD005563

12 Lindroth H, Bratzke L, Purvis S, et al. Systematic review of prediction models for delirium in the older adult inpatient. *BMJ Open* 2018; **8**: e019223

13 Sanders RD. Hypothesis for the pathophysiology of delirium: role of baseline brain network connectivity and changes in inhibitory tone. *Medical Hypotheses* 2011; **77**: 140-3

14 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**: 833-42.e1-3

15 Shiloach M, Frencher SK, Jr., Steeger JE, et al. Toward robust information: data quality and inter-rater reliability in the American College of Surgeons National Surgical Quality Improvement Program. *Journal Of The American College Of Surgeons* 2010; **210**: 6-16

16 Helman SN, Brant JA, Moubayed SP, Newman JG, Cannady SB, Chai RL. Predictors of length of stay, reoperation, and readmission following total laryngectomy. *The Laryngoscope* 2017; **127**: 1339-44

17 Kubasiak JC, Landin M, Schimpke S, et al. The effect of tobacco use on outcomes of laparoscopic and open ventral hernia repairs: a review of the NSQIP dataset. *Surg Endosc* 2017; **31**: 2661-6

18 Chimukangara M, Helm MC, Frelich MJ, et al. A 5-item frailty index based on NSQIP data correlates with outcomes following paraesophageal hernia repair. *Surg Endosc* 2017; **31**: 2509-19

19 Guillamondegui OD, Gunter OL, Hines L, et al. Using the National Surgical Quality Improvement Program and the Tennessee Surgical Quality Collaborative to improve surgical outcomes. *Journal Of The American College Of Surgeons* 2012; **214**: 709-14; discussion 14-16

20 Hamilton GM, Wheeler K, Di Michele J, Lalu MM, McIsaac DI. A Systematic Review and Meta-analysis Examining the Impact of Incident Postoperative Delirium on Mortality. *Anesthesiology* 2017; **127**: 78-88

21 Diamond A. Executive functions. *Annu Rev Psychol* 2013; **64**: 135-68

22 Smith PJ, Attix DK, Weldon BC, Greene NH, Monk TG. Executive function and depression as independent risk factors for postoperative delirium. *Anesthesiology* 2009; **110**: 781-7 7p

23 Rudolph JL, Jones RN, Grande LJ, et al. Impaired executive function is associated with delirium after coronary artery bypass graft surgery. *Journal of the American Geriatrics Society* 2006; **54**: 937-41

24 Fong TG, Hshieh TT, Wong B, et al. Neuropsychological Profiles of an Elderly Cohort Undergoing Elective Surgery and the Relationship Between Cognitive Performance and Delirium. *Journal of the American Geriatrics Society* 2015; **63**: 977-82 6p

25 Inouye SK, van Dyck CH, Alessi CA, Balkin S, Siegal AP, Horwitz RI. Clarifying confusion: the confusion assessment method. A new method for detection of delirium. *Annals of internal medicine* 1990; **113**: 941-8

26 Strauss E. *A compendium of neuropsychological tests : administration, norms, and commentary*. 3rd ed. Oxford ; New York : Oxford University Press, 2006., 2006

27 Graf C. The Lawton Instrumental Activities of Daily Living (iADL) Scale. 2015. Available from http://consultgerirn.org/uploads/File/trythis/try_this_23.pdf (accessed April 7 2015)

28 Dennis M, Kadri A, Coffey J. Depression in older people in the general hospital: a systematic review of screening instruments. *Age and ageing*. England, 2012; 148-54

29 Delegates AHo. ASA Physical Status Classification System. 2014. Available from https://www.asahq.org/resources/clinical-information/asa-physical-status-classification-system (accessed June 3, 2018 2018)

30 Parenti N, Reggiani ML, Percudani D, Melotti RM. Reliability of American Society of Anesthesiologists physical status classification. *Indian J Anaesth* 2016; **60**: 208-14

31 Raats JW, van Hoof-de Lepper CCHA, Feitsma MT, et al. Current factors of fragility and delirium in vascular surgery. *Annals Of Vascular Surgery* 2015; **29**: 968-76

32 D'Agostino RB, Sr., Vasan RS, Pencina MJ, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation* 2008; **117**: 743-53

