

## Original Paper

# Predicting Postoperative Delirium in Older Patients Before Elective Surgery: Multicenter Retrospective Cohort Study

Shun-Chin Jim Wu<sup>1,2,3\*</sup>, MD, MSc; Nitin Sharma<sup>1,2\*</sup>, MSc; Anne Bauch<sup>1,2</sup>, PhD; Hao-Chun Yang<sup>1,2</sup>, PhD; Jasmine L Hect<sup>4</sup>, BSc; Christine Thomas<sup>1,5</sup>, MD; Sören Wagner<sup>6,7</sup>, MD; Bernd R Förstner<sup>8</sup>, PhD; Christine A F von Arnim<sup>9,10</sup>, MD; Tobias Kaufmann<sup>1,2,11</sup>, PhD; Gerhard W Eschweiler<sup>1,2,12</sup>, MD; Thomas Wolfers<sup>1,2,11</sup>, PhD; PAWEL Study Group<sup>1,2\*</sup>

<sup>1</sup>Department of Psychiatry and Psychotherapy, University of Tübingen, Tübingen, Germany

<sup>2</sup>German Center for Mental Health (DZPG), partner site Tübingen, Tübingen, Germany

<sup>3</sup>Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom

<sup>4</sup>Department of Neurological Surgery, University of Pittsburgh School of Medicine, Pittsburgh, PA, United States

<sup>5</sup>Department of Geriatric Psychiatry and Psychotherapy, Klinikum Stuttgart, Stuttgart, Germany

<sup>6</sup>Department of Anesthesiology, Klinikum Stuttgart, Stuttgart, Germany

<sup>7</sup>Department of Anesthesiology and Intensive Care, School of Medicine and Health, Klinikum Rechts der Isar, Technical University of Munich, Munich, Germany

<sup>8</sup>Social and Preventive Medicine, Department of Sports and Health Sciences, University of Potsdam, Potsdam, Germany

<sup>9</sup>Department of Geriatrics, University Medical Center Göttingen, Göttingen, Germany

<sup>10</sup>German Center for Cardiovascular Research (DZHK), Göttingen, Germany

<sup>11</sup>Norwegian Centre for Mental Disorders Research, University of Oslo, Oslo, Norway

<sup>12</sup>Geriatric Center, Universitätsklinikum Tübingen, Tübingen, Germany

\*these authors contributed equally

## Corresponding Author:

Shun-Chin Jim Wu, MD, MSc

Department of Psychiatry

University of Cambridge

Cambridge CB2 1TN

United Kingdom

Phone: 44 07423706114

Email: [jim81098@gmail.com](mailto:jim81098@gmail.com)

## Abstract

**Background:** Elective surgeries for older adults are increasing. Machine learning could enhance risk assessment, influencing surgical planning and postoperative care. Preoperative cognitive assessment may facilitate early detection and management of postoperative delirium (POD).

**Objective:** This study aims to assess machine learning models' predictive ability for POD, focusing on the added predictive value of the neuropsychological assessments before elective surgery.

**Methods:** This retrospective cohort study analyzed data from the multicenter PAWEL (Patient safety, Efficiency and Life quality in elective surgery) and PAWEL-R (risk) studies, encompassing older patients ( $\geq 70$  y) undergoing elective surgeries from July 2017 to April 2019. A total of 1624 patients (52.3% male,  $N=850$ ; age: mean 77.9, SD 4.9 years) were included, with a POD diagnosis made before discharge. Sociodemographic, clinical, surgical, and neuropsychological features were collected pre- and intraoperatively by care providers. Machine learning models' performance was evaluated using the area under the receiver operating characteristic curve (AUC), with permutation testing for significance, and Shapley Additive Explanations to identify effective neuropsychological assessments.

**Results:** Predicting POD before surgery with a random forest model achieved an AUC of 0.760. Incorporating all pre- and intraoperative features into the model yielded a slightly higher AUC of 0.783, with no statistically significant difference observed ( $P=.24$ ). While cognitive factors alone were not strong predictors ( $AUC=0.617$ ), specific tests within neuropsychological assessments, such as the Montreal Cognitive Assessment and Trail Making Tests, showed high feature attribution and played a crucial role in further enhancing prediction before surgery.

**Conclusions:** Preoperative risk prediction for POD can increase risk awareness in presurgical assessment and improve perioperative management in older patients at a high risk for delirium.

**Trial Registration:** German Clinical Trials Register (Germany) DRKS12797; <https://drks.de/search/en/trial/DRKS00012797/details> and DRKS13311; <https://drks.de/search/en/trial/DRKS00013311/details>

**International Registered Report Identifier (IRRID):** RR2-10.1186/s13063-018-3148-8

*JMIR Aging* 2025;8:e67958; doi: [10.2196/67958](https://doi.org/10.2196/67958)

**Keywords:** postoperative delirium; predictive model; preoperative cognition; geriatrics; elective surgery

## Introduction

In an aging society, there is a rising demand for elective surgeries due to the changing health care needs of older people [1-3]. However, this increase in elective surgeries raises concerns about additional adverse outcomes, particularly given the unique challenges posed by aging, such as pre-existing health conditions, disease sequelae, and diminished physiological reserves [4,5]. Meeting the growing demand for elective surgeries among older adults requires a comprehensive strategy, including thorough preoperative assessments, personalized care plans, and continuous postoperative support [6-8].

Postoperative delirium (POD), characterized by acute and fluctuating inattention with alterations in thinking or consciousness after surgery, affects 12% to 51% of older patients, with incidence varying by surgical procedures and regions [9]. POD in older patients is linked to heightened rehospitalization rates, persistent postoperative cognitive dysfunction, increased incidence of dementia, and elevated mortality [10-12]. Some studies have shown significant associations between POD and factors collected before or during surgery (pre- or intraoperatively) [13,14]. These factors could aid in predicting the POD [15,16]. However, different pre- and intraoperative feature categories may vary in their importance for POD risk prediction. Sociodemographic factors, such as age and sex, are critical for assessing POD risk [17-24]. Clinical data, including blood samples and chronic disease medication, is also predictive [25-27]. The type and duration of surgery and anesthesia have been shown to be indispensable for POD prediction [18,28]. Although preoperative neuropsychological assessments, like the Mini-Mental State Examination (MMSE) and others [29-31], help identify at-risk patients for early risk mitigation [32-34], these evaluations are not yet incorporated into clinical routine despite experts' recommendations [35]. The early identification of POD predictors enables clinicians to proactively assess risks to mitigate the occurrence of POD [36] and might affect patients' decisions before nonemergent surgery. Therefore, identifying preoperative predictors for POD is crucial in facilitating personalized surgical risk assessment before surgery [37,38]. Moreover, examining POD risks remains challenging due to its multifactorial origins [18,22,39]. The precise categories of preoperative and intraoperative features with superior predictive capabilities for POD remain unknown and unvalidated [40].

In this study, we (1) predict POD by using machine learning approaches using diverse pre- and intraoperative features from a large multicenter cohort of older patients and (2) conduct a comparative analysis of independent models using various predictor categories, including sociodemographic, clinical, surgical, and neuropsychological features. By featuring one of the largest cohorts to date, this multicenter study provides an extensive assessment of POD predictors in both cardiac and noncardiac surgeries through machine learning techniques. This approach can enhance presurgical risk assessment and perioperative management in older patients.

## Methods

### Participants

Building on previous research [13,41,42], this study comprehensively investigates the predictive capabilities of various feature categories of the entire cohort from the PAWEL (Patientensicherheit, Wirtschaftlichkeit und Lebensqualität bei elektiven Operationen, English: Patient safety, Efficiency and Life quality in elective surgery) study, originally recruited between November 21, 2017, and April 12, 2019. In addition, it also includes additional participants from the add-on PAWEL-R study (R for risk estimation), which extended the recruitment period from July 11, 2017, to January 15, 2019. By incorporating both datasets, our study expanded the cohort size, strengthening the predictive model's reliability. Machine learning models were used to predict POD, departing from original statistical non-cross-validated methods in the previous reports [13]. Patients were recruited from 5 major medical institutions in Germany (3 university hospitals: Tübingen, Freiburg, and Ulm, and 2 tertiary medical centers: Stuttgart and Karlsruhe). The study adheres to CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) [43] and STROBE (Strengthening the Reporting of Observational studies in Epidemiology) [44] guidelines (the CONSORT-EHEALTH checklist is provided in [Checklist 1](#)). This study was approved by ethical votes under the PAWEL-R study, and ethical approval (Ethical Committee 233/2017BO1) was provided by the Ethical Committee of Tübingen University Hospitals, Tübingen, Germany on June 6, 2017.

