Prediction of Postoperative Delirium in Geriatric Hip Fracture Patients: A Clinical Prediction Model Using Machine Learning Algorithms

Jacobien H. F. Oosterhoff, MD[1,2,3] , Aditya V. Karhade, MD, MBA[1],
Tarandeep Oberai, PT[3], Esteban Franco-Garcia, MD[4], Job N. Doornberg, MD, PhD[3,5],
and Joseph H. Schwab, MD, MS[1]

Abstract

Introduction: Postoperative delirium in geriatric hip fracture patients adversely affects clinical and functional outcomes and increases costs. A preoperative prediction tool to identify high-risk patients may facilitate optimal use of preventive interventions. The purpose of this study was to develop a clinical prediction model using machine learning algorithms for preoperative prediction of postoperative delirium in geriatric hip fracture patients. **Materials & Methods:** Geriatric patients undergoing operative hip fracture fixation were queried in the American College of Surgeons National Surgical Quality Improvement Program database (ACS NSQIP) from 2016 through 2019. A total of 28 207 patients were included, of which 8030 (28.5%) developed a postoperative delirium. First, the dataset was randomly split 80:20 into a training and testing subset. Then, a random forest (RF) algorithm was used to identify the variables predictive for a postoperative delirium. The machine learning-model was developed on the training set and the performance was assessed in the testing set. Performance was assessed by discrimination (c-statistic), calibration (slope and intercept), overall performance (Brier-score), and decision curve analysis. **Results:** The included variables identified using RF algorithms were (1) age, (2) ASA class, (3) functional status, (4) preoperative dementia, (5) preoperative delirium, and (6) preoperative need for mobility-aid. The clinical prediction model reached good discrimination (c-statistic = .79), almost perfect calibration (intercept = −.01, slope = 1.02), and excellent overall model performance (Brier score = .15). The clinical prediction model was deployed as an open-access web-application: https://sorg-apps.shinyapps.io/hipfxdelirium/. **Discussion & Conclusions:** We developed a clinical prediction model that shows promise in estimating the risk of postoperative delirium in geriatric hip fracture patients. The clinical prediction model can play a beneficial role in decision-making for preventative measures for patients at risk of developing a delirium. If found to be externally valid, clinicians might use the available web-based application to help incorporate the model into clinical practice to aid decision-making and optimize preoperative prevention efforts.

Keywords

geriatric trauma, hip fracture, delirium, clinical prediction model, machine learning, personalized medicine

[1]Department of Orthopaedic Surgery, Harvard Medical School, Massachusetts General Hospital, Boston, MA, USA
[2]Department of Orthopaedic Surgery, University of Amsterdam, Amsterdam Movement Sciences, Amsterdam University Medical Centers, Amsterdam, the Netherlands
[3]Department of Orthopaedics & Trauma Surgery, Flinders Medical Centre and Flinders University, Adelaide SA Australia
[4]Division of Palliative Care & Geriatric Medicine, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA

[5]Department of Orthopaedic Surgery, University of Groningen, University Medical Centre Groningen, Groningen, the Netherlands

Submitted August 26, 2021. Accepted November 2, 2021.

Corresponding Author:
Jacobien H. F. Oosterhoff, Yawkey Building 3, Massachusetts General Hospital, 55 Fruit Street, Boston, MA 02114, USA.
Email: joosterhoff@mgh.harvard.edu

Introduction

The most common complication following geriatric hip fracture surgery is a postoperative delirium.[1-5] Occurrence of a postoperative delirium is associated with reduced cognitive performance, other major postoperative complications, loss of functional independence, increased morbidity, higher patient costs, prolonged hospitalization, and decreased overall quality of life.[6] Patients developing postoperative delirium have 2.5 times greater costs than those without, with additional costs ranging $16,000 to $64,000 per patient-year.[7]

Delirium is a complex neuropsychiatric syndrome characterized by acute and fluctuating course, inattention, altered level of consciousness, and evidence of disorganized thinking.[8] The results of a recent meta-analysis examining risk factors for delirium showed that patients with existing cognitive impairment, advancing age, living in an institution, heart failure, total hip arthroplasty, multiple comorbidities, and morphine use were more likely to experience delirium after hip surgery.[9]

Preoperative risk stratification for delirium is needed to propose effective prevention measures. Previous studies have identified risk factors for delirium, but few have developed models for preoperative prediction of postoperative delirium. Numerous preventative strategies exist, including preoperative optimization of clinical condition, careful surgical and medical co-management, and appropriate postoperative support.[10,11] Despite this knowledge regarding prevention and management of this important and debilitating condition it is under-recognized and mis-managed, which is perturbing as delirium can be prevented with focused multicomponent interventions amongst at-risk hospitalized elderly patients.[12-14]

To our knowledge, there are no studies that have explored machine learning (ML) for prediction of delirium in the hip fracture population, although such methodology has shown promise for other outcomes in orthopedic surgery.[15-22] In orthopedic trauma, there is growing interest in applying ML as a statistical method for developing a clinical prediction model. The computer trains an existing human-created algorithm to recognize patterns in the data and optimizes their own performance.