33 D'Agostino V, Pencina, Wolf, Cobain, Massaro, Kannel. Framingham Heart Study; Cardiovascular Disease (10-year risk). 2008. Available from https://www.framinghamheartstudy.org/fhs-risk-functions/cardiovascular-disease-10-year-risk/# (accessed June 3 2018 2018)

34 American College of Surgeons NSQIP. Surgical Risk Calculator. 2018. Available from https://riskcalculator.facs.org/RiskCalculator/ (accessed December 2017-May 15th, 2018 2018)

35 Marcantonio ER, Ngo LH, O'Connor M, et al. 3D-CAM: derivation and validation of a 3-minute diagnostic interview for. *Annals of internal medicine* 2014; **161**: 554-61

36 Trzepacz PT, Mittal D, Torres R, Kanary K, Norton J, Jimerson N. Validation of the Delirium Rating Scale-revised-98: comparison with the delirium rating scale and the

cognitive test for delirium. *The Journal of neuropsychiatry and clinical neurosciences* 2001; **13**: 229-42

37 Ely EW, Inouye SK, Bernard GR, et al. Caring for the critically ill patient. Delirium in mechanically ventilated patients: validity and reliability of the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU). *JAMA: Journal of the American Medical Association* 2001; **286**: 2703

38 Ely EW, Margolin R, Francis J, et al. Evaluation of delirium in critically ill patients: validation of the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU). *Critical Care Medicine* 2001; **29**: 1370-9

39 Devereaux PJ, Mrkobrada M. NeuroVISION. Detection and neurological impact of cerebrovascular events in noncardiac surgical patients: a cohort evaluation. 2014: 1-51

40 Program HEL. Hospital Elder Life Program (HELP) for Prevention of Delirium. The Hospital Elder Life Program, 2018

41 Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. *Am J Epidemiol* 2007; **165**: 710-8

42 Green SB. HOW MANY SUBJECTS DOES IT TAKE TO DO A REGRESSION-ANALYSIS. *Multivariate Behav Res* 1991; **26**: 499-510

43 Roderick JAL. A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association* 1988; **83**: 1198-202

44 Ogundimu EO, Altman DG, Collins GS. Adequate sample size for developing prediction models is not simply related to events per variable. *J Clin Epidemiol* 2016; **76**: 175-82

45 Lockhart R, Taylor J, Tibshirani RJ, Tibshirani R. A SIGNIFICANCE TEST FOR THE LASSO. *Annals of statistics* 2014; **42**: 413-68

46 Tibshirani R, Bien J, Friedman J, et al. Strong rules for discarding predictors in lasso-type problems. *Journal of the Royal Statistical Society Series B, Statistical methodology* 2012; **74**: 245-66

47 Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med* 1997; **16**: 385-95

48 Mallows CL. Some Comments on C_P. *Technometrics* 1973; **15**: 661-75

49 King JE. Running a Best-Subsets Logistic Regression: An Alternative to Stepwise Methods. *Educational and Psychological Measurement* 2003; **63**: 392-403

50 Hosmer DW, Jovanovic B, Lemeshow S. Best Subsets Logistic Regression. *Biometrics* 1989; **45**: 1265-70

51 Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology (Cambridge, Mass)* 2010; **21**: 128-38

52 Arzideh F, Wosniok W, Haeckel R. Indirect reference intervals of plasma and serum thyrotropin (TSH) concentrations from intra-laboratory data bases from several German and Italian medical centres. *Clinical chemistry and laboratory medicine* 2011; **49**: 659-64

53 Steyerberg EW, Eijkemans MJ, Harrell FE, Jr., Habbema JD. Prognostic modeling with logistic regression analysis: in search of a sensible strategy in small data sets. *Medical decision making : an international journal of the Society for Medical Decision Making* 2001; **21**: 45-56

54 Pavlou M, Ambler G, Seaman SR, et al. How to develop a more accurate risk prediction model when there are few events. *BMJ (Clinical research ed)* 2015; **351**: h3868

Tables (including legends to tables)

Table 1: Description of sample and significant differences between no delirium and delirium