Participants included patients aged 70 years and older undergoing elective surgery (joint, spine, vessels, heart, lung, abdomen, urogenital system, and other organs) with an

expected surgical duration exceeding 60 minutes. Exclusion criteria covered patients unable to communicate effectively in German, those undergoing emergency surgery, severe dementia with MMSE <15 or Montreal Cognitive Assessment (MoCA) <8, or an estimated survival time of less than 15 months. A stepped-wedge cluster randomized design was used for equitable intervention group allocation [41,42]. The study, initially including 1631 patients, conducted thorough postoperative assessments within 1 week after surgery but before discharge. To ensure accuracy in predicting POD diagnosis, patients without sufficient information for POD diagnosis before discharge (N=7) were excluded, resulting in 1624 patients included in prediction models (Figure S1). The study reported 23.1% of patients diagnosed with POD, with detailed between-group comparisons available in Tables S1 and S2.

## Measures

The PAWEL and PAWEL-R studies extensively evaluated elective surgery patients to identify risk factors and outcomes related to POD. Diagnosis involved the Confusion Assessment Method (I-CAM) algorithm and chart review within the first postoperative week or until discharge. Features were categorized into 4 feature groups: sociodemographic, neuropsychological, clinical, and surgical information.

Sociodemographic data, including age, sex, education, alcohol and smoking habits, living arrangements, and hospital location, were collected preoperatively. Neuropsychological assessments such as MoCA, Trail Making Test (TMT) parts A and B, digit span backward, Subjective Memory Impairment (SMI), and Patient Health Questionnaire-4 (PHQ-4) were conducted during admission. Clinical profiles included blood samples, past medical histories (including pre-existing mild or moderate dementia and previous delirium history), preoperative and intraoperative medication dose (including benzodiazepine, neuroleptics, opiates, and propofol), polypharmacy, multimorbidity, the mininutritional assessment, American Society of Anesthesiologists (ASA) physical status classification system, Charlson comorbidity index, Barthel index, clinical frailty scale, and sensory impairments (auditory, visual, and others). Surgical information covered types and duration of surgery and anesthesia, cardiopulmonary bypass, and intraoperative physiological changes. Preoperative and intraoperative features were defined to assess predictive performance achieved at 2 distinct time points: before and during surgery. Additional details can also be found in Table S1.

To identify effective predictors for POD diagnosis, the specific and combined predictive capacity of sociodemographic, clinical, and surgical features before (preoperative) and during surgery (intraoperative) were analyzed. Models were developed and tested with both features and preoperative features alone. In addition, models were compared with and without neuropsychological assessments to examine their contribution to predictive performance.

## Preprocessing, Imputation, and Features

Two features with excessive missingness were excluded, namely, albumin level (979/1624, 60.28%) and depth of anesthesia (1224/1624, 75.37%) were excluded [16]. Information about missing values is available in Table S3. Data imputation involved using multiple imputation (IterativeImputer) for continuous variables and random sampling from the original probability distribution for discrete and binary variables within cross-validation folds for all models, except for gradient boosting, which does not require imputation [45]. Continuous and discrete variables were scaled within cross-validation folds for all models. Potential outliers or data entry errors in clinical assessments were identified when a data point exceeded 5 SD from the mean and were subsequently imputed using the above method.

Blood samples, including hemoglobin, sodium levels, and C-reactive protein (CRP), were interpreted and discretized following clinical guidelines from Harrison's Principles of Internal Medicine, 20th version, in line with standard practice [46]. Specifically, hemoglobin levels of less than 12 g/dL were considered indicative of anemia, while values of 12 g/dL or higher were considered normal. Serum sodium levels were categorized as hyponatremia if below 135 mmol/L and hypernatremia if above 145 mmol/L, with values within this range considered normal. Similarly, CRP levels greater than 3 mg/L indicated elevated levels in adults aged 65 years and older [47], while values within the normal range were not considered clinically significant. Redundant features with a perfect correlation were excluded; therefore, use of heart-lung machine was excluded for its perfect correlation with cardiopulmonary bypass. Nonbinary categorical features, including location, SMI, types of anesthesia and surgery, and transformed blood samples were one-hot encoded because they had no natural ordinal relationship among their categories, and assigning numerical labels to them could introduce bias or incorrect assumptions in the model.

Fifty-nine features were used, divided into preoperative (48 features) and intraoperative (11 features) categories. Preoperative features encompassed sociodemographic (7 features), clinical (20 features), surgical categories (6 features), and neuropsychological assessments (15 features). Intraoperative features included clinical (4 features) and surgical (7 features) categories. Different combinations of features were compared, including preoperative only, preoperative and intraoperative, and each category of preoperative features. The study evaluated model performance, feature category effectiveness, and the additional benefits of preoperative neuropsychological assessments.

The study aimed to develop a prediction model in a naturalistic setting. Information regarding interventions was included only as a sensitivity analysis to demonstrate its potential impact on the prediction model. The potential imbalance of the dataset for all models was tested by oversampling with the Synthetic Minority Oversampling Technique [48]. Given that the overall sex distribution in the

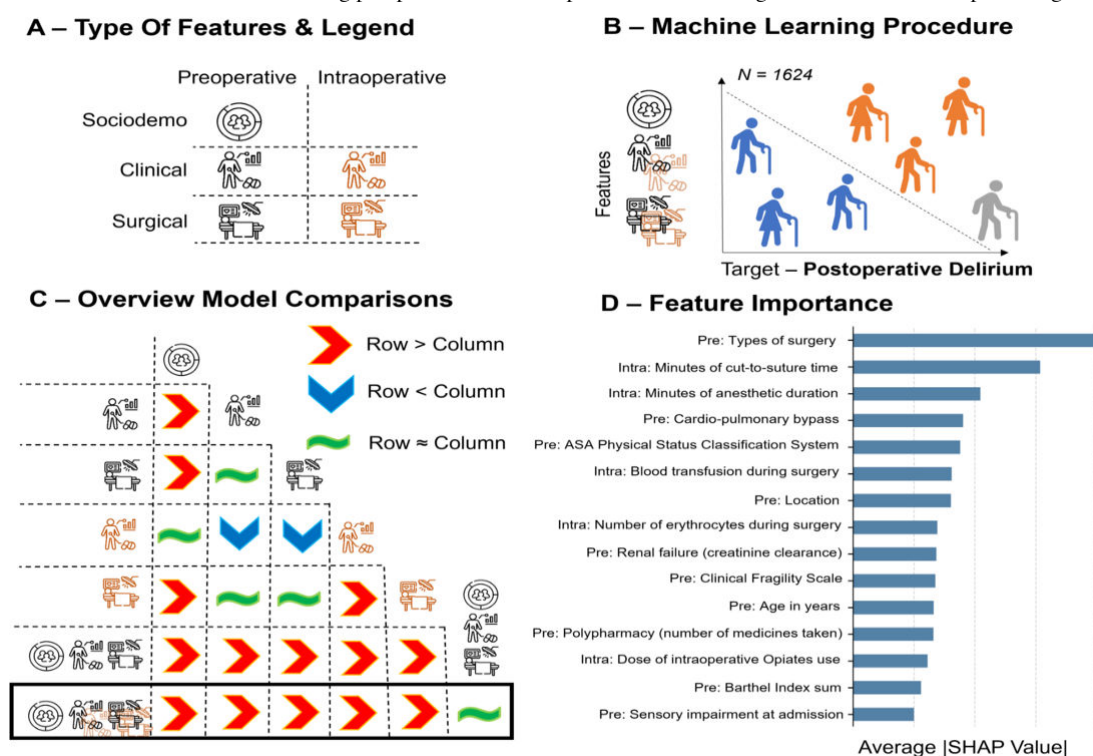
cohort was fairly balanced (774/1624, 47.66% female), we did not apply additional rebalancing for sex to avoid potential bias. In addition, no significant age differences were observed between patients with and without POD. Various sensitivity analyses described were performed.

