JNS SPINE

CLINICAL ARTICLE

J Neurosurg Spine 36:32–41, 2022

Preoperative prediction of postoperative urinary retention in lumbar surgery: a comparison of regression to multilayer neural network

Ken Porche, MD, MA,[1] Carolina B. Maciel, MD, MSCR,[2] Brandon Lucke-Wold, MD, PhD,[1] Steven A. Robicsek, MD, PhD,[3] Nohra Chalouhi, MD,[1] Meghan Brennan, MD, MS,[3] and Katharina M. Busl, MD, MS[2]

[1]Lillian S. Wells Department of Neurosurgery, [2]Departments of Neurology and Neurosurgery, and [3]Department of Anesthesiology, University of Florida, Gainesville, Florida

OBJECTIVE Postoperative urinary retention (POUR) is a common complication after spine surgery and is associated with prolongation of hospital stay, increased hospital cost, increased rate of urinary tract infection, bladder overdistention, and autonomic dysregulation. POUR incidence following spine surgery ranges between 5.6% and 38%; no reliable prediction tool to identify those at higher risk is available, and that constitutes an important gap in the literature. The objective of this study was to develop and validate a preoperative risk model to predict the occurrence of POUR following routine elective spine surgery.

METHODS The authors conducted a retrospective chart review of consecutive adults who underwent lumbar spine surgery between June 1, 2017, and June 1, 2019. Patient characteristics, preexisting ICD-10 codes, preoperative pain and opioid use, preoperative alpha-1 blocker use, details of surgical planning, development of POUR, and management strategies were abstracted from electronic medical records. A binomial logistic model and a multilayer perceptron (MLP) were optimized using training and validation sets. The models' performance was then evaluated on model-naïve patients (not a part of either cohort). The models were then stacked to take advantage of each model's strengths and to avoid their weaknesses. Four additional models were developed from previously published models adjusted to include only relevant factors (i.e., factors known preoperatively and applied to the lumbar spine).

RESULTS Overall, 891 patients were included in the cohort, with a mean of 59.6 ± 15.5 years of age, 52.7% male, BMI 30.4 ± 6.4, American Society of Anesthesiologists class 2.8 ± 0.6, and a mean of 5.6 ± 5.7 comorbidities. The rate of POUR was found to be 25.9%. The two models were comparable, with an area under the curve (AUC) of 0.737 for the regression model and 0.735 for the neural network. By combining the two models, an AUC of 0.753 was achieved. With a regression model probability cutoff of 0.24 and a neural network cutoff of 0.23, maximal sensitivity and specificity were achieved, with specificity 68.2%, sensitivity 72.9%, negative predictive value 88.2%, and positive predictive value 43.4%. Both models individually outperformed previously published models (AUC 0.516–0.645) when applied to the current data set.

CONCLUSIONS This predictive model can be a powerful preoperative tool in predicting patients who will be likely to develop POUR. By using a combination of regression and neural network modeling, good sensitivity, specificity, and NPV are achieved.

https://thejns.org/doi/abs/10.3171/2021.3.SPINE21189

KEYWORDS postoperative complications; risk factors; lumbar surgery; urinary catheterization; urinary retention

More than 1.62 million spine surgeries occur annually in the United States.[1,2] Postoperative urinary retention (POUR) is a common complication affecting 6%–38% of patients who undergo elective spine surgery.[3–5] POUR significantly increases hospital length of stay (LOS) by 1–2 days and has been found to impede functional recovery.[3,5–9] The occurrence of POUR increases hospitalization costs by $3565–$4100 per patient ($2260 for additional hospital LOS, $1103 for costs associated with complications, and $202–$737 for other hospital costs).[4,7,9–16]

Currently, there are no reliable methods to identify

ABBREVIATIONS ASA = American Society of Anesthesiologists; AUC = area under the curve; BMI = body mass index; LOS = length of stay; MLP = multilayer perceptron; NPV = negative predictive value; POUR = postoperative urinary retention; PPV = positive predictive value; UTI = urinary tract infection.
SUBMITTED February 1, 2021. **ACCEPTED** March 9, 2021.
INCLUDE WHEN CITING Published online September 10, 2021; DOI: 10.3171/2021.3.SPINE21189.

those at increased risk for POUR among patients who are awaiting elective spine surgery. Although several attempted prediction models have been published, the strength of the proposed predictors varies across studies.[3,17–19] Practically, for preoperative prediction, a single model that can be run and interpreted would need to be built into the preoperative workflow.[3,17–19] The preoperative setting is preferred to postoperative, because it provides time to arrange prophylactic therapies and to avoid known triggers both intraoperatively and immediately postoperatively. Factors reported to be associated with POUR include older age, male sex, obesity, history of back injury or poisoning, history of urinary retention, benign prostate hyperplasia, urinary tract infection (UTI), surgical time > 3 hours, fusion surgery, delayed ambulation > 48 hours, postoperative thoracic epidural analgesia, increased fluid volume > 1500 mL given intraoperatively, higher pain scores, opioid use, and use of glycopyrrolate or other anticholinergic medications.[6,9,13,17–21]

Our hypothesis in this study was that using regressive analysis and neural network modeling, which have been shown to be useful machine learning predictive techniques, could both individually and in combination create a preoperative tool with good predictive power for POUR. The objective of this study was to develop and validate a preoperative tool that best predicts the occurrence of POUR.

Methods

Part 1: Study Design and Setting

This was a retrospective review approved by the local IRB and was aimed at developing an optimal model to predict POUR. The first part comprised a retrospective review of consecutive adult patients who underwent lumbar spine surgery between June 1, 2017, and June 1, 2019, at the University of Florida. Patients were excluded if they required emergency surgery, were < 18 years old, or had surgery in a nonlumbar region (thoracic or cervical). The second part comprised development of two machine learning techniques: a binomial logistic regression and an artificial neural network classification. These models were furthermore combined to optimize prediction strength.

Case Ascertainment

POUR was defined according to previous literature as reinsertion of a Foley catheter based on retention urine volume > 400 mL, or requiring straight catheterization for urine volumes > 400 mL.[4,22–27] Urine volume was determined per standard of care with nurse-led bladder scanning.