Variable		Mean (SD) N=97	No Delirium N=66	Delirium N=31
Age		71.68 (4.55)	71.71 (4.79)	71.61 (4.09)
NSQIP-SC		17.56 (11.66)	13.93 (9.34)	25.27 (12.49)***
NSQIP-D		2.52 (3.87)	1.85 (3.45)	3.95 (4.38)**
Framingham CVD		35.86 (19.6)	35.99 (20.74)	26.52 (16.98)
ASA		2.64 (3.87)	2.56 (0.61)	2.84 (0.52)*
Preoperative SBP		133 (17)	133 (17)	135 (17)
Preoperative DBP		74 (10)	74 (11)	74 (10)
Preoperative PP		59 (16)	59 (16)	61 (17)
Preoperative MAP		94 (11)	94 (12)	94 (10)
Tobacco Pack Years (87)		18 (24)	16 (22)	23 (29)
GDS15 (93)		2.44 (2.51)	2.38 (2.52)	2.59 (2.53)
Preoperative TMTA (91)		41.39 (16.47)	38.47 (12.79)	47.66 (21.35)*
Preoperative TMTB (96)		98.5 (52.1)	89.54 (46.71)	117.48 (60.01)*
Peak DRS Total Score		11.5 (8.3)	6.82 (3.74)	21.48 (6.40)***
Delirium duration (median)				1 day
Delirium subtypes (%)				Hypoactive: 42 Mixed: 32 Hyperactive: 19 RASS 0: 6
Frequency (type, %)				
Sex		55 (male, 55%)	39 (59%)	16 (52%)
Type of Surgery-	Vascular	38 (vascular, 39%)	22 (33%)	16 (51%)*
	Other	12 (general, 12%) 36 (Spine, 37%) 11 (Urology, 11%)	44 (66%)	15 (48%)
Years of Education		3 (<12yrs, 3%) 29 (12yrs, 30%) 65 (>12yrs, 67%)	1 (1%) 19 (29%) 46 (70%)	2 (6%) 10 (32%) 19 (61%)
Current tobacco user Past tobacco user		19 (yes, 16%) 65 (yes, 67%)	10 (15%) 43 (65%)	6 (19%) 22 (71%)
Significance levels: * = p<0.05, **p<0.001, ***p<0.0001				
Abbreviations: ASA=American Society of Anesthesiologists classification score, DBP=Diastolic Blood Pressure, DRS=Delirium Rating Scale-98-R, Framingham CVD=Framingham Cardiovascular Risk Score, GDS15=Geriatric Depression Scale-15, MAP=Mean Arterial Pressure, NSQIP_SC=National Surgical Quality Improvement Program Risk for Serious Complications, NSQIP-D=National Surgical Quality Improvement Program Risk for Death, PP=Pulse Pressure, SBP=Systolic Blood Pressure, SD=Standard Deviation, TMTA=Trail Making Test A, TMTB=Trail Making Test B				

Table 2: Results of LASSO and Best Subsets Regression. Model Statistics shown below.

Delirium Incidence Prediction Model

LR chi2 p-value	Variable	Odds Ratio	Std. Error (95%CI)	p-value	AUROC (95%CI)	McKelvey& Zavoina's Adj R2
34.75	NSQIP-SC	1.10	0.03 (1.05-1.15)	<0.001	0.81 (0.72-0.90)	0.33
<0.0001	TMTB	1.01	0.01 (1.00-1.02)	<0.05		

Model Classification: Sensitivity: 0.52, Specificity: 0.88, PPV: 0.67, NPV: 0.79 AUROC 0.81(0.72-0.90)

Delirium Symptom Prediction Model

Transformed DRS

F-Statistic p-value	Variable	Coefficient	Std. Error (95%CI)	p-value	Std. Betas	Adj R-Squared
21.35	NSQIP-SC	0.02	0.004 (0.01-0.03)	<0.001	0.41	0.30
<0.0001	TMTB	0.003	0.001 (0.002-0.01)	<0.001	0.35	

Raw DRS

	Variable	Coefficient	Std. Error (95%CI)	p-value	Std. Betas	Adj R-Squared
34.73	NSQIP-SC	0.29	0.06 (0.17-0.42)	<0.001	0.41	0.29
<0.0001	TMTB	0.05	0.01 (0.02- 0.08)	<0.001	0.33	

Abbreviations: Adj. =adjusted, AUROC=Area Under the Receiver Curve Operator, CI=Confidence Interval, DRS=Delirium Rating Scale-R-98, LR=Likelihood ratio, NPV: Negative Predictive Value, PPV: Positive Predictive Value, Std. =Standard.

