

Saccade Tasks: A Noninvasive Approach for Predicting Postoperative Delirium in Elderly Arthroplasty Patients

Meng Kang, M.S., Xuan Lai, M.D., Junru Wu, Ph.D., Qian Xiang, M.S., Liwen Chen, M.D., Xiang Li, M.S., Jiawen Su, M.D., Zhe Ma, M.B., Yalin Wang, M.B., Wuji Zhao, Ph.D., Yang Li, M.D., Hua Zhang, M.D., Jiansuo Zhou, M.D., Mingsha Zhang, Ph.D., Xiangyang Guo, M.D., Yongzheng Han, M.D.

ANESTHESIOLOGY 2026; 144:622–33



EDITOR'S PERSPECTIVE

What We Already Know about This Topic

- Postoperative delirium may be preventable, but as many prevention strategies are labor intensive, appropriate risk stratification methods are needed
- Existing measures of physical performance, like gait speed and grip strength, may not offer much improvement to delirium prediction models

What This Article Tells Us That Is New

- Parameters like reaction time and degrees of error are readily measurable during saccade tasks, *i.e.*, experimental paradigms for evoking rapid (saccadic) eye movements, in surgical patients
- Compared to neuropsychological tests and serum/cerebrospinal fluid biomarkers, preoperative saccadic parameters—particularly

Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are available in both the HTML and PDF versions of this article. Links to the digital files are provided in the HTML text of this article on the Journal's Web site (www.anesthesiology.org). M.K. and X.Lai. contributed equally to this article. Y.H. and X.G. contributed equally to this article.

Submitted for publication May 13, 2025. Accepted for publication November 10, 2025. Published online first on December 4, 2025.

Copyright © 2025 The Author(s). Published by Wolters Kluwer Health, Inc., on behalf of the American Society of Anesthesiologists. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal. *ANESTHESIOLOGY* 2026; 144:622–33. DOI: 10.1097/ALN.0000000000005875
The article processing charge was funded by grant No. 2021ZD0204300 from the Scientific and Technological Innovation 2030.

Abbreviations: **ASA**, American Society of Anesthesiologists; **AUROC**, area under the receiver operating characteristic curve; **CRP**, C-reactive protein; **CSF**, cerebrospinal fluid; **FP**, fixation point; **IL**, interleukin; **MGS**, memory-guided saccade; **MLP**, multilayer perceptron; **MMSE**, Mini-Mental State Examination; **MoCA**, Montreal Cognitive Assessment; **NfL**, neurofilament light chain; **POD**, postoperative delirium; **PS**, pro-saccade; **RCS**, restricted cubic spline; **THA**, total hip arthroplasty; **TKA**, total knee arthroplasty

ABSTRACT

Background: Postoperative delirium (POD) is a prevalent complication in elderly surgical patients. It is associated with long-term cognitive impairment and increased dementia risk. However, reliable tools to predict POD are currently lacking.

Methods: The study enrolled 316 arthroplasty patients (aged 65 yr or older) in this study. Preoperative assessments comprised neuropsychological tests (*i.e.*, Mini-Mental State Examination [MMSE] and Montreal Cognitive Assessment [MoCA]), molecular biomarkers of serum/cerebrospinal fluid, and saccadic tasks. POD was diagnosed by expert persons based on the Confusion Assessment Method test. The effectiveness of abovementioned three types of assessments in predicting the occurrence of POD were compared.

Results: The incidence of POD was 8.2% (26 of 316). The MMSE and MoCA scales, serum neurofilament light chain levels, and five saccadic parameters values (reaction time, primary saccade error, saccadic gains in pro-saccades; peak velocity in anti-saccades and memory-guided saccades) differed significantly ($P < 0.05$) between POD and non-POD participants. The logistic regression classifier model revealed higher predictive accuracy when using saccadic parameters (area under the receiver operating characteristic curve [AUROC], 0.81; 95% CI, 0.70 to 0.92) than when using MMSE and MoCA scores (AUROC, 0.64; 95% CI, 0.53 to 0.76) or neurofilament light chain levels (AUROC, 0.61; 95% CI, 0.50 to 0.72). The multilayer perceptron machine learning classifier model further increased the accuracy (AUROC, 0.89; 95% CI, 0.82 to 0.94) by using saccadic parameters to predict POD occurrence.