Clinical Variables

The patient characteristics, including all preexisting ICD-10 codes associated with the patient; age; sex; body mass index (BMI); preoperative opioid use (morphine, methadone, fentanyl, oxycodone, hydrocodone, meperidine); preoperative urinary retention medication use (tamsulosin, doxazosin); planned surgery specifics; and POUR, were collected and assessed. Hospital LOS was also recorded.

Part 2: Statistical Analysis

Identification of Predictors

Univariate tests were used on the entire set to discover factors to include in multivariate analyses, and a Bonferroni correction was used to correct for multiple analyses. Mann-Whitney U-tests were used for continuous and nominal variables, whereas chi-square tests were used for categorical variables. Kruskal-Wallis tests were used to compare training, validation, and testing sets. Hospital LOS was not included in the final models because this was an outcome measure.

Model Derivation

The data were split into training, validation, and testing sets using an approximately 65:10:25 ratio.[28,29] A binomial logistic model—which estimates the probability that an outcome is present given the values of explanatory variables and is typically used for classification—was formed with backward elimination based on significant changes in likelihood ratios, using a 0.10 cutoff.[30,31] All patient demographics and surgical characteristics were included in the model, but only comorbidities that had significant correlations with POUR were included ($p < 0.05$ corrected for multiple comparisons).

We used a multilayer perceptron (MLP) neural network architecture—attractive because it demonstrates an ability to learn salient features of the data on its own—which consisted of two hidden layers terminating at an output layer.[19] The first hidden layer consisted of 38 fully connected nodes, whereas the second layer consisted of 21 fully connected nodes. Hidden layers used a sigmoid activation function with no dropout. The output layer used an identity activation function and a sum of squares error function. The stopping rule was 1 consecutive step with no decrease in error based on the validation set.

Three additional regression models were developed from previously published models adjusted to include only relevant factors, to derive a pragmatic preoperative risk assessment tool (i.e., factors known preoperatively and applied to the lumbar spine).

Model Validation

Using the same training/validation/testing split on both the regression and neural network models, optimal models were selected by maximizing both adjusted R^2 and validation set accuracy. Performance of the chosen model was then evaluated on the testing set. Performance was measured on validation and testing sets combined because there was no need for a validation step.

Model Stacking

The models were combined such that if a threshold cutoff point (different for each model) was exceeded by either model, the classification was declared to be positive (i.e., the patient is predicted to develop POUR).[32] Outcomes from all cutoff points from 0.01 to 0.99 for each model were compared to maximize each individual outcome measure and combinations of outcome measures (i.e., average sensitivity and positive predictive value [PPV], average specificity and negative predictive value [NPV], aver-

TABLE 1. Demographics of overall and training/validation/testing splits in 891 patients who underwent lumbar spine surgery

Variable	Overall	Training Set	Validation Set	Testing Set	p Value
Frequency	231	150	22	59	—
POUR rate (%)	25.9	26.7	23.4	25.1	0.754
Age, yrs	59.6 ± 15.5	58.7 ± 15.5	57.6 ± 15.1	60.2 ± 15.7	0.282
Male sex (%)	52.7	55.3	45.7	49.4	0.109
BMI, kg/m^2	30.4 ± 6.4	30.3 ± 6.3	30.7 ± 7.5	30.7 ± 6.2	0.684
ASA class	2.8 ± 0.6	2.7 ± 0.5	3.7 ± 0.6	2.7 ± 0.6	**<0.001**
No. of comorbidities	5.6 ± 5.7	6.1 ± 6.1	2.7 ± 4.1	5.7 ± 4.9	**<0.001**

Unless otherwise indicated, values are expressed as the mean ± SD. Boldface type indicates statistical significance.

age sensitivity and specificity, average NPV and PPV, and the average of all outcome measures).

All statistical analyses were performed with SPSS version 23 (IBM Corp.). Cutoff points were discovered using code developed with Strawberry Perl 5.30.2.1.

Results

Clinical Characteristics

There were a total of 1311 patients who underwent elective spine surgery in the study period. Of those, 891 were included in the analysis, with 369 excluded because they were cervical or thoracic surgeries and 51 excluded due to missing data. POUR rates were found to be 25.9% in the entire cohort. Differences in patient demographics among the training/validation/testing split are shown in Table 1. The mean age was 59.6 ± 15.5 years, 52.7% were male, the mean BMI was 30.4 ± 6.4, the mean American Society of Anesthesiologists (ASA) class was 2.8 ± 0.6, and there was a mean of 5.6 ± 5.7 comorbidities. The training, validation, and testing sets did not differ significantly among age, sex, or BMI. The validation set had significantly higher ASA class and fewer comorbidities. Conversely, the training and testing sets did not vary significantly between these factors. Patient demographics and their designated univariate analyses are demonstrated in Fig. 1. Patients who developed POUR were significantly older than those who did not (62.5 ± 15.1 years vs 58.6 ± 15.5 years; $p = 0.0003$) and were more likely to use preoperative opioids (36.8% vs 25.6%; $p = 0.001$) or preoperative urinary retention medications (5.2% vs 2.4%; $p = 0.048$). Male sex (53.7% vs 52.4%; $p = 0.760$) and BMI (30.2 ± 5.9 vs 30.5 ± 6.6; $p = 0.682$) were not found to be significantly different between groups. Additionally, rates of POUR were found to be significantly higher in cases associated with 65 separate preoperative ICD-10 codes or groups of ICD-10 codes, including history of urinary retention or UTIs. The differences in POUR rates for these traits are demonstrated in Supplementary Fig. 1. The differences in POUR rates for all nonsignificant ICD codes are demonstrated in Supplementary Fig. 2. Hospital LOS for patients with POUR was significantly longer than for those without POUR (6.9 ± 9.5 days vs 2.7 ± 3.0 days; $p < 0.0001$).