The aim of this study was to develop a clinical prediction model using ML algorithms for preoperative prediction of postoperative delirium in geriatric hip fracture patients. The secondary aim of this study was to deploy the best performing clinical prediction model as an open access web-application.

Materials and Methods

Guidelines

The study was conducted according to the Transparent Reporting of Multivariable Prediction Models for Individual Prognosis or Diagnosis Guideline (TRIPOD-Statement).[23]

Data Source

We utilized the 2016 through 2019 American College of Surgeons (ACS) and American Association of Orthopaedic Surgeons (AAOS) National Surgical Quality Improvement Program (NSQIP) Hip Fracture Procedure Targeted files. National Surgical Quality Improvement Program is a large clinical database that collects more than 150 variables (pre-, peri-, and postoperatively) up to 30 days following surgery of more than 680 US hospitals combined. The series undergoes routine auditing, which ensures high-quality data with reported inter-reviewer rate of less than 2%.[24] Hip Fracture Procedure Targeted Files includes additional factors that are disease and procedure specific for hip fracture patients. These files were queried to identify patients older than 60 years of age who underwent surgery for femoral neck, intertrochanteric, and subtrochanteric hip fractures.

Primary Outcome—Outcome of Interest

The primary outcome was postoperative delirium within 30 days following hip fracture surgery as defined by the Hip Fracture Targeted ACS-NSQIP files. A chart-based method has been used for determining delirium; this method has been previously validated in surgical and non-surgical specialties.[25-28] Data abstractors from ACS-NSQIP routinely examine the entire medical record, and are instructed to assign postoperative delirium if the medical record consists phrases characteristic of an acute confusional state: "mental status change, confusion, disorientation, agitation, delirium, inappropriate behavior, inattention, hallucination, combative (eg, pulling out lines or tubes)".[25] The chart-based method has the advantage to detect delirium which may be missed when administrating with a standardized tool, for a delirium that may occur during nights or weekends. The first 2 h postoperative were excluded and not coded as delirium in order to separate from emergence delirium.[29]

Candidate Input Variables

Variables that were considered potentially important for predicting delirium included age (years), gender (female/male), body mass index (BMI) (kg/m^2), fracture type (femoral neck Garden types 1 and 2—non-displaced/femoral neck Garden type 3 and 4-displaced/intertrochanteric/subtrochanteric), procedure (ORIF femoral neck, ORIF inter/subtrochanteric with plate/screw implant, ORIF inter/subtrochanteric with intramedullary implant), American Society of Anesthesiologist (ASA) class (I/II/III/IV), functional status *(independent* defined as patients do not require assistance from another person for any activities of daily living/*partially depended* defined as patients requires some assistance from another person for activities of daily

living (including patients who utilize prosthetics, equipment, or devices but still requires some assistance from another person)/*totally dependent* if the patient requires total assistance for all activities of daily living), preoperative dementia as having cognitive impairment, dementia, or predefined descriptors consistent with dementia documented by a nurse or doctor stated (yes/no), preoperative delirium assessed by the chart-based method (yes/no), preoperative bone protection medication prescription (yes/no), preoperative need of mobility aid, for example, cane, walker, wheelchair, or scooter (yes/no), medical co-management (no/yes, co-management throughout stay/ yes, partial co-management during stay), standardized hip fracture protocol (yes/no), diabetes (insulin dependent/ non-insulin dependent/no), smoking (yes/no), dyspnea (at rest/moderate exertion/no), chronic obstructive pulmonary disorder (COPD) (yes/no), congestive heart failure (yes/ no), hypertension requiring medication (yes/no), acute renal failure (yes/no), dialysis (yes/no), disseminated cancer (yes/no), wound infection (yes/no), preoperative steroid use (yes/no), weight loss >10% body weight in last 6 months (yes/no), bleeding disorder (yes/no), transfusion in 72 hours prior surgery (yes/no), systemic sepsis within 48 hours prior surgery (none/SIRS/sepsis), sodium (mg/ dL), creatinine (mg/dL), white blood cell ($\times 10^3$/μL) and hematocrit (%), platelet ($\times 10^3$/μL). Categorical variables will be described as absolute numbers with frequencies, and continuous variables as medians with interquartile ranges (IQR).

Variable Selection

As a first step, variables potentially associated with postoperative delirium were identified using random forest (RF) algorithms with recursive selection.