Legends to illustrations

Figure 1 illustrates the Cognitive Trajectory. The relationship between cognitive abilities (predisposing, y axis) and the precipitating event, *i.e.* surgery over time (x-axis) is shown, with each individual trajectory displayed with a horizontal line. The dashed line, situated above the x-axis of "time", represents the "Delirium Threshold." (A) Trajectory #1 (gray line, numbered 1) displays an individual with maximum cognitive abilities. They have a surgery, but do not cross the "Delirium Threshold." Trajectory #2 (blue line, numbered 2) contrasts #1 by showing an individual with decreased cognitive abilities. This individual undergoes the same surgery and crosses over the "Delirium Threshold" to experience delirium. (B) Trajectory #3 (black-dashed line, numbered 3) returns to an individual with maximum cognitive abilities. A sufficiently large precipitating event will push this individual across the "Delirium Threshold", inducing delirium. Trajectory #1 (gray line) is transposed onto this graph to show the difference in magnitude and impact of the precipitating event. (C) When developing a prediction model for delirium, it may be important to consider not only the predisposing risk factors, but also the influence of the precipitating event. A surgical risk score such as NSQIP combines both predisposing risk and the future-precipitating event into one score, which may be optimal for postoperative delirium prediction.

Figure 2 displays the inclusion and exclusion criteria and a flowchart detailing study screening, recruitment, consent, and attrition numbers.

Figure 3 illustrates the predictive ability of the NSQIP-SC and TMTB model for postoperative delirium incidence. (A) Displays the Area Under the Receiver Operator Curve statistic (AUROC). (B) Demonstrates the predicted probability of postoperative delirium incidence based on the % NSQIP-SC score. This is holding TMTB constant at zero.

Figure 4 illustrates the postoperative delirium symptom severity prediction model. Box A is a histogram showing the data distribution of the Peak Delirium Rating Scale Score (DRS). This value was transformed using the Box-Cox Method with an optimal lambda value of 0.35 achieving a near Gaussian distribution and is shown on the histogram in Box B. Boxes C-E display the predicted burden of delirium symptoms based on the NSQIP-SC and TMTB prediction model (Box C) and univariate analysis of NSQIP-SC (Box D) and TMTB (Box E). The statistics from each regression model are shown in the upper left hand corner of each box. The univariate NSQIP-SC regression model was analyzed with 97 participants. Due to one missing assessment of TMTB, Box C and E are analyzed with 96 participants.

Figure 1 illustrates the Cognitive Trajectory. The relationship between cognitive abilities (predisposing, y axis) and the precipitating event, *i.e.* surgery over time (x-axis) is shown, with each individual trajectory displayed with a horizontal line. The dashed line, situated above the x-axis of "time", represents the "Delirium Threshold." (A) Trajectory #1 (gray line, numbered 1) displays an individual with maximum cognitive abilities. They have a surgery, but do not cross the "Delirium Threshold." Trajectory #2 (blue line, numbered 2) contrasts #1 by showing an individual with decreased cognitive abilities. This individual undergoes the same surgery and crosses over the "Delirium Threshold" to experience delirium. (B) Trajectory #3 (black-dashed line, numbered 3) returns to an individual with maximum cognitive abilities. A sufficiently large precipitating event will push this individual across the "Delirium Threshold", inducing delirium. Trajectory #1 (gray line) is transposed onto this graph to show the difference in magnitude and impact of the precipitating event. (C) When developing a prediction model for delirium, it may be important to consider not only the predisposing risk factors, but also the influence of the precipitating event. A surgical risk score such as NSQIP combines both predisposing risk and the future-precipitating event into one score, which may be optimal for postoperative delirium prediction.

```
┌─────────────────────┐                    ┌──────────────────────────────────┐
│  1,054 Screened     │ ─────────────────→ │        574 Ineligible            │
│                     │                    │  • Age - 325                     │
└─────────────────────┘                    │  • Surgery type - 93             │
          │                                │  • Length of stay - 156          │
          ↓                                └──────────────────────────────────┘
┌─────────────────────┐                    ┌──────────────────────────────────┐
│  480 Eligible       │                    │        350 Excluded              │
│                     │ ─────────────────→ │  • Documented dementia -14       │
└─────────────────────┘                    │  • Residing in nursing home - 12 │
          │                                │  • MRI Contraindication - 52     │
          ↓                                │  • Communication barrier - 6     │
┌─────────────────────┐                    │  • Declined - 266                │
│  110 Consented      │                    └──────────────────────────────────┘
│                     │                    ┌──────────────────────────────────┐
└─────────────────────┘ ─────────────────→ │        10 Dropped               │
          │                                │  • Declined participation - 7    │
          ↓                                │  • Surgery cancelled - 3         │
┌─────────────────────┐                    └──────────────────────────────────┘
│ 100 Completed       │                    ┌──────────────────────────────────┐
│ Postoperatve        │ ─────────────────→ │        3 Dropped                │
│ Timepoint           │                    │  • Intraop/postop mortality - 2  │
└─────────────────────┘                    │  • ETOH induced delirium - 1     │
          │                                └──────────────────────────────────┘
          ↓
┌─────────────────────┐
│  97 Analyzed        │
└─────────────────────┘
```