Surgical Characteristics

The difference in POUR rates for individual surgical characteristics, as demonstrated in Fig. 2, revealed multiple surgical predictors of POUR. Rates of POUR were significantly lower in discectomies compared to patients whose spine surgery did not include discectomies (11.7% vs 30.7%; $p < 0.0001$) even as part of the operation. Rates of POUR were found to be higher overall in patients getting laminectomies (31.5% vs 17.4%; $p < 0.0001$) but not when only a laminectomy was performed (21.3% vs 27.6%; $p = 0.070$). Furthermore, rates were found to be significantly higher in multilevel laminectomies (34.5% vs 11.6%; $p < 0.0001$) and significantly lower in single-level laminectomies (11.6% vs 34.5%; $p < 0.0001$). Similarly, rates of POUR were found to be higher in surgeries with a fusion component (35.7% vs 16.7%; $p < 0.0001$), except for single-level fusions (24.6% vs 26.3%; $p = 0.705$). Within the scope of fusion surgeries, posterolateral fusions showed significantly higher rates of POUR (involvement: 39.3% vs 21.2%; alone: 41.2% vs 23.3%; $p < 0.0001$ for both), as did interbody fusions (32.9% vs 22.7%; $p = 0.001$) and pelvic screw placement (41.2% vs 25.0%; $p = 0.014$). Minimally invasive techniques demonstrated a significantly lower rate of POUR (16.1% vs 27.4%; $p = 0.009$). The average number of vertebral levels operated on was found to be 1.8 ± 1.8 in those who did not develop POUR and 2.9 ± 2.8 in those who did ($p < 0.0001$).

Outcomes

Binomial logistic multivariate model results are demonstrated in Table 2. Of the factors included in the model, only ICD-10 codes for diabetes (E11.9), abnormal heartbeat (R00), other general symptoms and signs (R68.89), altered mental status (R41.82), and screening for cardiovascular disorders (Z13.6) in addition to plans for a single laminectomy were found to be significant predictors of change in POUR. The ICD code for "other general symptoms and signs" and plans for only a single-level laminectomy were found to be significantly protective against POUR. For brevity, specific results of the neural network model were not included.

Comparison of the predictive outcomes of the two models and their combination is illustrated in receiver operating characteristic curves in Fig. 3. Table 3 reports the performance of the models on the training set, which can be viewed as the expected ceiling performance of the models.[33] The regression model, individually, achieved an area under the curve (AUC—an aggregate measure of performance across all possible classification thresholds) of

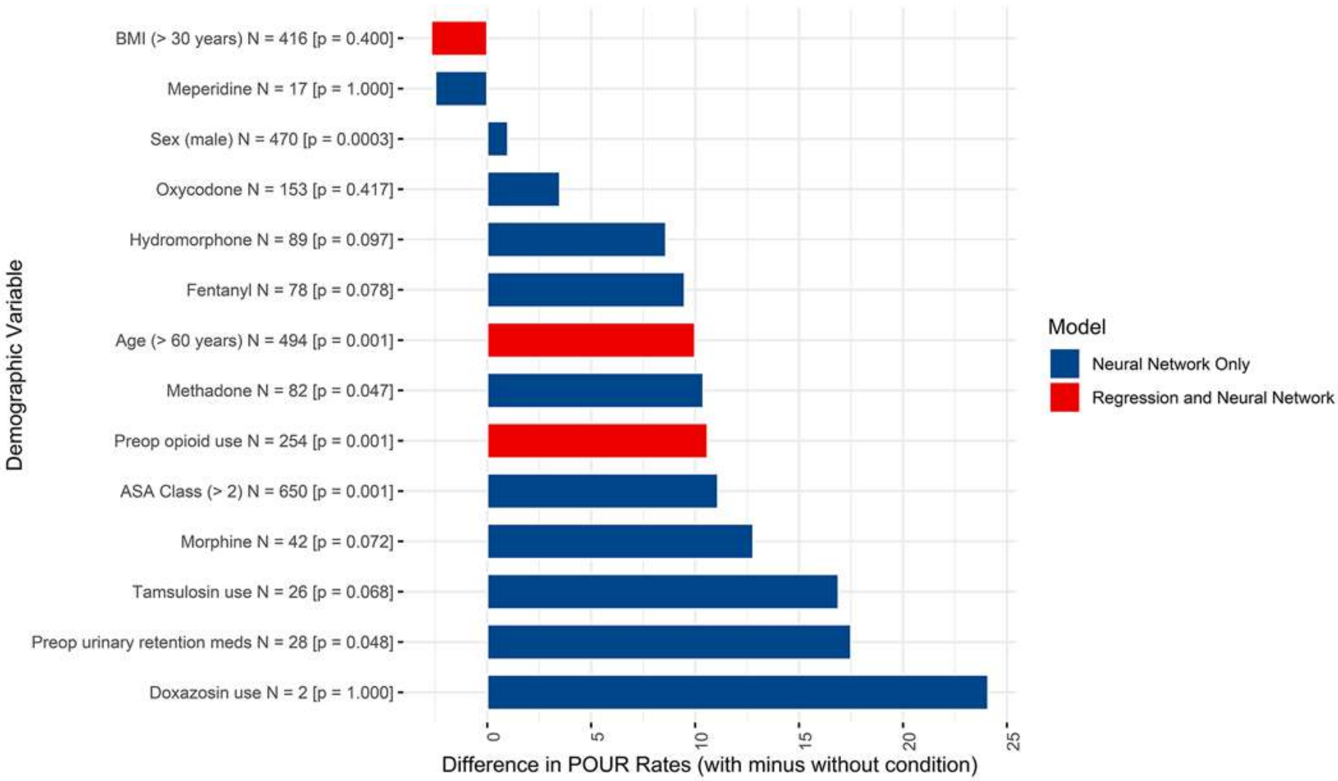

FIG. 1. Bar graph of the differences in rates of POUR based on patient demographics and medication use of the overall cohort (n = 891). Frequencies (N) and p values comparing those who did and did not develop urinary retention are listed in the label. *Red bars* = included in binomial logistic regression and neural network models. *Blue bars* = only included in neural network model. Figure is available in color online only.

0.737 (training set AUC 0.808); a probability cutoff of 0.34 maximized the average outcome parameters (specificity 85.2%, sensitivity 49.2%, NPV 83.3%, and PPV 52.7%). The neural network achieved an AUC of 0.735 (training set AUC 0.753); a probability cutoff of 0.21 maximized the average outcome parameters (specificity 54.5%, sensitivity 84.7%, NPV 91.4%, and PPV 38.5%). At the same cutoff point of 0.5, no significant differences between the models were noted in sensitivity, specificity, NPV, or PPV for the testing set, as shown in Table 3. When applying four previously published models to our data set, each provided high specificities (94.4%–98.8%) but low sensitivities (6.2%–7.4%); see Table 3 for detailed results. The AUC was lower than in our models, with a range of 0.516–0.645.