Conclusions: Saccadic parameters exhibited higher accuracy in predicting the occurrence of POD than MMSE and MoCA scores and molecular test results. Therefore, saccadic parameters may serve as a complementary behavioral biomarker for predicting the occurrence of POD in elderly arthroplasty patients.

(*ANESTHESIOLOGY* 2026; 144:622–33)

when included in a machine learning classifier model—predicted postoperative delirium in older arthroplasty patients with remarkable discrimination (area under the curve, 0.89)

- Patients who went on to develop delirium showed poorer saccadic reaction time, spatial error, and peak velocity, suggesting difficulty with spatial attention and motor control

Postoperative delirium (POD) is a common complication in older patients who have undergone surgical treatment (*e.g.*, total knee arthroplasty [TKA] or total hip arthroplasty [THA]).¹ POD typically occurs 24 to 72 h postoperatively and is associated with long-term cognitive dysfunction, higher mortality, and a poor long-term prognosis.² Notably, preoperative cognitive function has been shown to be significantly impaired in individuals with POD compared with those without POD.^{3–5} Therefore, developing sensitive techniques to predict the occurrence of POD by accurately evaluating preoperative cognitive function is crucial. However, current techniques used in the clinic have various shortcomings.

Neuropsychological tests, such as the Mini-Mental State Examination (MMSE) and the Montreal Cognitive Assessment (MoCA), are noninvasive and used frequently in clinical settings. However, they must be administered by trained professional personnel, which may introduce subjective bias.^{5–7} In contrast, tests of molecular biomarkers in peripheral serum and cerebrospinal fluid (CSF) are objective, yet invasive, and their relationship with the occurrence of POD is inconsistent.^{8–10}

Saccades during various tasks can reliably and objectively assess different aspects of cognitive function.^{11–13} For instance, the pro-saccade (PS) task primarily evaluates the function of attention; the anti-saccade task mainly evaluates the function of inhibition to suppress reflexive movements; and the memory-guided saccade (MGS) task primarily evaluates the function of working memory. Although numerous studies have found that saccades are strongly associated with cognitive impairments,^{11,14} whether saccades are associated with POD and could serve as a behavioral biomarker to predict the occurrence of POD remains unknown.

The purpose of the current study was to assess and compare the effectiveness of neuropsychological tests (the MMSE and MoCA), serum/CSF molecule biomarkers, and saccade parameters in predicting the occurrence of POD in elderly patients undergoing lower limb joint replacement surgery. The primary goal was to identify useful tools that could be

used to screen individuals at high risk of developing POD and accurately predict the occurrence of POD before surgery.

Materials and Methods

Ethics Approval

This prospective cohort study was conducted at Peking University Third Hospital (Beijing, China) from August 2022 to August 2024. Ethical approval for the study was obtained from Peking University Third Hospital Medical Science Research Ethics Committee (approval No. IRB00006761-M2022303, May 24, 2022) in Beijing, China. The study was registered with the Chinese Clinical Trial Registry (ChiCTR2200062483, August 8, 2022). Written informed consent was obtained from all participants or their legal representatives before their enrollment.

Participant Selection

The inclusion criteria were: (1) age 65 yr or older of either sex; (2) American Society of Anesthesiologists (ASA; Schaumburg, Illinois) classification of I to III; and (3) scheduled for elective TKA/THA under spinal anesthesia. Exclusion criteria were: (1) major psychiatric (*i.e.*, schizophrenia and epilepsy) or neuromuscular disorders (*i.e.*, myasthenia gravis); (2) severe sensory impairments in reading, writing, or audiovisual tasks; (3) active substance abuse or delirium risk factors (*i.e.*, alcohol dependence and benzodiazepine use); (4) ocular pathology precluding eye tracking (*i.e.*, cataracts, glaucoma, and retinal detachment); and (5) inability to comply with protocol requirements (*e.g.*, voluntary withdrawal, poor compliance, protocol violations, use of medications or methods affecting the trial outcomes, or failure to follow up).