Missing Data

After variable selection, variables with less than 30% missing values were imputed using multiple imputation with the missForest methodology.[30] Rates of missing data were as followed: BMI 4247 (14.2%), ASA class 45 (.2%), preoperative delirium 345 (1.2%), functional status 198 (.7%) and preoperative need for mobility-aid 1259 (4.2%). Of preoperative laboratory values the rates of missing data resulted: preoperative sodium 187 (.6%), preoperative creatinine 199 (.7%), white blood cell count 175 (.6%), hematocrit 144 (.5%), and platelet 200 (.7%).

Development of the Clinical Prediction Model

We trained and internally validated several ML algorithms based on prior research: Stochastic Gradient Boosting (SGM), RF, Support Vector Machine, Neural Network (NN) and Elastic-Net Penalized Logistic Regression (PLR). First, the dataset was randomly split in 80:20 into a training and testing subset. The training set (n = 22 563) was used for model training to predict postoperative delirium (ie, the outcome of interest). Then, we trained each model using 10-fold cross-validation in the training set, and the model performance was assessed the test set (n = 5641). Cross-validation means dividing the data into a selected number of groups, named folds. Results are subsequently averaged across all repetitions of this sequence.[31]

Model Performance

Model performance was evaluated according to a proposed framework for evaluation of a clinical prediction model[32] that includes: (1) discrimination with the c-statistic, (2) calibration with calibration slope and intercept (in line with the method by Cox[33]) and (3) the overall performance with the Brier score.

The c-statistic (area under the curve of a receiver operating characteristic curve) is a score ranging from .50 to 1.0 with 1.0 indicating the highest discrimination score and .50 indicating the lowest. The higher the discrimination score, the better the model's ability to distinguish patients who got the outcome from those who did not.[34]

A calibration plot plots the estimated vs the observed probabilities for the primary outcome. A perfect calibration plot has an intercept of 0 (<0 reflects overestimation, >0 reflects underestimating the probability of the outcome) and a slope of 1 (model is performing similarly in training and test-sets).[32,35] In a small dataset, slope is often <1 reflecting model overfitting; probabilities are too extreme (low probability too low, high probability too high).[34]

The Brier score calculates a composite of discrimination and calibration, with 0 indicating perfect prediction and a Brier score of 1 the poorest prediction.[32]

In addition, a decision curve analysis was undertaken and visualized to investigate the net benefit (weighted average of true positives and false positives) of the conducted algorithms over the range of predicted probabilities.[36]

The best model was deployed as an open-access web-application accessible on desktops, tablets, and smartphones.

Software

Data pre-processing and analysis were performed using R Version 5.3 ("R: A Language and Environment for Statistical Computing" The R Foundation, Vienna, Austria 2013) and R-studio Version 1.2.1335 (R-Studio, Boston, MA, USA).

Level of Evidence

Prognostic Level III

Table 1. Baseline Characteristics of NSQIP Hip Fracture Population, n = 28 207.

Variable	n (%) | Median (IQR)
Age (years)	
60+	3151 (11.2)
70+	6247 (22.1)
80+	11 691 (41.4)
90+	7118 (25.2)
Female sex	19 845 (70.4)
Body mass index (kg/m^2)	24.3 (21.6–27.3)
Fracture type	
Femoral neck fracture (sub capital, Garden types 1 and 2)-nondisplaced	2479 (8.8)
Femoral neck fracture (sub capital, Garden types 3 and 4)-displaced	8324 (29.5)
Intertrochanteric	15 761 (55.9)
Subtrochanteric	1642 (5.8)
ASA classification	
I	126 (.4)
II	4162 (14.8)
III	17 631 (62.5)
IV	6288 (22.3)
Functional status	
Independent	21 672 (76.8)
Partially dependent	5651 (20.0)
Totally dependent	884 (3.1)
Preoperative dementia	8668 (30.7)
Preoperative delirium	3714 (13.2)
Preoperative bone protective medication prescription	9047 (32.1)
Preoperative need for mobility aid	16 239 (57.6)
Preoperative pressure sore	971 (3.4)
Medical co-management	
No	3071 (10.9)
Yes-co-management throughout stay	20 576 (72.9)
Yes-partial co-management during stay	4560 (16.2)
Standardized hip fracture protocol	15 808 (56.0)
Diabetes	
Insulin dependent	2026 (7.2)
Non-insulin dependent	3035 (10.8)
No	23 146 (82.1)
Smoking	2848 (10.1)
Dyspnea	
At rest	273 (1.0)
Moderate exertion	1880 (6.7)
No	26 054 (92.4)
Chronic obstructive pulmonary disorder	3035 (10.8)
Congestive heart failure	1051 (3.7)
Hypertension requiring medication	19 120 (67.8)
Acute renal failure	150 (.5)
Dialysis	522 (1.9)
Disseminated cancer	407 (1.4)
Wound infection	1095 (3.9)
Preoperative steroid use	1476 (5.2)
Weight loss >10% body weight in last 6 months	384 (1.4)
Bleeding disorder	4796 (17.0)
Transfusion	1205 (4.3)

(continued)