The stacking of the two models outperformed the individual models, with an AUC of 0.753. Optimal cutoff points for each model are demonstrated in Table 4. With the neural network alone and a cutoff of 0.14, sensitivity (96.6%) and NPV (96.1%) were simultaneously maximized, but with sacrifices in specificity (27.8%) and PPV (31.0%). With a regression model probability cutoff of 0.24 and a neural network cutoff of 0.23, sensitivity (72.9%) and specificity (68.2%) were simultaneously maximized, but with milder sacrifices in PPV (43.4%). With a regression model probability cutoff of 0.54 and a neural network cutoff of 0.43, all outcome parameters were simultaneously maximized (specificity 99.4%, sensitivity 15.3%, NPV 77.8%, and PPV 90.0%); however, sensitivity was severely sacrificed. Figure 4 demonstrates graphically how the cutoff points are used to conjoin the models such that a positive prediction is derived when both the regression model predicted probability is greater than 0.54 and the neural network model predicted probability is greater than 0.43 as in Fig. 4A, or greater than 0.24 and 0.23, respectively, as in Fig. 4B. A spreadsheet (POUR Prediction Tool) is available in the Supplementary Materials, which assists in calculating the probability of developing POUR for each individual model and for the combined models.

Discussion

In this study, in which the aim was to develop a high-performing prediction tool for POUR in patients undergoing elective spine surgery, our combined model of binomial logistic regression and neural network outperformed each individual candidate strategy as well as four previously published models with better AUC. POUR is a common cause of postoperative morbidity and discomfort for patients undergoing lumbar spine surgery and has been shown to significantly increase hospital LOS.[9,13,20,21] This was found to be consistent in our study; hospital LOS was increased by 4.3 days in the POUR cohort on average, uncorrected for complexity of surgeries performed. POUR has been variably and inconsistently associated with age, sex, obesity, operating time, fusion surgery, delayed ambulation, postoperative thoracic epidural analgesia, increased

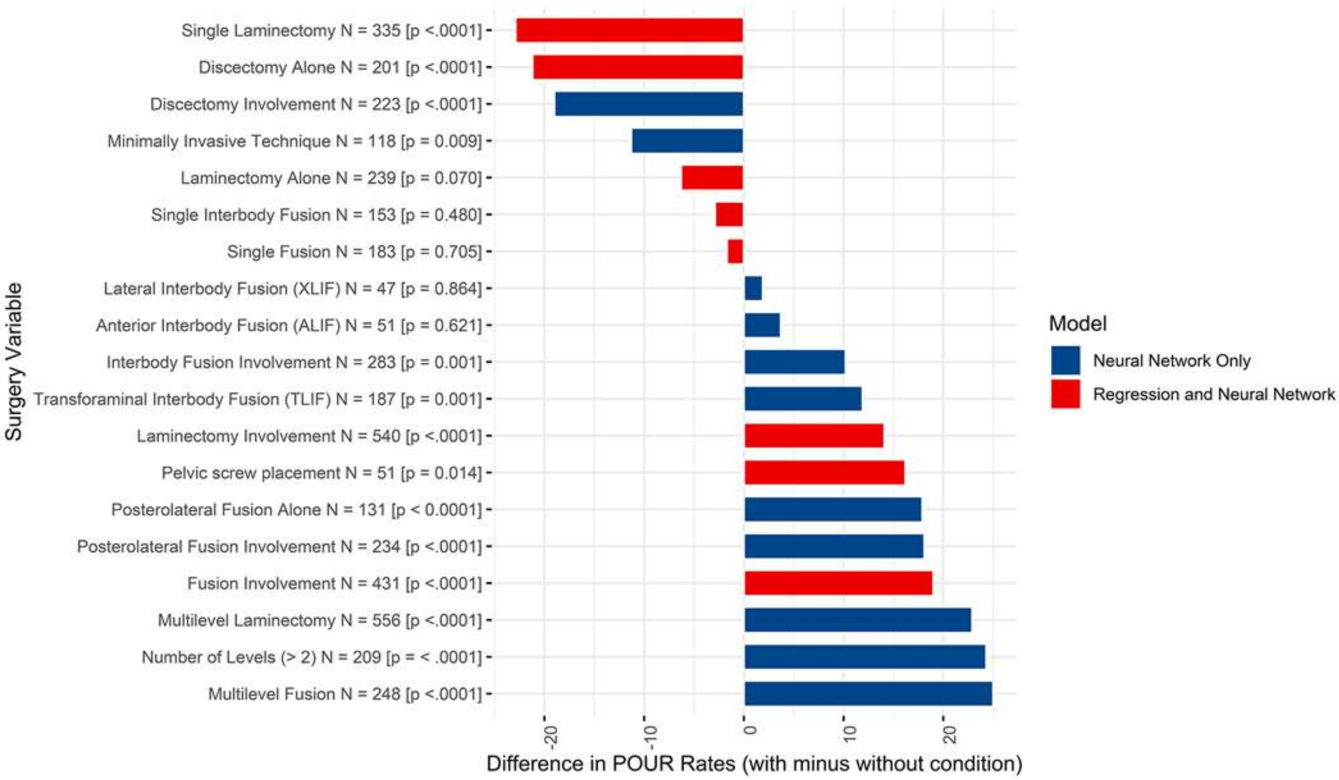

FIG. 2. Bar graph of the differences in rates of POUR based on planned surgical characteristics of the overall cohort (n = 891). Frequencies of the surgery type (N) and p values comparing those who did and did not develop urinary retention are listed in the label. *Red bars* = included in binomial logistic regression and neural network models. *Blue bars* = only included in neural network model. Figure is available in color online only.

fluid volume, higher pain scores, opioid consumption, glycopyrrolate use, history of urinary retention, benign prostate hyperplasia, and UTIs.[6,9,13,17,18,20,21] Although we did not find all these variables relevant, we found diabetes, heartbeat abnormalities, altered mental status, and prior screening for cardiovascular disorders to be significant predictors of POUR. General symptoms, not otherwise specified, and plans for a single laminectomy were found to be predictive of not developing POUR. Epidural analgesia is not commonly performed in the neurosurgery department at our institution, and was not performed on any of the patients in this study.