Demographic Data Collection and Neuropsychological Tests

Demographic data were collected before surgery during the initial interview with the participant and comprised age, sex,

Meng Kang, M.S.: Department of Anesthesiology, Peking University Third Hospital, Beijing, China.

Xuan Lai, M.D.: Department of Geriatrics, Peking University Third Hospital, Beijing, China.

Junru Wu, Ph.D.: College of Preschool Education, Capital Normal University, Beijing, China.

Qian Xiang, M.S.: Department of Anesthesiology, Peking University Third Hospital, Beijing, China.

Liwen Chen, M.D.: Department of Geriatrics, Peking University Third Hospital, Beijing, China.

Xiang Li, M.S.: Department of Anesthesiology, Peking University Third Hospital, Beijing, China.

Jiawen Su, M.D.: Department of Anesthesiology, Peking University Third Hospital, Beijing, China.

Zhe Ma, M.B.: Peking University Health Science Center, Beijing, China.

Yalin Wang, M.B.: Peking University Health Science Center, Beijing, China.

Wuji Zhao, Ph.D.: State Key Laboratory of Cognitive Neuroscience and Learning; IDG/McGovern Institute for Brain Research, Division of Psychology, Beijing Normal University, Beijing, China.

Yang Li, M.D.: Department of Orthopedics, Peking University Third Hospital, Beijing, China.

Hua Zhang, M.D.: Research Center of Clinical Epidemiology, Peking University Third Hospital, Beijing, China.

Jiansuo Zhou, M.D.: Department of Laboratory Medicine, Peking University Third Hospital, Beijing, China.

Mingsha Zhang, Ph.D.: College of Preschool Education, Capital Normal University, Beijing, China.

Xiangyang Guo, M.D.: Department of Anesthesiology, Peking University Third Hospital, Beijing, China.

Yongzheng Han, M.D.: Department of Anesthesiology, Peking University Third Hospital, Beijing, China.

weight, height, educational level, ASA classification, perioperative visual analog scale for pain, Charlson Comorbidity Index score, depression score (assessed using the Patient Health Questionnaire-9), and anxiety score (assessed using the Generalized Anxiety Disorder-7). Laboratory test results, medical history, and comorbidities were extracted from the patients' medical records. Patients' preoperative cognitive status was evaluated using the MMSE and MoCA, which were administered by experienced geriatricians.

Anesthesia and Perioperative Pain Management

Standard monitoring procedures included electrocardiogram, noninvasive blood pressure monitoring, heart rate monitoring, and pulse oximetry. Spinal anesthesia was administered *via* L2–3 or L3–4 puncture. After successful puncture, 10 to 15 mg 0.5% bupivacaine was injected. Standard medications of ephedrine, urapidil hydrochloride, atropine, and esmolol were selected to regulate blood pressure and heart rate within $\pm 20\%$ of baseline values.

Femoral nerve block with 20 ml 0.5% ropivacaine for TKA and iliac fascial block with 30 ml 0.3% ropivacaine for THA were provided for postoperative pain management. Acute postoperative pain was managed with intravenous patient-controlled analgesia with 100 μ g sufentanil and 20 mg metoclopramide in 100 ml saline.

Molecular Data Collection and Processing

Peripheral blood samples were collected the day before surgery between 2:00 and 4:00 PM in the ward, and CSF samples were obtained after successful spinal anesthesia puncture in the operating room. All samples were promptly transported to the testing center for centrifugation (3,000g for 10 min at 4°C), and the supernatant was stored at -80°C for subsequent analysis.