Table 1. (continued)

Variable	n (%) | Median (IQR)
Systemic inflammatory response syndrome (SIRS)	
None	25 478 (90.3)
SIRS	2566 (9.1)
Sepsis	163 (.6)
Sodium (mg/dL)	138.0 (136.0–140.0)
Creatinine (mg/dL)	.88 (.70–1.13)
White blood cell (×10^3/μL)	9.60 (7.60–11.90)
Hematocrit (%)	35.0 (31.0–38.6)
Platelet (×10^3/μL)	196.0 (156.0–244.0)
Postoperative delirium	8030 (28.5)

n = number; IQR = interquartile range; ASA = American Society of Anesthesiologist.
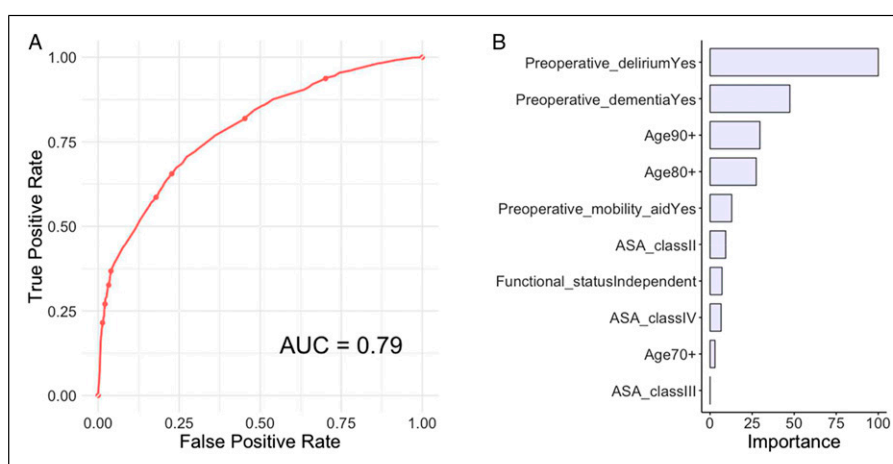
Figure 1. (A) Receiver operating curve and (B) global variable importance for the elastic-net penalized logistic regression for prediction of postoperative delirium in the testing set, n = 5641.

Results

In total, 28 207 patients underwent surgery for a hip fracture of which 8030 (28.5%) developed a delirium within 30 days after surgery (Table 1). The majority of patients was female (n = 19 845, 70.4%) and aged 80 years or older (n = 18 809, 66.6%).

Variables Selection

Variables selection identified the following predictive variables for prediction of postoperative delirium: (1) age, (2) ASA class, (3) functional status, (4) preoperative dementia, (5) preoperative delirium, and (6) preoperative need for mobility-aid (Figure 1).

Model Performance

The model performance of the conducted algorithms varied as measured by c-statistic from .67 to .79. The calibration slopes ranged from .44 to 1.02, and the calibration intercepts ranged from −.01 to 1.06. The Null model Brier score was .20, calculated based on an incidence of 28.5%. The Brier scores ranged from .15 to .18 (Table 2).

Although model performance metrics were comparable, except the RF algorithm, the PLR model had the most clinically meaningful variable importance and was therefore chosen as the best algorithm and therefore the final model (Figure 1).

Decision curve analysis of the PLR model revealed that decision changes based on the model outperformed not only the default strategies of decision change looking at all patients or no patients, but also that the decision changes based on presence of only a preoperative delirium alone (Figure 2).

Available Web-Application

The PLR algorithm was incorporated into a web-based application and deployed as an open-access available tool for clinicians: https://sorg-apps.shinyapps.io/hipfxdelirium/

Table 2. Machine Learning Model Performance Assessment in the Testing Set, n = 5641.

Metric	Stochastic Gradient Boosting	Random Forest	Support Vector Machine	Neural Network	Elastic-Net Penalized Logistic Regression
C-statistic	.79 (.77, .80)	.71 (.73, .77)	.67 (.68, .71)	.79 (.77, .80)	.79 (.77, .80)
Intercept	−.01 (−.08, .06)	1.06 (.95, 1.17)	−.01 (−.01, .01)	−.01 (−.08, .06)	−.01 (−.07, .06)
Slope	.97 (.91, 1.03)	.44 (.39, .49)	.92 (.85, .98)	.96 (.90, 1.02)	1.02 (.96, 1.09)
Brier	.15	.18	.16	.15	.15