┌──┐
│ Incklusion: │
│ • Age ≥ 65 years   │
│ • Anticipated length of stay following surgery of 2 days or greater   │
│ • Written informed consent to participant prior to noncardiac surgery   │
│   │
│ Exclusion: │
│ • Contraindication to MRI or EEG   │
│ • Unable or unwilling to attend follow-up study appointments   │
│ • Documented history of dementia   │
│ • Residing in a nursing home   │
│ • Undergoing carotid artery surgery or intracranial surgery   │
│ • Unable to complete neuropsychological tests or communicate with   │
│   research staff due to language or sensory barriers   │
└──┘

Figure 2 displays the inclusion and exclusion criteria and a flowchart detailing study screening, recruitment, consent, and attrition numbers.
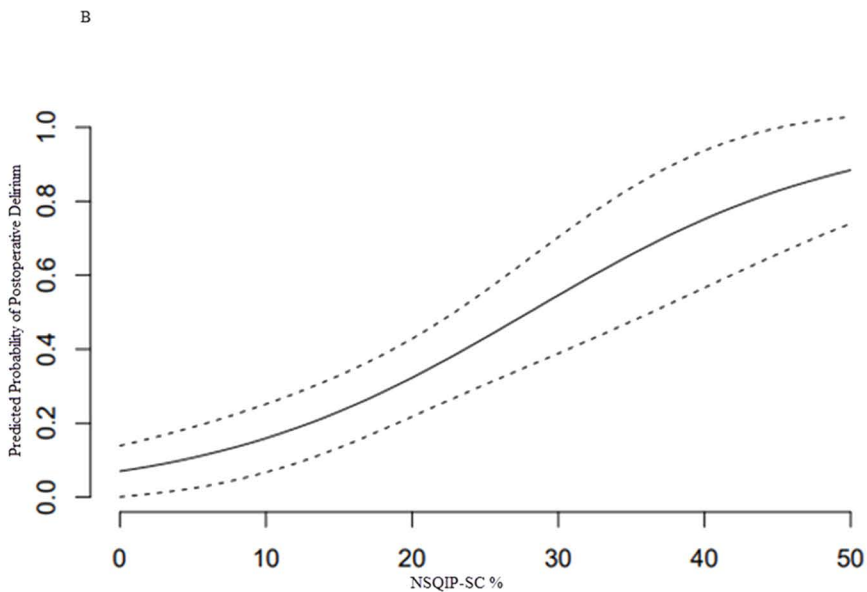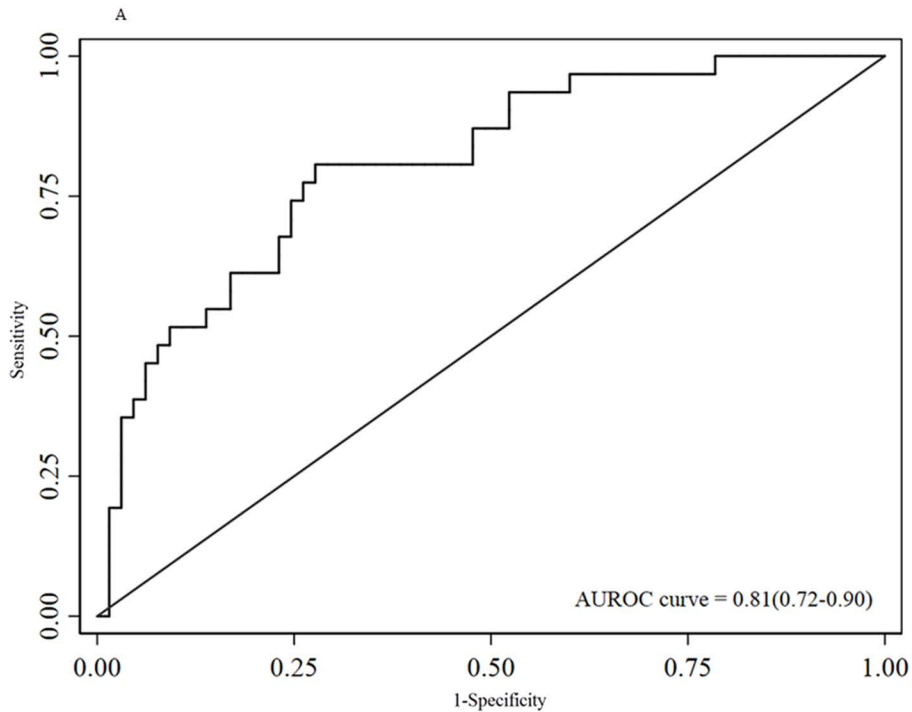
Figure 3 illustrates the predictive ability of the NSQIP-SC and TMTB model for postoperative delirium incidence. (A) Displays the Area Under the Receiver Operator Curve statistic (AUROC). (B) Demonstrates the predicted probability of postoperative delirium incidence based on the % NSQIP-SC score. This is holding TMTB constant at zero.
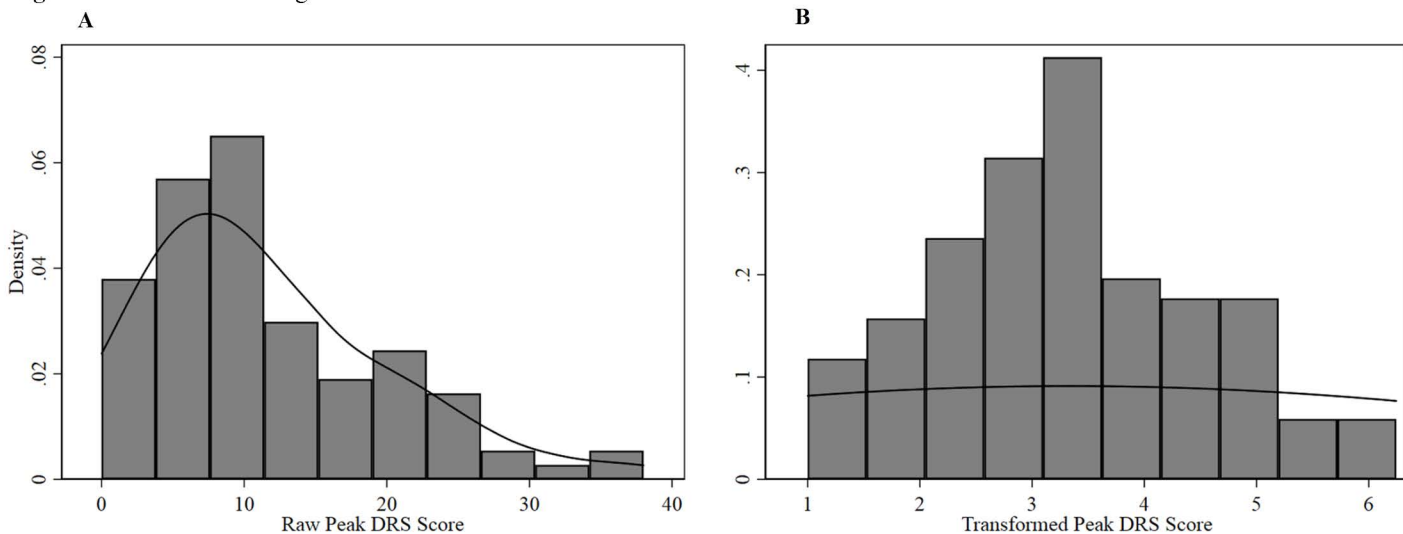
Figure 4-1 Peak DRS Histograms