CSF and blood specimens were thawed, thoroughly mixed before assay, and then sent to an automatic

chemiluminescence immunoanalyzer for assays of A β 40, A β 42, p-Tau217, neurofilament light chain (NfL), and glial fibrillary acidic protein (GFAP; LiCA 800, Chemcllin, China). In addition, the samples were sent to the Clinical Laboratory Department at Peking University Third Hospital for comprehensive analysis, which included quantification of interleukin (IL)-6 levels (*via* electrochemiluminescent immunoassay using a chemiluminescent immunoassay analyzer [Cobas e801, Roche Diagnostics, Switzerland]) and C-reactive protein (CRP) concentrations (*via* immunoturbidimetry using a biochemical analyzer [IMMAGE 800; Beckman Coulter, USA]).

Saccade Data Collection

Saccadic data were collected 1 day before the operation. Participants were seated in a quiet room, with their heads stabilized on a chin rest to minimize movement. Eye tracking was performed using an infrared video-based eye tracker (EM 2000 Series; Chengdu Jasmines BioTech Inc., China), which recorded the position of both eyes at a sampling rate of 1 kHz. Visual stimuli were presented on a 27-inch liquid-crystal display monitor with a resolution of 1,920 \times 1,080 pixels and a refresh rate of 100 Hz, positioned at a distance of 57 cm from the participants. MATLAB (R2009b; MathWorks, USA) integrated with Psychtoolbox (PTB-3) was used to present visual stimuli and collect the behavioral data. Details regarding the saccadic calibration procedures, systematic error control, and saccadic parameter determination are described in the Supplemental Digital Content (<https://links.lww.com/ALN/E314>). The three saccadic tasks are illustrated in figure 1.¹¹

PS Task. Each trial starts with the presentation of a fixation point (FP), a white cross with a size of $1^\circ \times 1^\circ$, in the center of the screen. Participants were instructed to focus on the FP as soon as it appeared. Upon maintaining fixation within

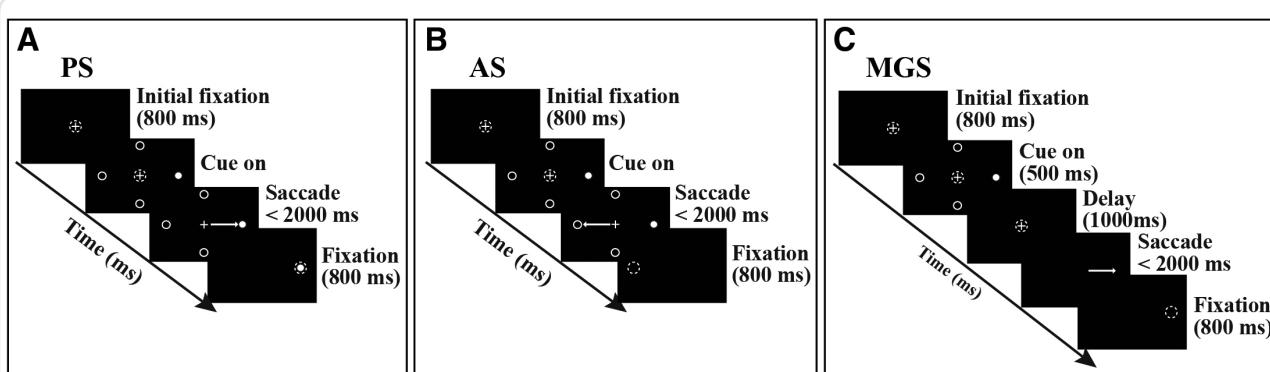


Fig. 1. Schematic illustration of the saccade tasks. (A) Pro-saccade (PS) task. (B) Anti-saccade (AS) task. (C) Memory-guided saccade (MGS) task. The cross represents the fixation point. The dashed circle represents the required fixation location. The filled circle represents the cue. The open circles represent potential cue locations. The arrows represent the required saccade trajectory.

a 4° radius check window for 800 ms, the FP disappeared, and a visual stimulus (a white 1° diameter circle) appeared randomly at one of four peripheral locations (*i.e.*, left, right, up, or down, at an eccentricity of 10°). Participants were asked to execute a saccade to the visual stimulus as quickly and precisely as possible. The visual stimulus disappeared when both eyes entered and remained within a 4° radius check window for 800 ms. If participants failed to enter the fixation window within 2,000 ms of FP onset or did not initiate a saccadic eye movement within 2,000 ms of the visual cue onset, the trial was terminated. A blank screen was displayed for 800 ms between each trial.