Values are given in with the 95% confidence interval in parentheses. Null model Brier score = .20.
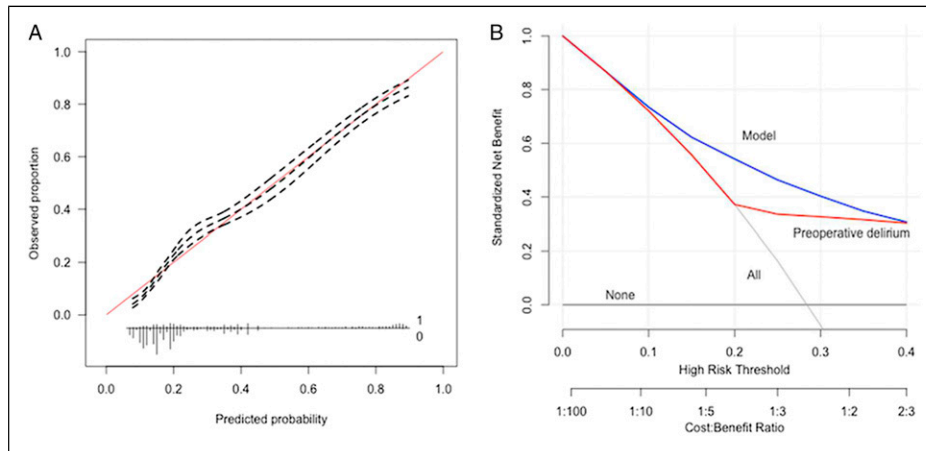

Figure 2. (A) Calibration plot and (B) decision curve analysis for the elastic-net penalized logistic regression for prediction of postoperative delirium in the testing set, n = 5641. Decision curve analysis with net benefit achieve by management changes based on the PLR algorithm relative to default strategies and for those based on solely presence of preoperative delirium.
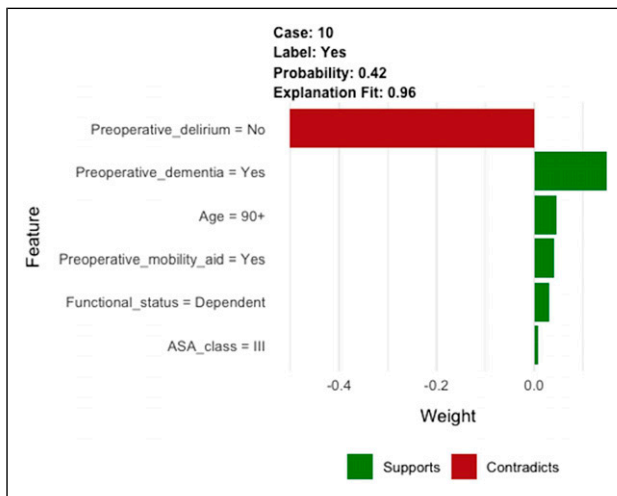

Figure 3. Example of individual patient-level explanation for postoperative delirium.

As an example, a patient older than 90 years of age is scheduled for hip fracture surgery. The patient is classified as ASA III, the patient lives dependent, and has preoperative need for mobile-aid. The patient has a history of dementia, but does not have preoperatively delirium. After filling out these values in the clinical prediction model in the available tool, this patient has a 42% chance of developing a delirium postoperatively. An explanation of the prediction is given by the available tool as shown in Figure 3, to overcome the "black box" of the developed model (How did this decision support tool came to this prediction?). In our example, factors increasing the likelihood were preoperative dementia, age, preoperative need for mobility-aid, dependent functional status, and ASA class. However, the lack of preoperative delirium reduced the likelihood of a delirium following hip fracture surgery.

Discussion and Conclusions

We have developed a clinical prediction model using ML algorithms for prediction of delirium in geriatric patients undergoing hip fracture surgery with almost perfect performance metrics. The clinical prediction model is based on patient characteristics and may guide clinicians to identify high-risk patients at a preoperative stage. Variables predictive for postoperative delirium were the following: age; ASA class; functional status; preoperative

dementia; preoperative delirium; and preoperative need for mobility-aid. The clinical prediction model with the best performance (PLR) showed good discrimination (c-statistic = .79), almost perfect calibration metrics (intercept = −.01, slope = 1.02), and good overall performance (Brier score = .15).

To the best of our knowledge, this is the first study applying ML algorithms for prediction of postoperative outcome following hip fracture surgery. In addition, this is the largest sample size thus far in predicting postoperative delirium in geriatric hip fracture surgery.[37] Predicting postoperative delirium is a challenging problem, but multicomponent intervention strategies can be effective in reducing rates and length of postoperative delirium.[13,38] The decision curve analysis allows for the comparison of selectively changing management for patients based on the clinical prediction model's predicted probabilities as compared to the default strategies of changing management for all patients or for no patients, or for patients screened based on the presence of postoperative delirium alone. In our opinion, the results of this study show that the developed ML prediction tool could play a beneficial role in decision-making for preventative measures for patients at risk of developing a delirium.