## Machine Learning Models, Performance Evaluation, and Feature Importance

Machine learning models were used to predict POD, including logistic regression, support vector machines, random forest, and gradient boosting without hyperparameter tuning using the scikit-learn library version 1.2.2 [49] and the Xgboost library version (1.7.3) [50]. Independent variables were feature variables, while the dependent variable was POD diagnosis, as illustrated in Figure 1B. Feature selection was performed using the SelectFromModel function from scikit-learn, leveraging model-based feature selection, to retain the most predictive variables. Given the sample size relative to the complexity of the models, we did not anticipate substantial improvements from hyperparameter tuning. However, we conducted additional sensitivity analyses to assess its impact on model performance via nested cross-validation (Tables S12-S15 in Multimedia Appendix 1). To ensure model stability and interpretability, we used default parameters from the software for the primary analysis

(Table S11 in Multimedia Appendix 1). 5-fold cross-validation, with balanced labels across folds, measured model performance at testing using the area under the receiver operating characteristic curve (AUC) as the primary metric. In order to robustly evaluate model performance against chance and between models, permutation testing was used to assess whether the AUC of each model was significantly greater than expected by chance. Specifically, POD diagnostic labels were randomly shuffled, and AUC values were recalculated 1000 times to generate a null distribution. The *P* value was derived by comparing the observed AUC to this null distribution (Table S4 in Multimedia Appendix 1). Furthermore, to compare AUC differences between models, permutation testing was also used to generate a null distribution of AUC differences by randomly permuting the labels 1000 times while preserving data dependencies. The *P* value for model comparison was computed based on the observed difference relative to this null distribution [51,52]. Lastly, to obtain a robust measure of variability in model performance, 95% CI for AUC using bootstrapping with 1000 resamples was estimated. Additional metrics included precision, recall, sensitivity, specificity, balanced accuracy, and area under the precision-recall curve presented in Table S5 in Multimedia Appendix 1.

**Figure 1.** Individual feature categories perform worse than combined feature categories pre-, intraoperatively, and postoperative delirium (POD) can effectively be predicted prior to surgery (see back box). (A) The symbols are consistently used in this study and represent the different types of features. (B) A random forest classifier with 5-fold cross-validation was used to forecast POD onset in 1624 older individuals. Among them, 23.1% experienced POD. Orange indicates patients with POD, blue represents those without POD, and gray signifies patients predicted to have POD by the algorithm. (C) A comparative analysis of different feature combinations based on their respective area under the receiver operating characteristic curve (AUC) values, calculated over 1000 permutations was conducted. Combined feature sets consistently outperformed individual feature types. (D) Shapley Additive Explanations (SHAP) was used to assess feature importance. The feature importance was based on the average absolute value of the aggregated SHAP values across folds. Among preoperative and intraoperative features, surgical data is critical for predicting POD.





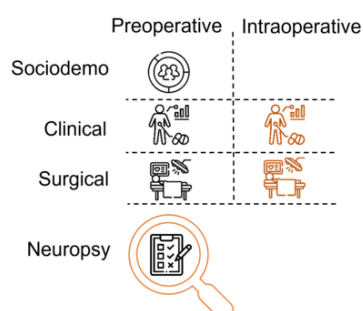
The Shapley Additive Explanations (SHAP) values were used to assess feature importance. Positive SHAP values increased the probability of POD, while negative values decreased it. SHAP values were computed across all cross-validation folds and aggregated over 5-fold cross-validation with 5 shuffles to ensure stability and robustness in feature importance rankings. To facilitate explanation of the main models, we presented the top 15 features contributing to model performance and complete feature attributions in major feature sets: preoperative features with and without neuropsychological assessments, and combined preoperative and intraoperative features with and without neuropsychological assessments (Figures 1D and 2D, Figure S2 in Multimedia Appendix 1, and Tables S7-S10 in Multimedia Appendix 1). The SHAP

library version 0.41.0 in Python was used [53]. To evaluate the calibration of our models, calibration plots were generated (applying Platt scaling if needed), and the number of patients with high-confidence positive predictions ( $>0.9$ ) and high-confidence negative predictions ( $<0.1$ ) was calculated for each model. This analysis was conducted across different feature sets to assess whether the addition of features affected the model's probability estimates.

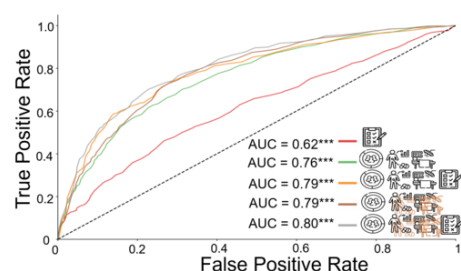
All preprocessing steps and analyses are available on GitHub upon publication from Sharma forked to our laboratory's page, and data can be requested by addressing the PAWEL consortium as indicated in the data availability section.

**Figure 2.** Cognition, in isolation, demonstrates limited predictive power. However, when combined with preoperative features, the performance matches that of all pre- and intraoperative features (see back box). (A) The predictive capability of preoperative neuropsychological tests was assessed for postoperative delirium (POD) individually and in combination with other features. (B) Our analysis, as depicted in the receiver operating characteristic curve, demonstrated the successful prediction of POD by preoperative neuropsychological tests. However, when combined with pre- and intraoperative data, the performance gains were only marginal. Asterisks (\*) denote the Area Under the receiver operating characteristic Curve (AUC) values, which were assessed against chance permutation testing, revealing a significance ( $**P \leq .001$ ). (C) The AUC values of neuropsychological tests were compared to those integrated into the pre- and intraoperative features. Adding preoperative neuropsychological tests to the preoperative model improves POD prediction to a level comparable to the best-performing model, similar to the effect of incorporating intraoperative features. In contrast, using preoperative features alone without neuropsychological tests results in lower performance. (D) Specific neuropsychological tests, such as the Montreal Cognitive Assessment (MoCA) and Trail Making Tests (TMTs), were particularly influential in POD prediction.

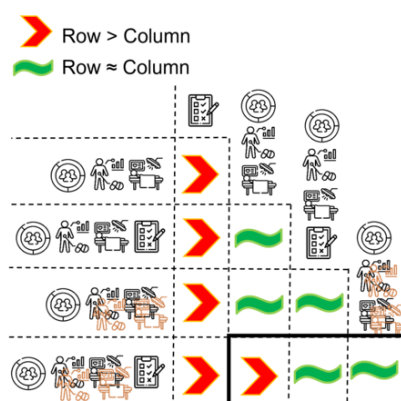
#### A – Type Of Features & Legend



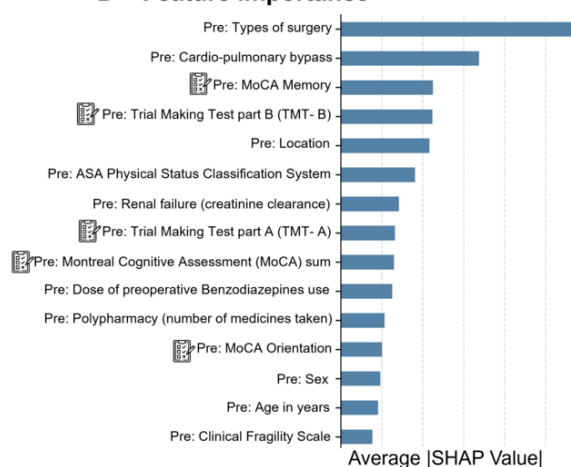
#### B – The Role Of Cognition



#### C – Overview Model Comparisons



#### D – Feature Importance



## Ethical Considerations

This study was approved by the Ethics Commission of the Faculty of Medicine of the University of Tübingen (233/2017BO1, October 12, 2017) and the Ethics

Commission of the University of Potsdam (38/2017, December 11, 2017), and was registered in the German Clinical Trials Register (DRKS00012797, July 2017).

All participants provided written informed consent for participation and for the publication of any potentially identifiable data. Data were pseudonymized, stored securely, and handled in accordance with applicable privacy and confidentiality regulations. Participants were not financially compensated but were reimbursed for travel expenses when applicable.

## Results

### *Predicting POD With Combined, Pre-, and Intraoperative Features*

The models incorporating combined and independent pre- and intraoperative features exhibited robust performance, as evidenced by AUC values surpassing chance levels, all with  $P < .002$  (Table S4 in [Multimedia Appendix 1](#)) and Table S5 in [Multimedia Appendix 1](#) displays the performance of these models through receiver operating characteristic curves with the random forest. The combined model using only preoperative features (Pre-Op) achieved an AUC of 0.760, comparable to a model incorporating both pre- and intraoperative features (Pre- and Intra-Op) with an AUC of 0.783, showing no statistically significant difference ([Figure 1C](#) and Table S6 in [Multimedia Appendix 1](#)). Independent models exclusively using preoperative clinical, preoperative surgical, or intraoperative surgical features also demonstrated high AUC values (0.691, 0.664, and 0.670, respectively) without statistically significant difference. Notably, key predictors of POD included surgery type and cardiopulmonary bypass (preoperative surgical; recruiting hospital and age (preoperative sociodemographic); ASA status, Clinical Frailty Scale, Barthel Index, polypharmacy, and creatinine clearance (preoperative clinical); cut-to-suture time, anesthesia duration, and blood loss (intraoperative surgical) shown in [Figure 1D](#) and [Figure S2-A](#) in [Multimedia Appendix 1](#). Further details, including all pairwise model comparisons and  $P$  values for differences in AUC values, can be found in Table S6 in [Multimedia Appendix 1](#). A detailed list of the feature importance can be found in Tables S7-S8 in [Multimedia Appendix 1](#).