Preventive strategies for POUR exist but are associated with important risks—thus, refining patient selection for these interventions is a highly relevant clinical quest. Although intraoperative bladder catheter placement has been shown to reduce POUR incidence in elective spine surgeries, intraoperative catheters are commonly placed only in procedures lasting > 2–3 hours, because their use is associated with higher risk of UTI, increased operating room time, and increased patient discomfort (overall outweighing the benefit).[34] A modification of this approach, immediate postoperative bladder scans and bladder catherization if bladder volume is > 450 mL, can significantly reduce the number of urine catheter placements required postoperatively.[35] Furthermore, the initiation of opioid-sparing postoperative analgesia—specifically gabapentin—can reduce postoperative opioid consumption, which in turn may reduce the occurrence of POUR.[36] Despite these approaches enacted at our institution (including a stricter bladder scan cutoff of 400 mL), rates of POUR remain at 25.9%. Another successful option is a short course of detrusor relaxants such as alpha-1 antagonists, which are effective in reducing POUR in at-risk patients;[37] however, these are associated with significant adverse effects (i.e., hypotension, syncope, arrhythmia, hypersensitive syndrome, Stevens-Johnson syndrome, exfoliative dermatitis, priapism, and somnolence). Determining which patients are to be labeled at risk, who would qualify for prophylactic initiation of such targeted medication, remains to be elucidated.

In this study, we developed two preoperative risk assessment tools that could predict POUR in lumbar surgery: one based on regression modeling and one based on neural network modeling. The first, binomial logistic regression, estimates the probability that an outcome is present given the values of explanatory variables and is typically used for classification.[31] The second, an artificial neural network, is attractive because it demonstrates an ability to learn salient features of the data on its own as well as solve very complex problems.[19] Factors included in model development consisted of preoperative patient factors and planned surgical approaches, all of which are known to care providers prior to the surgical episode, hence allowing for feasibility of the use of the model preoperatively. The regression model is the simpler tool, requires only information easily available via medical re-

TABLE 2. Multivariate binomial logistic regression analysis for development of the POUR tool

Variable	Effect	SE	p Value	OR	95% CI for OR Lower	95% CI for OR Upper
Age (yrs)	0.002	0.008	0.780	1.002	0.986	1.019
Preop opioid use	−0.121	0.273	0.658	0.886	0.520	1.512
BMI (kg/m^2)	−0.006	0.019	0.759	0.994	0.958	1.032
Diabetes (E11.9)	0.953	0.485	**0.050**	2.593	1.001	6.713
Cardiomegaly (I51.7)	1.070	0.571	0.061	2.916	0.952	8.933
Hypotension (I95.9)	0.842	0.535	0.115	2.322	0.813	6.629
Ileus (K56.7)	1.240	0.844	0.142	3.455	0.660	18.077
Constipation (K59.00)	1.192	0.657	0.070	3.293	0.909	11.926
Other intestinal disease (K63.89)	0.725	0.653	0.267	2.064	0.574	7.420
Spondylolisthesis (M43.16)	0.408	0.312	0.191	1.504	0.815	2.772
UTI (N39.0)	1.405	1.001	0.161	4.075	0.573	28.996
Abnormalities of heartbeat (R00)	0.982	0.446	**0.028**	2.670	1.114	6.395
Other general symptoms & signs (R68.89)	−1.605	0.705	**0.023**	0.201	0.050	0.800
Altered mental status (R41.82)	3.013	1.186	**0.011**	20.356	1.990	208.226
Urinary retention (R33.9)	1.103	0.759	0.146	3.013	0.681	13.330
Pain (R52)	1.204	0.729	0.099	3.334	0.798	13.925
Encounter for other preprocedural exam (Z01.818)	0.464	0.494	0.348	1.591	0.604	4.190
Encounter for screening for cardiovascular disorders (Z13.6)	1.587	0.599	**0.008**	4.889	1.512	15.807
Persons w/ potential health hazards related to family & personal history & certain conditions influencing health status (Z77–Z99)	0.285	0.287	0.321	1.329	0.757	2.334
Planned laminectomy	0.479	0.265	0.071	1.614	0.960	2.714
Planned single fusion	−0.221	0.474	0.642	0.802	0.317	2.031
Planned pelvic screw	−0.540	0.467	0.247	0.583	0.233	1.454
Planned single laminectomy	−0.975	0.322	**0.003**	0.377	0.201	0.710
Planned single interbody fusion	−0.439	0.506	0.386	0.645	0.239	1.738
Constant	−1.583	0.889	0.075	0.205		

SE = standard error.
Boldface type indicates statistical significance.

cords, and was found to have an adequate AUC of 0.737. The neural network model was also found to have an adequate AUC of 0.735; however, its use entails listing all individual variables and performing matrix multiplication. Although it is a useful tool for prediction, the black-box nature of deep neural networks is a well-documented obstacle, and although much research and progress have been made on this front, it is outside of the scope of the current study to interpret the results of the model.[38] Stacking models (i.e., combining them) has been shown to optimize predictive power by leveraging the strengths of individual models and avoiding their weaknesses.[32,39] By combining

TABLE 3. Prediction outcomes for the POUR tool

Sets & Models	AUC	Specificity	Sensitivity	NPV	PPV
Training set (probability cutoff = 0.5)					
Regression	0.808	95.4	42.0	81.9	76.8
Neural network	0.753	94.4	26.0	77.8	62.9
Testing set (probability cutoff = 0.5)					
Regression	0.737	94.3	25.4	79.0	60.0
Neural network	0.735	94.9	20.3	78.0	57.1
Aiyer et al., 2018[18]	0.645	95.6	6.2	75.7	31.3
Mormol et al., 2021[17]	0.638	94.4	7.4	75.7	30.0
Nickerson et al., 2016[19]	0.559	98.8	6.2	76.3	62.5
Altschul et al., 2017[3]	0.516	95.6	7.4	76.0	35.3