Anti-saccade Task. The sequence of events in the anti-saccade task was the same as that in the PS task, except that the participants were instructed to execute a saccade directed toward the mirrored location of the visual stimulus.

MGS Task. The trial begins with the appearance of an FP in the center of the screen. Participants were asked to look at the FP as soon as it appeared. When the participant's eyes entered the stayed within 4° radius check window for 800 ms, a visual stimulus randomly appeared at one of the four peripheral locations (same as those in the PS task) for 500 ms. Participants were instructed to maintain their fixation on the FP and remember the location of the visual stimulus. The stimulus then disappeared, followed by a delay period of 1,000 ms, during which participants were required to maintain their focus on the FP. The disappearance of FP served as a "go signal," whereby participants needed to make a saccade toward the remembered stimulus location as accurately and rapidly as possible. The trial ended when both eyes entered and stayed within the 4° radius check window for 800 ms. If the participant's eyes did not enter the fixation window within 2,000 ms after FP onset or did not make a saccadic eye movement within 2,000 ms after the go signal, the trial was aborted. If the participants moved their eyes out of the 4° radius check window of FP while the FP was displayed, the trial was aborted as a "fixation break." At the end of each trial, a blank screen appeared for an intertrial interval of 800 ms.

POD Assessment

Starting from the first postoperative day, participants were assessed for delirium using the Confusion Assessment Method and Memorial Delirium Assessment Scale twice daily (at 8:00 AM and 8:00 PM) for 3 consecutive days by two geriatricians. If there were significant discrepancies in assessments between the two geriatricians, a panel discussion was organized by a supervisor of the two geriatricians, and final judgment was made based on a comprehensive analysis of the patient's clinical symptoms, medical history, and laboratory test results. Detailed records on the occurrence, duration, severity, and subtype of POD were maintained.

Quality Management and Blinding

The data were collected by researchers with professional competency qualifications who had undergone standardized training. The MMSE and MoCA tests, molecular biomarker measurements, and saccadic tasks were conducted by experienced technicians blinded to the POD diagnosis, strictly following the manufacturer's protocols. Surgical procedures and anesthesia administration were conducted in a standardized manner by the same team of physicians to minimize procedural variability. Additionally, individuals involved in the data collection were independent from those who determined patient group allocation and conducted the data analysis. These measures ensured a robust framework to safeguard data integrity and enhance the reliability of study outcomes.

Statistical Analysis and Prediction Models

The sample size estimation was based on the comparison of the area under the receiver operating characteristic curve (AUROC), as derived from the pilot study. The findings of the preliminary experiment showed a POD incidence of 8.4% (8 of 95). The MMSE-based data set revealed an AUROC of 0.67 for the prediction of POD occurrence, and the saccade tasks showed a higher AUROC of 0.85. With a significance level of $\alpha = 0.05$ and a statistical power of 0.8 ($1 - \beta$), we calculated a sample size using the PASS 11.0 software (NCSS, LLC, USA) of 22 participants in the POD group and 241 participants in the non-POD group to detect significant group differences in diagnostic tests. Accounting for an expected dropout rate of 20%, we estimated a total required sample size of 316 participants.

Data analysis was conducted using the SPSS Statistics 26.0 software (IBM, USA). The Kolmogorov–Smirnov test for normality was first performed to determine the distribution of the measurement data. For normally distributed data, we calculated means and standard deviations, and the POD and non-POD groups were compared using the *t* test. For data with a skewed distribution, we calculated medians and interquartile ranges (IQRs), and the POD and non-POD groups were compared using the Mann–Whitney test. Categorical data, such as sex, surgical procedure, and education level, were analyzed using the chi-square test. Given that age and sex may be potential confounding factors for molecular biomarkers, an analysis of covariance was performed to adjust for these effects.¹¹ Statistical significance was set at $P < 0.05$.