### *Addition of Preoperative Neuropsychological Assessments*

The model using preoperative neuropsychological assessments exclusively exhibited AUC values of 0.617. Integrating neuropsychological assessments into the model, using both pre- and intraoperative features (Pre- and Intra-Op+NeuroPsy), led to a slight improvement in the AUC, reaching 0.803 ([Figure 2B](#) and Table S6 in [Multimedia Appendix 1](#)). Adding neuropsychological assessments to the preoperative model (Pre-Op+ NeuroPsy) improved the AUC from 0.760 to 0.787, which resulted in a model that matched the above best-performing model ([Figure 2C](#) and Table S6 in [Multimedia Appendix 1](#)). Specifically, the MoCA scores and TMTs before surgery were important for predicting POD, as illustrated in [Figure 2D](#) and [Figure S2-B](#) in [Multimedia Appendix 1](#). For detailed pairwise comparisons of AUC values and corresponding  $P$  values, please refer to Table

S6 in [Multimedia Appendix 1](#). A detailed list of the feature importance can be found in Tables S9-S10 in [Multimedia Appendix 1](#).

### *Robustness of Predictive Models in Performance Evaluation*

Evaluating the performance of 4 classifiers with tuned hyperparameters revealed comparable results (Tables S12-S15 in [Multimedia Appendix 1](#)). The random forest model, demonstrating marginally better performance in AUC values in models with combined features, was highlighted above. In addition, random forest models demonstrated good calibration, as indicated by a consistent distribution of high-confidence predictions across different feature sets. The inclusion of intraoperative and neuropsychological information did not compromise calibration, ensuring the model's reliability in distinguishing high- and low-risk patients ([Figure S3](#) in [Multimedia Appendix 1](#)).

Including intervention allocation information had no discernible impact on predictive performance (Tables S5 and S16 in [Multimedia Appendix 1](#)). Finally, there was no statistically significant difference in AUC between models with and without oversampling (Table S17 in [Multimedia Appendix 1](#)).

## Discussion

### *Principal Findings*

Leveraging machine learning, we predicted the occurrence of POD after elective surgeries through a combination of preoperative and intraoperative features with a large multicenter cohort, achieving an AUC above 0.8. This performance exceeds that of traditional scoring systems such as the Delirium Risk Assessment Tool, Delirium Risk Assessment Score, and Delirium Elderly At-Risk, which have demonstrated AUC values between 0.5 and 0.7 in a large cohort [54-56]. Surgical information both before and during the surgery was critical in predicting POD. Integrating neuropsychological tests into preoperative features enhanced the AUC to a level comparable to the best-performing model, effectively replacing intraoperative features for predicting POD before surgery. This improvement was primarily driven by MoCA and TMTs A and B, as elucidated by model explanations. This integrated analysis improves conventional clinical risk profiling, furnishing superior predictive capacity with promising implications for surgical planning in the era of machine learning-assisted health care and empowering the prioritization of pivotal features in future work.

### *Critical Surgical Information*

Surgical features, both preoperative and intraoperative, emerged as good solo predictors category for POD ([Figure 1C](#)). While baseline clinical profiles, including ASA physical status, renal function, frailty, and polypharmacy, are important, intraoperative surgical features such as individual surgery and anesthesia duration and factors related to blood loss were particularly important for predicting POD

(Figure 1D and Figure S2-A in [Multimedia Appendix 1](#)). In addition, preoperative surgical information such as types of surgery and use of cardiopulmonary bypass were also associated with a higher risk of POD, aligning with a higher overall incidence of cardiac surgery in patients with POD relative to those without POD (206/375, 54.9% vs 264/1249, 21.1% as indicated in Table S1 in [Multimedia Appendix 1](#)). Consistent with our findings, a 30-minute increase in surgery duration corresponded to a 6% rise in POD risk [28], and this risk is further elevated in prolonged cardiac surgeries using cardiopulmonary bypass [57], potentially leading to hypoperfusion or microembolism [58]. This suggests a potential cumulative effect, particularly in patients with poor preoperative clinical profiles. Overall, our findings highlight the critical importance of cardiovascular surgical risk measures in POD prediction.

Looking ahead, the integration of real-time predictive technology into surgical workflows holds promise. This advancement could potentially facilitate on-the-fly predictions during surgery, enabling timely adjustments to medication or nonpharmacological intervention to mitigate potential adverse outcomes associated with surgical interventions. Our study emphasizes the substantial value of information gleaned from measures taken during surgery, shedding light on their crucial role in enhancing our understanding and prediction of POD.

### **Enhanced POD Prediction Before Elective Surgery Through Neuropsychological Assessments**

With a larger cohort and a more comprehensive battery of preoperative neuropsychological assessments, this study demonstrated slightly improved predictive performance compared to a previous model that used a smaller sample and a limited neuropsychological assessment (MoCA) [37]. Consistent with prior research [16], our results indicate that preoperative models can achieve comparable performance to those incorporating intraoperative features in predicting POD. Although incorporating intraoperative data slightly improved POD predictions, they do not diminish the value of neuropsychological testing for early risk stratification, especially in elective surgeries where timing allows for actionable planning. A key limitation of intraoperative data prevents surgical planning and decision-making beforehand, allowing only adjustments in real time. These findings are particularly relevant to older patients undergoing elective surgery, as they have sufficient preoperative time for noninvasive neuropsychological assessments and for adjusting surgical strategies accordingly. Therefore, augmenting the predictive performance by incorporating data that can be gathered before a surgical procedure is important, as it allows for potential prehabilitation strategies, integrated surgical planning, and informed decision making prior to any invasive or surgical procedure.

In our study, preoperative neuropsychological assessments were predictive above chance (Figure 2B). To achieve performance comparable to the model combining all available features, adding neuropsychological tests to the preoperative

model can effectively replace intraoperative features to predict POD before surgery (Figure 2C), which is critical given that surgical and postoperative management could be optimized. Preoperative models with neuropsychological tests effectively predicted POD before surgery substantiating previous observations [13,29,59]. This could be attributed to preoperative neuropsychological tests revealing subtle cognitive deficits that are not captured by dementia or delirium history. Fewer than 2% of patients in this cohort reported a diagnosis of mild or moderate dementia before elective surgery (Table S1 in [Multimedia Appendix 1](#)), and neither dementia nor delirium history was influential predictors with consistently almost zero feature attributions (Tables S7-S10 in [Multimedia Appendix 1](#)). These subtle deficits may progress into POD. In line with this explanation, timely preoperative cognitive interventions can mitigate the risk of POD and long-term cognitive dysfunction after cardiac surgeries [60,61]. In addition, patients with pre-existing cognitive decline face increased risks of other postoperative complications [62,63]. Consequently, baseline neuropsychological assessments are valuable for improving the prediction of POD beyond what clinical history alone can offer.

Selecting suitable neuropsychological assessments for clinical use is crucial. Our study identified the MoCA and TMTs as effectively indicated by their average absolute SHAP values (Figure 2D and Figure S2-B in [Multimedia Appendix 1](#)). Low scores on the MoCA and longer test times on the TMT indicate poor cognitive performance and executive dysfunction. These tests are crucial for predicting POD risk, as demonstrated in a previous prediction study [15]. Our findings, while focused on prediction [64], can complement previous etiological studies that show patients with mild cognitive impairment at baseline are more likely to develop POD [26], while good preoperative cognitive performance is protective against POD [18]. Previous studies have often used the MMSE [29,32,65], Clock-Drawing Test [20], or MoCA score as preoperative risk factors [66]. Critically, we replicated the strong association between baseline MoCA and POD risk in the previous theory-driven etiological PAWEL-R study with a larger and extended cohort [13]. These findings offer a thorough understanding of the efficacy of individual preoperative neuropsychological tests in predicting POD. By conducting comprehensive assessments of predictors for pre-existing risks, we may unlock new avenues for optimizing surgical planning and postoperative management. Our study underscores the added, albeit moderate, advantage of evaluating cognition, emphasizing its importance and advocating for its inclusion in future developments aimed at refining preoperative risk assessments.