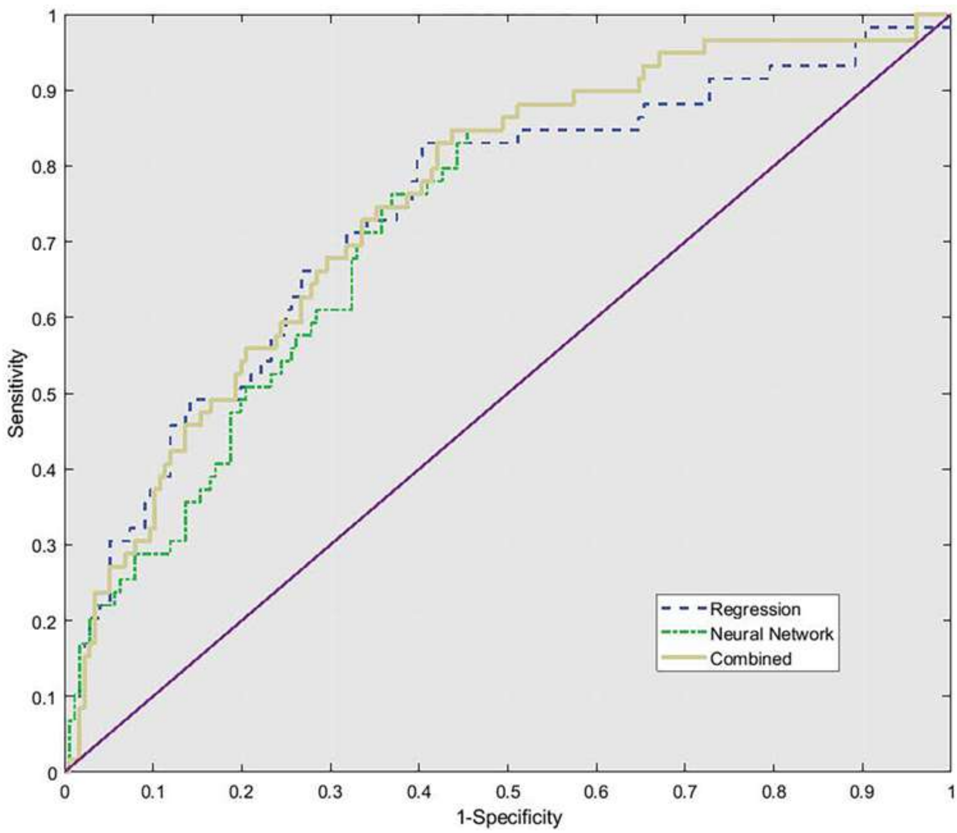

FIG. 3. Receiver operating characteristic curves for patients comparing regression model (*dashed blue line*), neural network model (*dashed-dotted green line*), and the stacked model (*solid yellow line*) combining the two. Figure is available in color online only.

the two models, a predictor tool with an AUC of 0.753 was created, making this a high-performing, acceptable tool.[40] Different cutoff points were furthermore revealed to optimize different predictive parameters; it is our opinion that cutoff points of 0.24 for the regression model and 0.23 for the neural network model, which maximize specificity and sensitivity, provide the most practical results. When compared to previously published models, each individual model and its combination outperform prior models, which had an AUC range of 0.516–0.645.[3,17–19]

Strengths of this study include the ability to separate the cohort into training, validation, and naïve testing sets. This afforded us the opportunity to validate our models as compared to reporting results that were strictly applied to the data from which the model was developed. Additionally, by combining the two predictive models, we were able to optimize predictive power by leveraging the strengths of each individual model and avoiding its weaknesses. Furthermore, by optimizing cutoff points, the reader is able to use the predictor tool to their individual specifications. As an example, one may wish to have a better specificity and NPV at the sacrifice of sensitivity and PPV. Last, the Excel document provided affords the reader an opportunity to easily use this complex model, with easy drop-down options and cut-paste forms.

Limitations of this study include its retrospective design and its untested external validity given that this study was conducted at a single institution. Because the model selection strategy was aimed at prediction and not interpretation, we cannot elaborate further on the meaning of specific predictors included in the models. Although no steps were taken to ameliorate overfitting (i.e., regulariza-

TABLE 4. Model cutoff points to maximize prediction outcomes for stacked (combined) model (AUC = 0.753)

Regression Cutoff	Neural Network Cutoff	Specificity (%)	Sensitivity (%)	NPV (%)	PPV (%)
—	0.14	27.8	96.6	96.1	31.0
0.54	0.43	99.4	15.3	77.8	90.0
0.24	0.23	72.9	68.2	88.2	43.4
0.54	0.43	99.4	15.3	77.8	90.0
0.54	0.43	99.4	15.3	77.8	90.0