The potential predictors of POD were classified into three categories: neuropsychological tests, molecular biomarkers, and saccadic parameters. We examined the predictive effect of these three categories using a binary logistic regression model. Restricted cubic splines (RCSs) were used to examine whether the three types of predictors were linearly associated with POD occurrence. For predictors with nonlinear association, RCSs were integrated into the

logistic regression model. The predictive performance of all models was evaluated using the AUROC, positive predictive value sensitivity curves, and calibration plots.

For machine learning, the data sets were standardized and balanced using the Synthetic Minority Oversampling Technique.¹⁵ The data set was randomly split into training (80%) and testing (20%) sets to ensure robust model evaluation. For model validation, we applied fivefold cross-validation on the training set and used a multilayer perceptron (MLP) classifier as the primary model, combined with a grid search (GridSearchCV) for hyperparameter optimization to improve model generalization. All parameters were tuned *via* grid search to ensure the best model performance on the training set. Finally, the optimal model was used to make predictions using the test set, and receiver operating characteristic curve were plotted. Detailed information regarding machine learning model construction is provided in the Supplemental Digital Content (<https://links.lww.com/ALN/E315>).

Results

We enrolled 316 participants, recorded their complete demographic profiles, and classified them into POD (n = 26, 8.2%) and non-POD groups (n = 290, 91.8%). Figure 2 illustrates the flow of the experimental procedures. In those with POD, POD emerged within 48 h postoperatively; 96.2% (25 of 26) of cases developed POD at day 1 after surgery, and 3.8% (1 of 26) of cases developed POD at day 2 after surgery. The onset of recovery from POD was within 48 h of surgery in 80.8% of cases, of whom 46.2% (12 of 26) started to recover within 24 h, 34.6% (9 of 26) started to recover within 24 to 48 h after surgery, and 19.2% (5 of 26) of cases experienced POD that lasted beyond 48 h. There were three subtypes of POD: hypoactive (65.4%, 17 of 26), hyperactive (23.1%, 6 of 26), and mixed (11.5%, 3 of 26). Additionally, the highest score for POD severity, as assessed using the Memorial Delirium Assessment Scale, was selected from the multiple