### **Limitations and Recommendations**

The study exhibits several limitations that require consideration. First, interpreting and using SHAP values warrants caution [67], as is generally the case for methods using model explanations in medicine [68,69]. Unstable explanations are not uncommon for complex models trained on large datasets [70,71]. While the ranking of importance may fluctuate, features with higher mean absolute SHAP values generally

maintain consistent attributions. To enhance stability and reduce dependence on specific cohort segments, we aggregated SHAP values across multiple folds with shuffling, ensuring a more reliable assessment of feature importance (Tables S7-S10 in [Multimedia Appendix 1](#)). Second, although permutation is a robust statistical method, its conservative nature means nonsignificance does not always indicate the absence of a difference. Third, to minimize temporal bias and avoid causal leakage, we excluded perioperative features whose timing could not be clearly verified to precede the onset of delirium. While such features may have clinical relevance, their inclusion risks introducing postoutcome information. We recommend future research explore causal-learning approaches to better define temporal relationships, though limitations of observational data must be considered [72]. Fourth, we harnessed cognitive data derived from standardized assessments. However, it is crucial to note that these procedures, while standardized, often remain nondigitalized. This presents significant untapped potential for future

advancements in the realm of risk prediction before surgical procedures. Fifth, while our study included a wide range of preoperative neuropsychological assessments, it is not an exhaustive list. As neuropsychological assessments are time-consuming and require training for assessors to conduct accurate tests and interpret results. Therefore, incorporating semiautomatic assessments of cognition [73] may prove advantageous and a relevant direction for future research.

## Conclusion and Relevance

This study highlights the feasibility of predicting POD before elective surgery in adults aged 70 years and older by using a diverse set of features, including neuropsychological assessments. Our research advances the understanding of POD predictors, enabling a more targeted approach to POD risk prediction in clinical practice. The findings offer crucial insights into predictive features for POD, underscoring the importance of integrating these predictors into the digital transformation of preoperative risk assessment.

## Acknowledgments

This work was supported by grant VF1\_2016--201 from the Innovationsfonds (fund of the Gemeinsamer Bundesausschuss, GBA) as well as the German Research Foundation (DFG) Emmy Noether with reference 513851350 (TW) and the BMBF/DLR Project FEDORA: 01EQ2403G (TW). This work was supported by the BMBF-funded de.NBI Cloud within the German Network for Bioinformatics Infrastructure (de.NBI) (031A532B, 031A533A, 031A533B, 031A534A, 031A535A, 031A537A, 031A537B, 031A537C, 031A537D, 031A538A).

We would like to thank the PAWEL (Patientensicherheit, Wirtschaftlichkeit und Lebensqualität bei elektiven Operationen) and PAWEL-R (Patientensicherheit, Wirtschaftlichkeit und Lebensqualität bei elektiven Operationen–Risk) Study group and the patients' participation in this study, making the work possible.

"PAWEL Study Group" includes members: Florian Metzger, Department of Psychiatry and Psychotherapy, University Hospital Tübingen; Andreas Straub, Department of Anesthesiology, University Hospital Tübingen; Tobias Krüger, Department of Heart Surgery, University Hospital Tübingen; Felix Bausenhardt, Department of Orthopedics, University Hospital Tübingen; Petra Renz, Nursing Department, University Hospital Tübingen; Andreas Walther, Department of Anaesthesiology and Intensive Care, Katharinen Hospital, Klinikum Stuttgart; Carola Bruns, Department of Old Age Psychiatry and Psychotherapy, Klinikum Stuttgart; Juliane Spank, Department of Old Age Psychiatry and Psychotherapy, Klinikum Stuttgart; Patricia Sabbah, Department of Old Age Psychiatry and Psychotherapy, Klinikum Stuttgart; Andreas Häusler, Department of Social and Preventive Medicine, University of Potsdam; Bernd Förstner, Department of Social and Preventive Medicine, University of Potsdam; Susanne Schulze, Department of Social and Preventive Medicine, University of Potsdam; Markus Martin, Department of Neurology and Neurophysiology, University of Freiburg; Bernhard Heimbach, Sebastian Voigt-Radloff, Center for Geriatric Medicine and Gerontology (ZGGF), University of Freiburg; Heiko Reichel, Department of Orthopedics, University Hospital Ulm; Andreas Liebold, Department of Cardiothoracic and Vascular Surgery, University Hospital Ulm; Simone Brefka, Agaplesion Bethesda Clinic, Geriatric Medicine, Ulm University and Geriatric Center Ulm; Stephan Kirschner, Department of Orthopedics, ViDia Christian Clinics Karlsruhe; Nina Stober, Geriatric Center Karlsruhe, ViDia Christian Clinics Karlsruhe; Uwe Mehlhorn, Helios Clinic for Cardiac Surgery, Study Center in Karlsruhe; Jürgen Wasem, Institute for Healthcare Management and Research, University Duisburg-Essen; Anja Neumann, Institute for Healthcare Management and Research, University Duisburg-Essen

## Data Availability

All preprocessing steps and analyses are available on GitHub [74] upon publication. The dataset analyzed during this study is available from the PAWEL study consortium on reasonable request.

## Conflicts of interest

CT received honoraria from serving on the scientific advisory board of Roche, a research grant from the gemeinsame Bundesausschuss der Krankenkassen and has received funding for travel and speaker honoraria from several hospitals for scientific education. GWE has nothing to disclose. CAFvA received honoraria from serving on the scientific advisory board of Biogen, Roche, Novo Nordisk, Biontech, MindAhead UG, Lilly and Dr. Willmar Schwabe GmbH & Co. KG and has received funding for travel and speaker honoraria from Lilly, Biogen, Roche diagnostics AG, Novartis, Medical Tribune Verlagsgesellschaft mbH, Landesvereinigung für Gesundheit und Akademie für Sozialmedizin Niedersachsen e. V., FomF



GmbH | Forum für medizinische Fortbildung and Dr. Willmar Schwabe GmbH & Co. KG and has received research support from Roche diagnostics AG.

## Multimedia Appendix 1

Additional figures and tables.

[\[DOCX File \(Microsoft Word File\), 725 KB-Multimedia Appendix 1\]](#)

## Checklist 1

CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth; version 1.6.1) checklist.

[\[PDF File \(Adobe File\), 3180 KB-Checklist 1\]](#)

## References

- Deng C, Mitchell S, Paine SJ, Kerse N. Retrospective analysis of the 13-year trend in acute and elective surgery for patients aged 60 years and over at Auckland City Hospital, New Zealand. *J Epidemiol Community Health*. Jan 2020;74(1):42-47. [doi: [10.1136/jech-2019-212283](#)] [Medline: [31649040](#)]
- Lee YZ, Dharmawan A, Zhang X, Chua DYC, Low JK. The changing landscape of general surgery in the elderly - trends over a decade in a tertiary centre in Singapore. *ANZ J Surg*. Sep 2022;92(9):2018-2024. [doi: [10.1111/ans.17674](#)] [Medline: [35373437](#)]
- Lutz W, Sanderson W, Scherbov S. The coming acceleration of global population ageing. *Nature New Biol*. Feb 7, 2008;451(7179):716-719. [doi: [10.1038/nature06516](#)] [Medline: [18204438](#)]
- Glance LG, Benesch CG, Holloway RG, et al. Association of time elapsed since ischemic stroke with risk of recurrent stroke in older patients undergoing elective nonneurologic, noncardiac surgery. *JAMA Surg*. Aug 1, 2022;157(8):e222236. [doi: [10.1001/jamasurg.2022.2236](#)] [Medline: [35767247](#)]
- Graham LA, Hawn MT. Managing competing risks for surgical patients with complex medical problems—considering confounding. *JAMA Surg*. Feb 1, 2024;159(2):149. [doi: [10.1001/jamasurg.2023.5952](#)]
- Partridge JSL, Harari D, Martin FC, Dhesi JK. The impact of pre-operative comprehensive geriatric assessment on postoperative outcomes in older patients undergoing scheduled surgery: a systematic review. *Anaesthesia*. Jan 2014;69 Suppl 1(s1):8-16. [doi: [10.1111/anae.12494](#)] [Medline: [24303856](#)]
- Kim KI, Park KH, Koo KH, Han HS, Kim CH. Comprehensive geriatric assessment can predict postoperative morbidity and mortality in elderly patients undergoing elective surgery. *Arch Gerontol Geriatr*. 2013;56(3):507-512. [doi: [10.1016/j.archger.2012.09.002](#)] [Medline: [23246499](#)]
- Aceto P, Antonelli Incalzi R, Bettelli G, et al. Perioperative Management of Elderly patients (PriME): recommendations from an Italian intersociety consensus. *Aging Clin Exp Res*. Sep 2020;32(9):1647-1673. [doi: [10.1007/s40520-020-01624-x](#)] [Medline: [32651902](#)]
- Whitlock EL, Vannucci A, Avidan MS. Postoperative delirium. *Minerva Anestesiol*. Apr 2011;77(4):448-456. [Medline: [21483389](#)]
- Witlox J, Eurelings LSM, de Jonghe JFM, Kalisvaart KJ, Eikelenboom P, van Gool WA. Delirium in elderly patients and the risk of postdischarge mortality, institutionalization, and dementia: a meta-analysis. *JAMA*. Jul 28, 2010;304(4):443-451. [doi: [10.1001/jama.2010.1013](#)] [Medline: [20664045](#)]
- Vasilevskis EE, Han JH, Hughes CG, Ely EW. Epidemiology and risk factors for delirium across hospital settings. *Best Practice & Research Clinical Anaesthesiology*. Sep 2012;26(3):277-287. [doi: [10.1016/j.bpa.2012.07.003](#)]
- Gleason LJ, Schmitt EM, Kosar CM, et al. Effect of delirium and other major complications on outcomes after elective surgery in older adults. *JAMA Surg*. Dec 2015;150(12):1134-1140. [doi: [10.1001/jamasurg.2015.2606](#)] [Medline: [26352694](#)]
- Eschweiler GW, Czornik M, Herrmann ML, et al. Presurgical screening improves RISK prediction for delirium in elective surgery of older patients: The PAWEL RISK study. *Front Aging Neurosci*. 2021;13:679933. [doi: [10.3389/fnagi.2021.679933](#)]
- 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. [doi: [10.1097/ANA.0000000000000566](#)]
- Sadlonova M, Hansen N, Esselmann H, et al. Preoperative delirium risk screening in patients undergoing a cardiac surgery: Results from the prospective observational FINDERI study. *Am J Geriatr Psychiatry*. Jul 2024;32(7):835-851. [doi: [10.1016/j.jagp.2023.12.017](#)]
- Xue B, Li D, Lu C, et al. Use of machine learning to develop and evaluate models using preoperative and intraoperative data to identify risks of postoperative complications. *JAMA Netw Open*. Mar 1, 2021;4(3):e212240. [doi: [10.1001/jamanetworkopen.2021.2240](#)] [Medline: [33783520](#)]