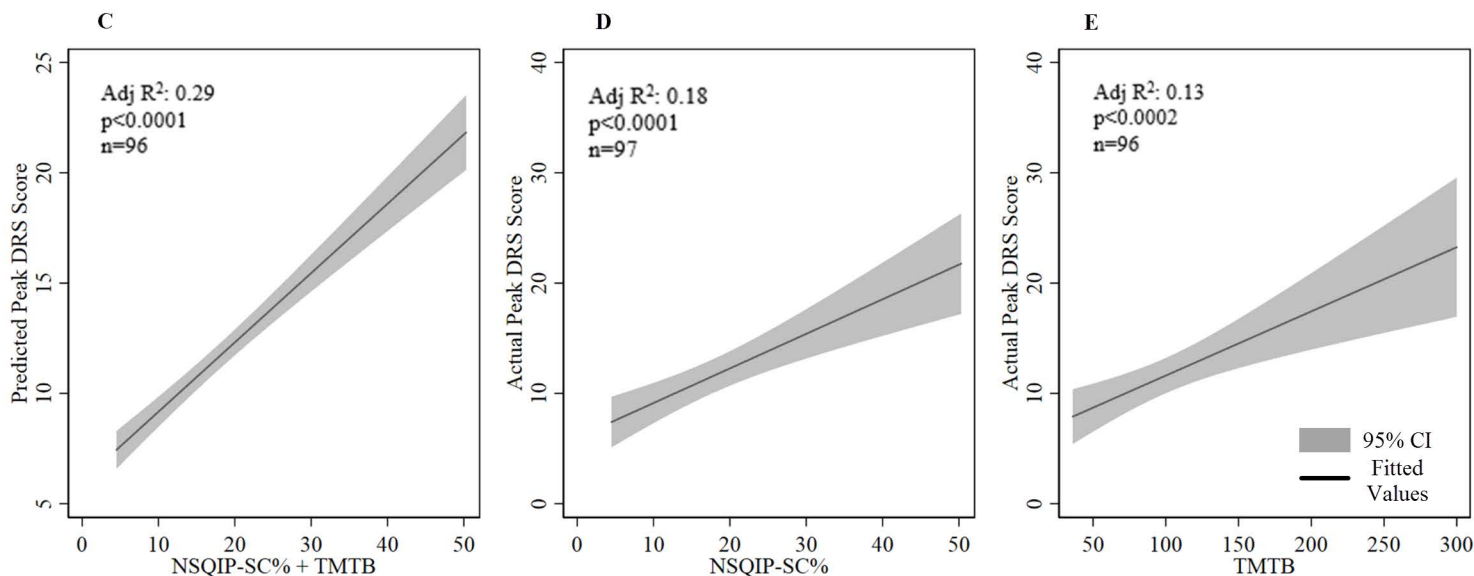**Figure 4-2** Delirium Severity Prediction Model

Figure 4 illustrates the postoperative delirium symptom severity prediction model. Box A is a histogram showing the data distribution of the Peak Delirium Rating Scale Score (DRS). This value was transformed using the Box-Cox Method with an optimal lambda value of 0.35 achieving a near Gaussian distribution and is shown on the histogram in Box B. Boxes C-E display the predicted burden of delirium symptoms based on the NSQIP-SC and TMTB prediction model (Box C) and univariate analysis of NSQIP-SC (Box D) and TMTB (Box E). The statistics from each regression model are shown in the upper left hand corner of each box. The univariate NSQIP-SC regression model was analyzed with 97 participants. Due to one missing assessment of TMTB, Box C and E are analyzed with 96 participants.