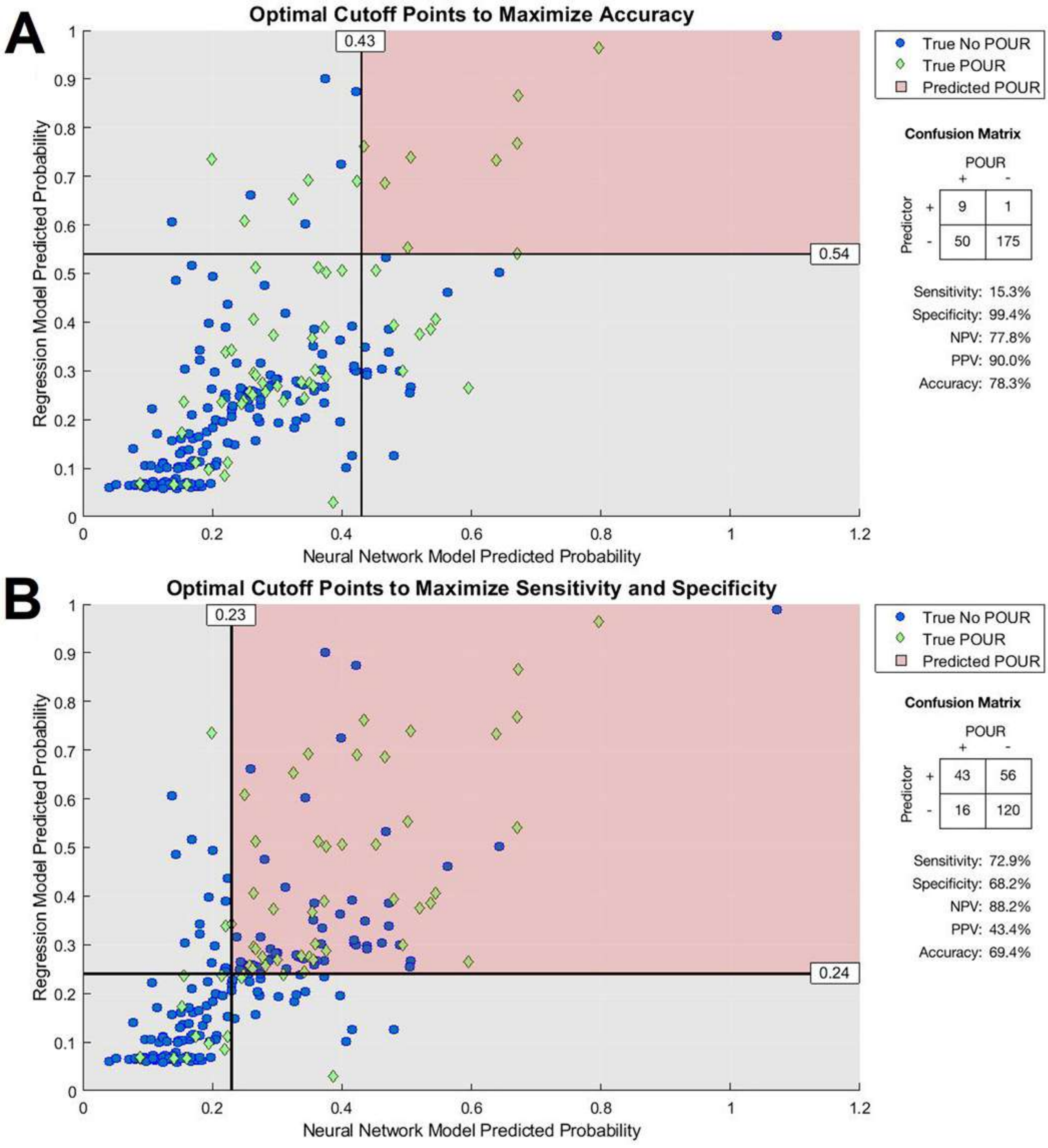

FIG. 4. Scatterplot for the testing set (N = 235) of the predicted probabilities of the regression model (y-axis) compared to the neural network model (x-axis) demonstrating (**A**) optimal cutoff points of 0.43 and 0.54, respectively, and (**B**) optimal cutoff points of 0.23 and 0.24, respectively, for combining the two models. By combining the two models using an "and" conjunction, the upper right quadrant (*red shaded area*) represents a positive test result. Thus, a *green diamond* in this area represents a true positive, whereas a *blue circle* in this area represents a false positive. A *green* or *blue* marker in the other quadrants represents a false negative or a true positive, respectively. Figure is available in color online only.

tion or dropout) in the neural network model, the notable similarity to its training set performance supports forgoing these efforts. The validation set was found to have a significantly higher ASA class and fewer comorbidities. However, this set was only used to train the models, and because the training and testing sets had no significant differences between these two factors, no further correction was deemed necessary. Due to an insufficient number of outpatient and 23-hour observations in the testing group, application of the prediction tool in this subpopulation was unable to be tested. This type of subanalysis will be easier to conduct with ongoing prospective studies.

Conclusions

Our predictive model can be a powerful preoperative tool in predicting patients who will be likely to develop POUR. By using a combination of regression and neural network modeling, good sensitivity, specificity, and NPV are achieved. Furthermore, our tool outperforms previously published models, with the caveat that these models were not designed to be used as a preoperative tool and originally incorporated intraoperative and postoperative factors that were removed for the purposes of this study. Ongoing and future efforts aim to validate the prediction tool in a prospective manner, broaden its focus to multiple institutions, apply the tool to test preventive strategies, and use its results to aid in discovery of a point-of-care test to enhance its predictive power.

Acknowledgments

Research reported in this publication was supported by the University of Florida Clinical and Translational Science Institute, which is supported in part by the NIH National Center for Advancing Translational Sciences under award no. UL1TR001427. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH.

References

1. Swann MC, Hoes KS, Aoun SG, McDonagh DL. Postoperative complications of spine surgery. *Best Pract Res Clin Anaesthesiol*. 2016;30(1):103–120.
2. How many spinal fusions are performed each year in the United States? iData Research. Accessed April 23, 2021. https://idataresearch.com/how-many-instrumented-spinal-fusions-are-performed-each-year-in-the-united-states/
3. Altschul D, Kobets A, Nakhla J, et al. Postoperative urinary retention in patients undergoing elective spinal surgery. *J Neurosurg Spine*. 2017;26(2):229–234.
4. Baldini G, Bagry H, Aprikian A, Carli F. Postoperative urinary retention: anesthetic and perioperative considerations. *Anesthesiology*. 2009;110(5):1139–1157.
5. Strickland AR, Usmani MF, Camacho JE, et al. Evaluation of risk factors for postoperative urinary retention in elective thoracolumbar spinal fusion patients. *Global Spine J*. 2021; 11(3):338–344.
6. Grass F, Slieker J, Frauche P, et al. Postoperative urinary retention in colorectal surgery within an enhanced recovery pathway. *J Surg Res*. 2017;207:70–76.
7. Boulis NM, Mian FS, Rodriguez D, et al. Urinary retention following routine neurosurgical spine procedures. *Surg Neurol*. 2001;55(1):23–28.
8. Balderi T, Mistraletti G, D'Angelo E, Carli F. Incidence of postoperative urinary retention (POUR) after joint arthroplasty and management using ultrasound-guided bladder catheterization. *Minerva Anestesiol*. 2011;77(11):1050–1057.
9. Cremins M, Vellanky S, McCann G, et al. Considering healthcare value and associated risk factors with postoperative urinary retention after elective laminectomy. *Spine J*. 2020;20(5):701–707.
10. Garg D, Agarwal A. Comment on "Early presentation of urinary retention in multiple system atrophy: can the disease begin in the sacral spinal cord?". *J Neurol*. 2020;267(3):665.
11. Agrawal K, Majhi S, Garg R. Post-operative urinary retention: review of literature. *World J Anesthesiol*. 2019;8(1): 1–12.
12. Mouchtouris N, Hines K, Fitchett EM, et al. Cost of postoperative urinary retention after elective spine surgery: significant variation by surgeon and department. *Neurosurgery*. 2020; 67(suppl 1):nyaa447_115.
13. Golubovsky JL, Ilyas H, Chen J, et al. Risk factors and associated complications for postoperative urinary retention after lumbar surgery for lumbar spinal stenosis. *Spine J*. 2018; 18(9):1533–1539.
14. Hospital adjusted expenses per inpatient day by ownership. KFF. Accessed April 23, 2021. https://www.kff.org/health-costs/state-indicator/expenses-per-inpatient-day-by-ownership/
15. Sullivan NM, Sutter VL, Mims MM, et al. Clinical aspects of bacteremia after manipulation of the genitourinary tract. *J Infect Dis*. 1973;127(1):49–55.
16. Estimating the additional hospital inpatient cost and mortality associated with selected hospital-acquired conditions. Agency for Health Research and Quality. Accessed April 23, 2021. https://www.ahrq.gov/hai/pfp/haccost2017-results.html
17. Mormol JD, Basques BA, Harada GK, et al. Risk factors associated with development of urinary retention following posterior lumbar spinal fusion: special attention to the use of glycopyrrolate in anesthesia reversal. *Spine (Phila Pa 1976)*. 2021;46(2):E133–E138.
18. Aiyer SN, Kumar A, Shetty AP, et al. Factors influencing postoperative urinary retention following elective posterior lumbar spine surgery: a prospective study. *Asian Spine J*. 2018;12(6):1100–1105.
19. Nickerson P, Tighe P, Shickel B, Rashidi P. Deep neural network architectures for forecasting analgesic response. *Annu Int Conf IEEE Eng Med Biol Soc*. 2016;2016:2966–2969.
20. Balabaud L, Pitel S, Caux I, et al. Lumbar spine surgery in patients 80 years of age or older: morbidity and mortality. *Eur J Orthop Surg Traumatol*. 2015;25(suppl 1):S205–S212.
21. Knight BA, Bayne AP, Zusman N, et al. Postoperative management factors affect urinary retention following posterior spinal fusion for adolescent idiopathic scoliosis. *Spine Deform*. 2020;8(4):703–709.
22. Petros JG, Bradley TM. Factors influencing postoperative urinary retention in patients undergoing surgery for benign anorectal disease. *Am J Surg*. 1990;159(4):374–376.
23. Petros JG, Rimm EB, Robillard RJ, Argy O. Factors influencing postoperative urinary retention in patients undergoing elective inguinal herniorrhaphy. *Am J Surg*. 1991;161(4): 431–434.
24. Petros JG, Rimm EB, Robillard RJ. Factors influencing urinary tract retention after elective open cholecystectomy. *Surg Gynecol Obstet*. 1992;174(6):497–500.
25. Petros JG, Mallen JK, Howe K, et al. Patient-controlled analgesia and postoperative urinary retention after open appendectomy. *Surg Gynecol Obstet*. 1993;177(2):172–175.
26. Petros JG, Alameddine F, Testa E, et al. Patient-controlled analgesia and postoperative urinary retention after hysterectomy for benign disease. *J Am Coll Surg*. 1994;179(6): 663–667.
27. Faas CL, Acosta FJ, Campbell MDR, et al. The effects of spinal anesthesia vs epidural anesthesia on 3 potential post-