17. Chaiwat O, Chanidnuan M, Pancharoen W, et al. Postoperative delirium in critically ill surgical patients: incidence, risk factors, and predictive scores. *BMC Anesthesiol*. Mar 20, 2019;19(1):39. [doi: [10.1186/s12871-019-0694-x](https://doi.org/10.1186/s12871-019-0694-x)] [Medline: [30894129](https://pubmed.ncbi.nlm.nih.gov/30894129/)]
18. Greaves D, Psaltis PJ, Davis DHJ, et al. Risk factors for delirium and cognitive decline following coronary artery bypass grafting surgery: A systematic review and meta-analysis. *J Am Heart Assoc*. Nov 17, 2020;9(22):e017275. [doi: [10.1161/JAHA.120.017275](https://doi.org/10.1161/JAHA.120.017275)] [Medline: [33164631](https://pubmed.ncbi.nlm.nih.gov/33164631/)]
19. Jankowski CJ, Trenerry MR, Cook DJ, et al. Cognitive and functional predictors and sequelae of postoperative delirium in elderly patients undergoing elective joint arthroplasty. *Anesth Analg*. May 2011;112(5):1186-1193. [doi: [10.1213/ANE.0b013e318211501b](https://doi.org/10.1213/ANE.0b013e318211501b)] [Medline: [21415433](https://pubmed.ncbi.nlm.nih.gov/21415433/)]
20. Vasilian CC, Tamasan SC, Lungeanu D, Poenaru DV. Clock-drawing test as a bedside assessment of post-operative delirium risk in elderly patients with accidental hip fracture. *World J Surg*. May 2018;42(5):1340-1345. [doi: [10.1007/s00268-017-4294-y](https://doi.org/10.1007/s00268-017-4294-y)] [Medline: [29030678](https://pubmed.ncbi.nlm.nih.gov/29030678/)]
21. Wang CG, Qin YF, Wan X, Song LC, Li ZJ, Li H. Incidence and risk factors of postoperative delirium in the elderly patients with hip fracture. *J Orthop Surg Res*. Jul 27, 2018;13(1):186. [doi: [10.1186/s13018-018-0897-8](https://doi.org/10.1186/s13018-018-0897-8)] [Medline: [30049276](https://pubmed.ncbi.nlm.nih.gov/30049276/)]
22. Wu J, Yin Y, Jin M, Li B. The risk factors for postoperative delirium in adult patients after hip fracture surgery: a systematic review and meta-analysis. *Int J Geriatr Psychiatry*. Jan 2021;36(1):3-14. [doi: [10.1002/gps.5408](https://doi.org/10.1002/gps.5408)] [Medline: [32833302](https://pubmed.ncbi.nlm.nih.gov/32833302/)]
23. Wang H, Guo X, Zhu X, et al. Gender differences and postoperative delirium in adult patients undergoing cardiac valve surgery. *Front Cardiovasc Med*. 2021;8:751421. [doi: [10.3389/fcvm.2021.751421](https://doi.org/10.3389/fcvm.2021.751421)]
24. Ansaloni L, Catena F, Chattat R, et al. Risk factors and incidence of postoperative delirium in elderly patients after elective and emergency surgery. *Br J Surg*. Feb 2010;97(2):273-280. [doi: [10.1002/bjs.6843](https://doi.org/10.1002/bjs.6843)] [Medline: [20069607](https://pubmed.ncbi.nlm.nih.gov/20069607/)]
25. Ayob F, Lam E, Ho G, Chung F, El-Beheiry H, Wong J. Pre-operative biomarkers and imaging tests as predictors of post-operative delirium in non-cardiac surgical patients: a systematic review. *BMC Anesthesiol*. Feb 23, 2019;19(1):25. [doi: [10.1186/s12871-019-0693-y](https://doi.org/10.1186/s12871-019-0693-y)] [Medline: [30797230](https://pubmed.ncbi.nlm.nih.gov/30797230/)]
26. Kazmierski J, Banys A, Latek J, et al. Mild cognitive impairment with associated inflammatory and cortisol alterations as independent risk factor for postoperative delirium. *Dement Geriatr Cogn Disord*. 2014;38(1-2):65-78. [doi: [10.1159/000357454](https://doi.org/10.1159/000357454)] [Medline: [24603477](https://pubmed.ncbi.nlm.nih.gov/24603477/)]
27. Kassie GM, Nguyen TA, Kalisch Ellett LM, Pratt NL, Roughead EE. Do risk prediction models for postoperative delirium consider patients' preoperative medication use? *Drugs Aging*. Mar 2018;35(3):213-222. [doi: [10.1007/s40266-018-0526-6](https://doi.org/10.1007/s40266-018-0526-6)] [Medline: [29423780](https://pubmed.ncbi.nlm.nih.gov/29423780/)]
28. Ravi B, Pincus D, Choi S, Jenkinson R, Wasserstein DN, Redelmeier DA. Association of duration of surgery with postoperative delirium among patients receiving hip fracture repair. *JAMA Netw Open*. Feb 1, 2019;2(2):e190111. [doi: [10.1001/jamanetworkopen.2019.0111](https://doi.org/10.1001/jamanetworkopen.2019.0111)] [Medline: [30794305](https://pubmed.ncbi.nlm.nih.gov/30794305/)]
29. Liu J, Li J, He J, Zhang H, Liu M, Rong J. The age-adjusted Charlson Comorbidity Index predicts post-operative delirium in the elderly following thoracic and abdominal surgery: A prospective observational cohort study. *Front Aging Neurosci*. 2022;14:979119. [doi: [10.3389/fnagi.2022.979119](https://doi.org/10.3389/fnagi.2022.979119)]
30. Lin X, Liu F, Wang B, et al. Subjective cognitive decline may be associated with post-operative delirium in patients undergoing total hip replacement: The PNDABLE study. *Front Aging Neurosci*. 2021;13:680672. [doi: [10.3389/fnagi.2021.680672](https://doi.org/10.3389/fnagi.2021.680672)]
31. Segernäs A, Skoog J, Ahlgren Andersson E, Almerud Österberg S, Thulesius H, Zachrisson H. Prediction of postoperative delirium after cardiac surgery with a quick test of cognitive speed, Mini-Mental State Examination and Hospital Anxiety and Depression Scale. *Clin Interv Aging*. 2022;17:359-368. [doi: [10.2147/CIA.S350195](https://doi.org/10.2147/CIA.S350195)] [Medline: [35400995](https://pubmed.ncbi.nlm.nih.gov/35400995/)]
32. Veliz-Reissmüller G, Agüero Torres H, van der Linden J, Lindblom D, Eriksdotter Jönhagen M. Pre-operative mild cognitive dysfunction predicts risk for post-operative delirium after elective cardiac surgery. *Aging Clin Exp Res*. Jun 2007;19(3):172-177. [doi: [10.1007/BF03324686](https://doi.org/10.1007/BF03324686)] [Medline: [17607083](https://pubmed.ncbi.nlm.nih.gov/17607083/)]
33. Cao SJ, Chen D, Yang L, Zhu T. Effects of an abnormal mini-mental state examination score on postoperative outcomes in geriatric surgical patients: a meta-analysis. *BMC Anesthesiol*. May 15, 2019;19(1):74. [doi: [10.1186/s12871-019-0735-5](https://doi.org/10.1186/s12871-019-0735-5)] [Medline: [31092206](https://pubmed.ncbi.nlm.nih.gov/31092206/)]
34. Sadeghirad B, Dodsworth BT, Schmutz Gelsomino N, et al. Perioperative factors associated with postoperative delirium in patients undergoing noncardiac surgery: An individual patient data meta-analysis. *JAMA Netw Open*. Oct 2, 2023;6(10):e2337239. [doi: [10.1001/jamanetworkopen.2023.37239](https://doi.org/10.1001/jamanetworkopen.2023.37239)] [Medline: [37819663](https://pubmed.ncbi.nlm.nih.gov/37819663/)]
35. Peden CJ, Miller TR, Deiner SG, Eckenhoff RG, Fleisher LA, Members of the Perioperative Brain Health Expert Panel. Improving perioperative brain health: an expert consensus review of key actions for the perioperative care team. *Br J Anaesth*. Feb 2021;126(2):423-432. [doi: [10.1016/j.bja.2020.10.037](https://doi.org/10.1016/j.bja.2020.10.037)] [Medline: [33413977](https://pubmed.ncbi.nlm.nih.gov/33413977/)]