Our findings are in line with previous research. A meta-analysis from 2017 found moderate evidence for greater probability of the occurrence of postoperative delirium following hip fracture surgery for patients aged 80 years or older, higher ASA class (>2), functional status and presence of pre-admission diagnosis of dementia.[37] A similar conclusion was reached by studies published more recently, where increasing age and cognitive impairment were also found as predictors.[1,13,39] Mossello et al[40] have shown that moderate renal impairment (estimated by eGFR) was independently associated with the occurrence of delirium in older fracture patients (>75 years), based on serum creatinine and/or cystatin C in preoperative laboratory results.

When aiming to develop a prediction model that is applicable in daily practice, variables should be included in the trained algorithm that are readily available and use of definitions that are in line with daily practice should be followed. In this study, variables derived from variable selection are clinically readily available and in line with daily practice, where only preoperative creatinine level should be carried out which is mostly standardized collected prior surgery.

The results of this study should be viewed in light of several limitations. First, this was a retrospective study beholden to limitations inherent to such a research design. Second, the data was derived from the NSQIP Targeted Hip Fracture database and results may not be generalized to the (inter)national population. Third, the NSQIP Targeted Hip Fracture database does not provide information on patients treated with hemiarthroplasty or total hip arthroplasty which is often indicated in older and dependent patients and is associated with higher delirium rate.[9] This could lead to underestimation of the developed algorithm when applying to an independent population for prediction of postoperative delirium after hip fracture surgery in patients older than 60 years of age. Because of aforementioned limitations, external validation of the algorithm will support generalizability of results. Fourth, NSQIP collects data up to 30 days postoperatively and delirium may occur beyond this time-frame. Fifth, by nature of selection from participating ACS NSQIP hospitals, data may be subject to selection bias. Sixth, NSQIP is a large clinical database and collects more than 150 variables; however, the NSQIP does not account for specific factors such as preoperative use of medication and the use of cognitive screening instruments and/or interventions. Seventh, development of a prediction tool should be developed with the ability to enhance clinical care. Predicting patients at high risk is important, which may already be anticipated by the clinician, but maximum value may be obtained for those at moderate risk with the use of a prediction model.[41] However, treating all patients as at-risk with non-pharmacological and multicomponent interventions can do no harm, since use of medications to prevent delirium is not yet supported in prevention of developing a delirium following hip fracture surgery.[42] Eighth, preoperative risk stratification for delirium is needed to propose early effective prevention measures, although intraoperative and postoperative factors associated with postoperative delirium may be confounding with delirium prediction after surgery.[43] Lastly, current AUC of .79 is considered good, suggesting 79% chance that the model will correctly distinguish between patients who developed a postoperative delirium from those who did not.[44] In addition, the model showed almost perfect calibration which is considered more important in prognostic settings.[45] External validation is essential before testing and implementing in clinical practice. External validation can be carried out with temporal, geographical or fully independent validation.[46]

Conclusion

In summary, the developed SORG model effectively predicted delirium in geriatric hip fracture patients with good model performance. Our model would likely improve the efficiency of a screening program aimed to identify patients at risk for delirium since it outperformed the default strategy of screening all patients, screening no patients, and screening based on the presence of preoperative delirium alone. We will seek to externally validate our model in an independent data set.