operative complications: pain, urinary retention, and mobility following inguinal herniorrhaphy. *AANA J*. 2002;70(6): 441–447.
28. Larsen J, Goutte C. On optimal data split for generalization estimation and model selection. In: *Neural Networks for Signal Processing IX: Proceedings of the 1999 IEEE Signal Processing Society Workshop*. IEEE; 1999:225–234.
29. Draelos R. Best use of train/val/test splits, with tips for medical data. Glass Box. Published September 15, 2019. Accessed April 23, 2021. https://glassboxmedicine.com/2019/09/15/best-use-of-train-val-test-splits-with-tips-for-medical-data/
30. Bursac Z, Gauss CH, Williams DK, Hosmer DW. Purposeful selection of variables in logistic regression. *Source Code Biol Med*. 2008;3:17.
31. Harrell FE Jr. Binary logistic regression. In: Harrell FE Jr, ed. *Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*. 2nd ed. Springer Series in Statistics. Springer International Publishing; 2015:219–274.
32. Bisong E. Ensemble methods. In: Bisong E, ed. *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*. Apress; 2019:269–286.
33. Bisong E. Principles of learning. In: Bisong E, ed. *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*. Apress; 2019:171–197.
34. Meddings J, Skolarus TA, Fowler KE, et al. Michigan Appropriate Perioperative (MAP) criteria for urinary catheter use in common general and orthopaedic surgeries: results obtained using the RAND/UCLA Appropriateness Method. *BMJ Qual Saf*. 2019;28(1):56–66.
35. Hoke N, Bradway C. A clinical nurse specialist-directed initiative to reduce postoperative urinary retention in spinal surgery patients. *Am J Nurs*. 2016;116(8):47–52.
36. Turan A, Karamanlioğlu B, Memiş D, et al. Analgesic effects of gabapentin after spinal surgery. *Anesthesiology*. 2004; 100(4):935–938.
37. Madani AH, Aval HB, Mokhtari G, et al. Effectiveness of tamsulosin in prevention of post-operative urinary retention: a randomized double-blind placebo-controlled study. *Int Braz J Urol*. 2014;40(1):30–36.
38. Fan F, Xiong J, Li M, Wang G. On interpretability of artificial neural networks: a survey. *ArXiv*. Preprint posted online November 30, 2020. http://arxiv.org/abs/2001.02522
39. Pavlyshenko B. Using stacking approaches for machine learning models. In: *2018 IEEE Second International Conference on Data Stream Mining Processing (DSMP)*. IEEE; 2018:255–258.
40. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol*. 2010;5(9): 1315–1316.

Disclosures

The authors report no conflict of interest concerning the materials or methods used in this study or the findings specified in this paper.

Author Contributions

Conception and design: Porche, Maciel, Busl. Acquisition of data: Porche. Analysis and interpretation of data: Porche. Drafting the article: Porche, Lucke-Wold. Critically revising the article: Porche, Maciel, Busl. Reviewed submitted version of manuscript: Porche, Maciel, Lucke-Wold, Robicsek, Chalouhi, Busl. Approved the final version of the manuscript on behalf of all authors: Porche. Statistical analysis: Porche, Brennan. Administrative/technical/material support: Porche, Robicsek, Chalouhi, Busl. Study supervision: Robicsek, Busl.

Supplemental Information

Online-Only Content

Supplemental material is available with the online version of the article.

Supplementary Materials. https://thejns.org/doi/suppl/10.3171/2021.3.SPINE21189.

Correspondence

Ken Porche: University of Florida, Gainesville, FL. ken.porche@neurosurgery.ufl.edu.