36. Jin Z, Hu J, Ma D. Postoperative delirium: perioperative assessment, risk reduction, and management. *Br J Anaesth*. Oct 2020;125(4):492-504. [doi: [10.1016/j.bja.2020.06.063](https://doi.org/10.1016/j.bja.2020.06.063)] [Medline: [32798069](https://pubmed.ncbi.nlm.nih.gov/32798069/)]
37. Menzenbach J, Kirfel A, Guttenthaler V, et al. PRe-Operative Prediction of postoperative DELirium by appropriate SCReening (PROPDESC) development and validation of a pragmatic POD risk screening score based on routine preoperative data. *J Clin Anesth*. Jun 2022;78:110684. [doi: [10.1016/j.jclinane.2022.110684](https://doi.org/10.1016/j.jclinane.2022.110684)] [Medline: [35190344](https://pubmed.ncbi.nlm.nih.gov/35190344/)]
38. Bishara A, Chiu C, Whitlock EL, et al. Postoperative delirium prediction using machine learning models and preoperative electronic health record data. *BMC Anesthesiol*. Jan 3, 2022;22(1):8. [doi: [10.1186/s12871-021-01543-y](https://doi.org/10.1186/s12871-021-01543-y)] [Medline: [34979919](https://pubmed.ncbi.nlm.nih.gov/34979919/)]
39. Chen H, Mo L, Hu H, Ou Y, Luo J. Risk factors of postoperative delirium after cardiac surgery: a meta-analysis. *J Cardiothorac Surg*. Apr 26, 2021;16(1):113. [doi: [10.1186/s13019-021-01496-w](https://doi.org/10.1186/s13019-021-01496-w)] [Medline: [33902644](https://pubmed.ncbi.nlm.nih.gov/33902644/)]
40. Bramley P, McArthur K, Blayney A, McCullagh I. Risk factors for postoperative delirium: An umbrella review of systematic reviews. *Int J Surg*. Sep 2021;93:106063. [doi: [10.1016/j.ijsu.2021.106063](https://doi.org/10.1016/j.ijsu.2021.106063)] [Medline: [34411752](https://pubmed.ncbi.nlm.nih.gov/34411752/)]
41. Sánchez A, Thomas C, Deeken F, et al. Patient safety, cost-effectiveness, and quality of life: reduction of delirium risk and postoperative cognitive dysfunction after elective procedures in older adults-study protocol for a stepped-wedge cluster randomized trial (PAWEL Study). *Trials*. Jan 21, 2019;20(1):71. [doi: [10.1186/s13063-018-3148-8](https://doi.org/10.1186/s13063-018-3148-8)] [Medline: [30665435](https://pubmed.ncbi.nlm.nih.gov/30665435/)]
42. Deeken F, Sánchez A, Rapp MA, et al. Outcomes of a delirium prevention program in older persons after elective surgery: A stepped-wedge cluster randomized clinical trial. *JAMA Surg*. Feb 1, 2022;157(2):e216370. [doi: [10.1001/jamasurg.2021.6370](https://doi.org/10.1001/jamasurg.2021.6370)] [Medline: [34910080](https://pubmed.ncbi.nlm.nih.gov/34910080/)]
43. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. *J Med Internet Res*. Dec 31, 2011;13(4):e126. [doi: [10.2196/jmir.1923](https://doi.org/10.2196/jmir.1923)] [Medline: [22209829](https://pubmed.ncbi.nlm.nih.gov/22209829/)]
44. Cuschieri S. The STROBE guidelines. *Saudi J Anaesth*. Apr 2019;13(Suppl 1):S31-S34. [doi: [10.4103/sja.SJA\\_543\\_18](https://doi.org/10.4103/sja.SJA_543_18)] [Medline: [30930717](https://pubmed.ncbi.nlm.nih.gov/30930717/)]
45. Dong Y, Peng CYJ. Principled missing data methods for researchers. *Springerplus*. Dec 2013;2(1):222. [doi: [10.1186/2193-1801-2-222](https://doi.org/10.1186/2193-1801-2-222)] [Medline: [23853744](https://pubmed.ncbi.nlm.nih.gov/23853744/)]
46. Jameson JL, Fauci AS, Kasper DL, Hauser SL, Longo DL, Loscalzo J, editors. *Harrison's Principles of Internal Medicine*. New York, NY: McGraw-Hill Education; 2018.
47. Wyczalkowska-Tomasik A, Czarkowska-Paczek B, Zielenkiewicz M, Paczek L. Inflammatory markers change with age, but do not fall beyond reported normal ranges. *Arch Immunol Ther Exp (Warsz)*. Jun 2016;64(3):249-254. [doi: [10.1007/s00005-015-0357-7](https://doi.org/10.1007/s00005-015-0357-7)] [Medline: [26283530](https://pubmed.ncbi.nlm.nih.gov/26283530/)]
48. Chawla N, Bowyer K, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *ArXiv*. Preprint posted online on 1813. [doi: [10.1613/jair.953](https://doi.org/10.1613/jair.953)]
49. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. *J Mach Learn Res*. 2011;12:2825-2830. URL: <https://www.jmlr.org/papers/v12/pedregosa11a.html>
50. Chen t, Guestrin C. XGBoost: a scalable tree boosting system. Preprint posted online on 2016.
51. Rutherford S, Barkema P, Tso IF, et al. Evidence for embracing normative modeling. *Elife*. Mar 13, 2023;12:e85082. [doi: [10.7554/eLife.85082](https://doi.org/10.7554/eLife.85082)] [Medline: [36912775](https://pubmed.ncbi.nlm.nih.gov/36912775/)]
52. Ojala M, Garriga GC. Permutation tests for studying classifier performance. Presented at: 2009 Ninth IEEE International Conference on Data Mining; Dec 6-9, 2009; Miami Beach, FL, USA. [doi: [10.1109/ICDM.2009.108](https://doi.org/10.1109/ICDM.2009.108)]
53. Lundberg S, Lee SI. A unified approach to interpreting model predictions. Preprint posted online on 2017.
54. Singh M, Sahhar M, Nassar JE, et al. Analysis of delirium risk assessment tools for prediction of postoperative delirium following lumbar spinal fusion. *Spine (Phila Pa 1976)*. Jan 30, 2025;9900. [doi: [10.1097/BRS.0000000000005271](https://doi.org/10.1097/BRS.0000000000005271)] [Medline: [39882654](https://pubmed.ncbi.nlm.nih.gov/39882654/)]
55. Vreeswijk R, Kalisvaart I, Maier AB, Kalisvaart KJ. Development and validation of the delirium risk assessment score (DRAS). *Eur Geriatr Med*. Apr 2020;11(2):307-314. [doi: [10.1007/s41999-019-00287-w](https://doi.org/10.1007/s41999-019-00287-w)] [Medline: [32297197](https://pubmed.ncbi.nlm.nih.gov/32297197/)]
56. Freter SH, Dunbar MJ, MacLeod H, Morrison M, MacKnight C, Rockwood K. Predicting post-operative delirium in elective orthopaedic patients: the Delirium Elderly At-Risk (DEAR) instrument. *Age Ageing*. Mar 2005;34(2):169-171. [doi: [10.1093/ageing/afh245](https://doi.org/10.1093/ageing/afh245)] [Medline: [15713861](https://pubmed.ncbi.nlm.nih.gov/15713861/)]
57. O'Neal JB, Billings FT 4th, Liu X, et al. Risk factors for delirium after cardiac surgery: a historical cohort study outlining the influence of cardiopulmonary bypass. *Can J Anaesth*. Nov 2017;64(11):1129-1137. [doi: [10.1007/s12630-017-0938-5](https://doi.org/10.1007/s12630-017-0938-5)] [Medline: [28718100](https://pubmed.ncbi.nlm.nih.gov/28718100/)]
58. Levin P. Chapter 223 - postoperative delirium. In: Atlee JL, editor. *Complications in Anesthesia*. Philadelphia: W.B. Saunders; 2007:888-889.