ORCID iD

Jacobien H. F. Oosterhoff 🔘 https://orcid.org/0000-0002-3782-2791

References

1. Malik AT, Quatman CE, Phieffer LS, Ly TV, Khan SN. Incidence, risk factors and clinical impact of postoperative delirium following open reduction and internal fixation (ORIF) for hip fractures: An analysis of 7859 patients from the ACS-NSQIP hip fracture procedure targeted database. *Eur J Orthop Surg Traumatol*. 2019;29(2):435-446. doi:10.1007/s00590-018-2308-6
2. Brauer C, Morrison RS, Silberzweig SB, Siu AL. The cause of delirium in patients with hip fracture. *Arch Intern Med*. 2000; 160(12):1856-1860. doi:10.1001/archinte.160.12.1856
3. Holmes JD, House AO. Psychiatric illness in hip fracture. *Age Ageing*. 2000;29(6):537-546. doi:10.1093/ageing/29.6.537
4. Kyziridis TC. Post-operative delirium after hip fracture treatment - a review of the current literature. *Psycho Soc Med*. 2006;3:Doc01.
5. Edlund A, Lundstrom M, Brannstrom B, Bucht G, Gustafson Y. Delirium before and after operation for femoral neck fracture. *J Am Geriatr Soc*. 2001;49(10):1335-1340. doi:10.1046/j.1532-5415.2001.49261.x
6. Merchant RA, Lui KL, Ismail NH, Wong HP, Sitoh YY. The relationship between postoperative complications and outcomes after hip fracture surgery. *Ann Acad Med Singapore*. 2005;34(2):163-168.
7. Leslie DL, Marcantonio ER, Zhang Y, Leo-Summers L, Inouye SK. One-year health care costs associated with delirium in the elderly population. *Arch Intern Med*. 2008; 168(1):27-32. doi:10.1001/archinternmed.2007.4
8. Marcantonio ER. In the clinic. Delirium. *Ann Intern Med*. 2011; 154(11):6-16. doi:10.7326/0003-4819-154-11-201106070-01006
9. Yang Y, Zhao X, Dong T, Yang Z, Zhang Q, Zhang Y. Risk factors for postoperative delirium following hip fracture repair in elderly patients: A systematic review and meta-analysis. *Aging Clin Exp Res*. 2017;29(2):115-126. doi:10.1007/s40520-016-0541-6
10. Zaki H-AE, Mousa SM, El Said SMS, Mortagy AK. Morbidity and mortality following surgery for hip fractures in elderly patients. *J Aging Res*. 2019;2019:7084657. doi:10.1155/2019/7084657
11. Inouye S, Robinson T, Blaum C. Postoperative delirium in older adults: Best practice statement from the American Geriatrics Society. *J Am Coll Surg*. 2015;220(2):136-148.e1. doi:10.1016/j.jamcollsurg.2014.10.019
12. Bohlken J, Kostev K. Prevalence and risk factors for delirium diagnosis in patients followed in general practices in Germany. *Int psychogeriatrics*. 2018;30(4):511-518. doi:10.1017/S1041610217002587
13. Oberai T, Laver K, Crotty M, Killington M, Jaarsma R. Effectiveness of multicomponent interventions on incidence of delirium in hospitalized older patients with hip fracture: A systematic review. *Int Psychogeriatrics*. 2018;30(4):481-492. doi:10.1017/S1041610217002782
14. Power C, Duffy R, Bates H, et al. The detection, diagnosis, and impact of cognitive impairment among inpatients aged 65 years and over in an Irish general hospital - a prospective observational study. *Int psychogeriatrics*. 2017;29(11): 1879-1888. doi:10.1017/S1041610217001326
15. Topol EJ. High-performance medicine: The convergence of human and artificial intelligence. *Nat Med*. 2019;25(1): 44-56. doi:10.1038/s41591-018-0300-7
16. Panch T, Szolovits P, Atun R. Artificial intelligence, machine learning and health systems. *J Glob Health*. 2018;8(2): 20303. doi:10.7189/jogh.08.020303
17. Tran B, Vu G, Ha G, et al. Global evolution of research in artificial intelligence in health and medicine: A Bibliometric Study. *J Clin Med*. 2019;8(3):360. doi:10.3390/jcm8030360
18. Fontana MA, Lyman S, Sarker GK, Padgett DE, MacLean CH. Can machine learning algorithms predict which patients will achieve minimally clinically important differences from total joint arthroplasty? *Clin Orthop Relat Res*. 2019;477(6):1267-1279. doi:10.1097/CORR.0000000000000687

19. Ogink PT, Karhade AV, Thio QCBS, et al. Development of a machine learning algorithm predicting discharge placement after surgery for spondylolisthesis. *Eur Spine J*. 2019;28(8): 1775-1782. doi:10.1007/s00586-019-05936-z

20. Karhade AV, Thio QCBS, Ogink PT, et al. Development of machine learning algorithms for prediction of 30-day mortality after surgery for spinal metastasis. *Clin Neurosurg*. 2019;85(1):E83-E91. doi:10.1093/neuros/nyy469

21. Shah AA, Karhade AV, Bono CM, Harris MB, Nelson SB, Schwab JH. Development of a machine learning algorithm for prediction of failure of nonoperative management in spinal epidural abscess. *Spine J*. 2019;19(10):1657-1665. doi:10.1016/j.spinee.2019.04.022

22. Bongers MER, Thio QCBS, Karhade AV, et al. Does the SORG algorithm predict 5-year survival in patients with chondrosarcoma? An external validation. *Clin Orthop Relat Res*. 2019;477(10):2296-2303. doi:10.1097/corr.0000000000000748

23. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *BMJ*. 2015;350:g7594. doi:10.1136/bmj.g7594

24. Ingraham AM, Cohen ME, Bilimoria KY, et al. Association of surgical care improvement project infection-related process measure compliance with risk-adjusted outcomes: Implications for quality measurement. *J Am Coll Surg*. 2010;211(6):705-714. doi:10.1016/j.jamcollsurg.2010.09.006

25. Berian JR, Zhou L, Russell MM, et al. Postoperative Delirium as a target for surgical quality improvement. *Ann Surg*. 2018; 268(1):93-99. https://journals.lww.com/annalsofsurgery/Fulltext/2018/07000/Postoperative_Delirium_as_a_Target_for_Surgical.16.aspx.

26. Mitasova A, Kostalova M, Bednarik J, et al. Poststroke delirium incidence and outcomes: Validation of the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU). *Crit Care Med*. 2012;40(2):484-490. doi:10.1097/CCM.0b013e318232da12

27. Hestermann U, Backenstrass M, Gekle I, et al. Validation of a German version of the Confusion Assessment Method for delirium detection in a sample of acute geriatric patients with a high prevalence of dementia. *Psychopathology*. 2009; 42(4):270-276. doi:10.1159/000224151

28. McNicoll L, Pisani MA, Ely EW, Gifford D, Inouye SK. Detection of delirium in the intensive care unit: Comparison of confusion assessment method for the intensive care unit with confusion assessment method ratings. *J Am Geriatr Soc*. 2005;53(3):495-500. doi:10.1111/j.1532-5415.2005.53171.x

29. Card E, Pandharipande P, Tomes C, et al. Emergence from general anaesthesia and evolution of delirium signs in the post-anaesthesia care unit. *BJA Br J Anaesth*. 2014;115(3): 411-417. doi:10.1093/bja/aeu442

30. Stekhoven DJ, Buhlmann P. MissForest–non-parametric missing value imputation for mixed-type data. *Bioinformatics*. 2012; 28(1):112-118. doi:10.1093/bioinformatics/btr597

31. Ogink PT, Karhade AV, Thio QCBS, et al. Predicting discharge placement after elective surgery for lumbar spinal stenosis using machine learning methods. *Eur spine J Off Publ Eur Spine Soc Eur Spinal Deform Soc Eur Sect Cerv Spine Res Soc*. 2019;28(6):1433-1440. doi:10.1007/s00586-019-05928-z

32. Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: A framework for traditional and novel measures. *Epidemiology*. 2010;21(1): 128-138. doi:10.1097/EDE.0b013e3181c30fb2

33. Cox DR. Two further applications of a model for binary regression. *Biometrika*. 1958;45(3/4):562-565. doi:10.2307/2333203

34. Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: Seven steps for development and an ABCD for validation. *Eur Heart J*. 2014;35(29):1925-1931. doi:10.1093/eurheartj/ehu207

35. Van Calster B, Vickers AJ. Calibration of risk prediction models: Impact on decision-analytic performance. *Med Decis Making*. 2015;35(2):162-169. doi:10.1177/0272989X14547233

36. Vickers AJ, Elkin EB. Decision curve analysis: A novel method for evaluating prediction models. *Med Decis Making*. 2006;26(6):565-574. doi:10.1177/0272989X06295361

37. Smith TO, Cooper A, Peryer G, Griffiths R, Fox C, Cross J. Factors predicting incidence of post-operative delirium in older people following hip fracture surgery: A systematic review and meta-analysis. *Int J Geriatr Psychiatry*. 2017; 32(4):386-396. doi:10.1002/gps.4655

38. Lundstrom M, Olofsson B, Stenvall M, et al. Postoperative delirium in old patients with femoral neck fracture: A randomized intervention study. *Aging Clin Exp Res*. 2007; 19(3):178-186. doi:10.1007/bf03324687

39. Arshi A, Lai WC, Chen JB, Bukata SV, Stavrakis AI, Zeegen EN. Predictors and Sequelae of Postoperative Delirium in Geriatric Hip Fracture Patients. *Geriatr Orthop Surg Rehabil*. 2018;9:2151459318814823. doi:10.1177/2151459318814823

40. Mossello E, Rivasi G, Tortù V, et al. Renal function and delirium in older fracture patients: Different information from different formulas? *Eur J Intern Med*. 2019;71:70-75. doi:10.1016/j.ejim.2019.10.022

41. Lindroth H, Bratzke L, Purvis S, et al. Systematic review of prediction models for delirium in the older adult inpatient. *BMJ Open*. 2018;8(4):e019223. doi:10.1136/bmjopen-2017-019223

42. Marcantonio ER, Flacker JM, Wright RJ, Resnick NM. Reducing Delirium after hip fracture: A randomized trial. *J Am Geriatr Soc*. 2001;49(5):516-522. doi:10.1046/j.1532-5415.2001.49108.x

43. Susano MJ, Scheetz SD, Grasfield RH, et al. Retrospective analysis of perioperative variables associated with postoperative delirium and other adverse outcomes in older patients after spine surgery. *J Neurosurg Anesthesiol*. 2019;31(4):385-391.

https://journals.lww.com/jnsa/Fulltext/2019/10000/Retrospective_Analysis_of_Perioperative_Variables.7.aspx.

44. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol.* 2010;5(9): 1315-1316. doi:10.1097/JTO.0b013e3181ec173d

45. Goff DC, Lloyd-Jones DM, Bennett G, et al. 2013 ACC/ AHA guideline on the assessment of cardiovascular risk: A report of the American College of Cardiology/American Heart Association Task force on practice guidelines. *J Am Coll Cardiol.* 2014;63(25, Part B):2935-2959. doi:10.1016/ j.jacc.2013.11.005

46. Steyerberg EW. Validation of prediction models. In: *Clinical Prediction Models. A Practical Approach to Development, Validation, and Updating.* Springer; 2019:309-323.