59. Kaźmierski J, Miler P, Pawlak A, et al. Lower preoperative verbal memory performance is associated with delirium after coronary artery bypass graft surgery: A prospective cohort study. *Arch Clin Neuropsychol*. Jan 21, 2023;38(1):49-56. [doi: [10.1093/arclin/acac064](https://doi.org/10.1093/arclin/acac064)] [Medline: [35915987](https://pubmed.ncbi.nlm.nih.gov/35915987/)]
60. Butz M, Meyer R, Gerriets T, et al. Increasing preoperative cognitive reserve to prevent postoperative delirium and postoperative cognitive decline in cardiac surgical patients (INCORE): Study protocol for a randomized clinical trial on cognitive training. *Front Neurol*. 2022;13:1040733. [doi: [10.3389/fneur.2022.1040733](https://doi.org/10.3389/fneur.2022.1040733)] [Medline: [36578306](https://pubmed.ncbi.nlm.nih.gov/36578306/)]
61. Moller J, Cluitmans P, Rasmussen L, et al. Long-term postoperative cognitive dysfunction in the elderly: ISPOCD1 study. *The Lancet*. Mar 1998;351(9106):857-861. [doi: [10.1016/S0140-6736\(97\)07382-0](https://doi.org/10.1016/S0140-6736(97)07382-0)]
62. Weiss Y, Zac L, Refaeli E, et al. Preoperative Cognitive Impairment and Postoperative Delirium in Elderly Surgical Patients: A Retrospective Large Cohort Study (The CIPOD Study). *Ann Surg*. Jul 1, 2023;278(1):59-64. [doi: [10.1097/SLA.0000000000005657](https://doi.org/10.1097/SLA.0000000000005657)] [Medline: [35913053](https://pubmed.ncbi.nlm.nih.gov/35913053/)]
63. Scholz AFM, Oldroyd C, McCarthy K, Quinn TJ, Hewitt J. Systematic review and meta-analysis of risk factors for postoperative delirium among older patients undergoing gastrointestinal surgery. *Br J Surg*. Jan 2016;103(2):e21-8. [doi: [10.1002/bjs.10062](https://doi.org/10.1002/bjs.10062)] [Medline: [26676760](https://pubmed.ncbi.nlm.nih.gov/26676760/)]
64. Ramspek CL, Steyerberg EW, Riley RD, et al. Prediction or causality? A scoping review of their conflation within current observational research. *Eur J Epidemiol*. Sep 2021;36(9):889-898. [doi: [10.1007/s10654-021-00794-w](https://doi.org/10.1007/s10654-021-00794-w)] [Medline: [34392488](https://pubmed.ncbi.nlm.nih.gov/34392488/)]
65. Zhao J, Liang G, Hong K, et al. Risk factors for postoperative delirium following total hip or knee arthroplasty: A meta-analysis. *Front Psychol*. 2022;13:993136. [doi: [10.3389/fpsyg.2022.993136](https://doi.org/10.3389/fpsyg.2022.993136)]
66. Frei BW, Woodward KT, Zhang MY, et al. Considerations for clock drawing scoring systems in perioperative anesthesia settings. *Anesth Analg*. 2019;128(5):e61-e64. [doi: [10.1213/ANE.0000000000004105](https://doi.org/10.1213/ANE.0000000000004105)]
67. Fryer D, Strumke I, Nguyen H. Shapley values for feature selection: The good, the bad, and the axioms. *IEEE Access*. 2021;9:144352-144360. [doi: [10.1109/ACCESS.2021.3119110](https://doi.org/10.1109/ACCESS.2021.3119110)]
68. Petch J, Di S, Nelson W. Opening the black box: The promise and limitations of explainable machine learning in cardiology. *Can J Cardiol*. Feb 2022;38(2):204-213. [doi: [10.1016/j.cjca.2021.09.004](https://doi.org/10.1016/j.cjca.2021.09.004)] [Medline: [34534619](https://pubmed.ncbi.nlm.nih.gov/34534619/)]
69. Oh S, Park Y, Cho KJ, Kim SJ. Explainable machine learning model for glaucoma diagnosis and its interpretation. *Diagnostics (Basel)*. Mar 13, 2021;11(3):510. [doi: [10.3390/diagnostics11030510](https://doi.org/10.3390/diagnostics11030510)] [Medline: [33805685](https://pubmed.ncbi.nlm.nih.gov/33805685/)]
70. Amparore E, Perotti A, Bajardi P. To trust or not to trust an explanation: using LEAF to evaluate local linear XAI methods. *PeerJ Comput Sci*. 2021;7:e479. [doi: [10.7717/peerj-cs.479](https://doi.org/10.7717/peerj-cs.479)] [Medline: [33977131](https://pubmed.ncbi.nlm.nih.gov/33977131/)]
71. Bordt S, von Luxburg U. From shapley values to generalized additive models and back. *arXiv*. URL: <https://arxiv.org/abs/2209.04012> [Accessed 2025-08-08] [doi: [10.48550/arXiv.2209.04012](https://doi.org/10.48550/arXiv.2209.04012)]
72. Sokolova E, von Rhein D, Naaijen J, et al. Handling hybrid and missing data in constraint-based causal discovery to study the etiology of ADHD. *Int J Data Sci Anal*. 2017;3(2):105-119. [doi: [10.1007/s41060-016-0034-x](https://doi.org/10.1007/s41060-016-0034-x)] [Medline: [28691055](https://pubmed.ncbi.nlm.nih.gov/28691055/)]
73. Moon KJ, Son CS, Lee JH, Park M. The development of a web-based app employing machine learning for delirium prevention in long-term care facilities in South Korea. *BMC Med Inform Decis Mak*. Aug 17, 2022;22(1):220. [doi: [10.1186/s12911-022-01966-8](https://doi.org/10.1186/s12911-022-01966-8)] [Medline: [35978303](https://pubmed.ncbi.nlm.nih.gov/35978303/)]
74. MHM-lab/PAWEL-delirium-prediction. *GitHub*. URL: <https://github.com/MHM-lab/PAWEL-Delirium-Prediction> [Accessed 2025-08-08]

## Abbreviations

**ASA:** American Society of Anesthesiologists  
**AUC:** area under the receiver operating characteristic curve  
**CONSORT-EHEALTH:** Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth  
**CRP:** C-reactive protein  
**I-CAM:** Confusion Assessment Method  
**MMSE:** Mini-Mental State Examination  
**MoCA:** Montreal Cognitive Assessment  
**PAWEL:** Patientensicherheit, Wirtschaftlichkeit und Lebensqualität bei elektiven Operationen  
**PAWEL-R:** Patientensicherheit, Wirtschaftlichkeit und Lebensqualität bei elektiven Operationen–Risk  
**PHQ-4:** Patient Health Questionnaire-4  
**POD:** postoperative delirium  
**SHAP:** Shapley Additive Explanations  
**SMI:** Subjective Memory Impairment  
**STROBE:** Strengthening the Reporting of Observational studies in Epidemiology



**TMT: Trail Making Test**

*Edited by Jenny Job; peer-reviewed by Christopher R King, Helmut Frohnhofen; submitted 24.10.2024; final revised version received 02.06.2025; accepted 26.06.2025; published 19.08.2025*

*Please cite as:*

*Wu SCJ, Sharma N, Bauch A, Yang HC, Hect JL, Thomas C, Wagner S, Förstner BR, von Arnim CAF, Kaufmann T, Eschweiler GW, Wolfers T, PAWEL Study Group*

*Predicting Postoperative Delirium in Older Patients Before Elective Surgery: Multicenter Retrospective Cohort Study*

*JMIR Aging* 2025;8:e67958

URL: <https://aging.jmir.org/2025/1/e67958>

doi: [10.2196/67958](https://doi.org/10.2196/67958)

© Shun-Chin Jim Wu, Nitin Sharma, Anne Bauch, Hao-Chun Yang, Jasmine L Hect, Christine Thomas, Sören Wagner, Bernd R Förstner, Christine A F von Arnim, Tobias Kaufmann, Gerhard W Eschweiler, Thomas Wolfers, PAWEL Study Group. Originally published in JMIR Aging (<https://aging.jmir.org>), 19.08.2025. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Aging, is properly cited. The complete bibliographic information, a link to the original publication on <https://aging.jmir.org>, as well as this copyright and license information must be included.