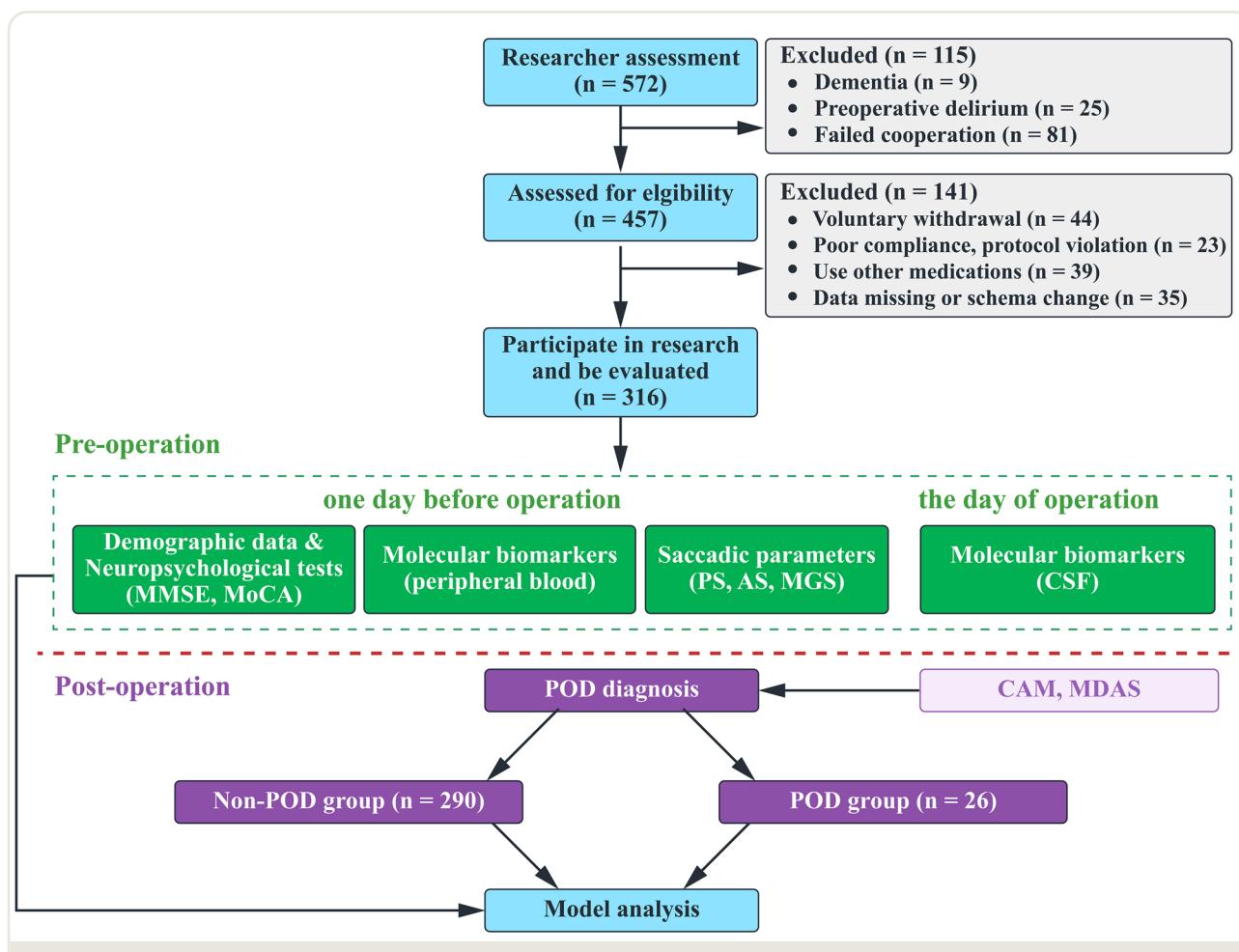


Fig. 2. Flowchart of the study. AS, anti-saccade; CAM, Confusion Assessment Method; CSF, cerebrospinal fluid; MDAS, Memorial Delirium Assessment Scale; MGS, memory-guided saccade; MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment; POD, postoperative delirium; PS, pro-saccade.

assessments performed during POD onset. This served as the standard for severity adjustment (highest score = 19, median score = 8, IQR = 2).

Comparison of Demographic and Clinical Data between POD and Non-POD Groups

As shown in table 1, the POD patients scored significantly lower preoperative on the MMSE ($P = 0.024$) and MoCA

($P = 0.021$) than non-POD patients. No other variable differed significantly between groups.

Comparison of the Preoperative Blood and CSF Molecular Biomarkers between POD and Non-POD Groups

Serum NfL levels in POD patients were significantly higher than in non-POD patients ($P = 0.039$), even after adjusting

Table 1. Demographic Data and Neuropsychological Tests of Study Participants

Characteristic	Non-POD (n = 290, 91.8%)	POD (n = 26, 8.2%)	P Value
Age, mean \pm SD, yr	71.03 \pm 4.39	72.12 \pm 5.70	0.354
Sex, No. (%)			0.494
Male	82 (28.28)	9 (34.62)	
Female	208 (71.72)	17 (65.38)	
BMI, mean \pm SD, kg/m ²	27.13 \pm 3.77	26.76 \pm 3.08	0.627
Duration of disease course, median (IQR), months	107.56 (96.06)	79.17 (59.92)	0.211
Type of operation, No. (%)			0.769
TKA	236 (81.38)	20 (76.92)	
THA	54 (18.62)	6 (23.08)	
Education level, No. (%)			0.600
Illiteracy	23 (7.93)	3 (11.54)	
Primary school	83 (28.62)	9 (34.62)	
Secondary school and above	184 (63.45)	14 (53.85)	
ASA Physical Status, No. (%)			0.354
II	258 (88.97)	21 (80.77)	
III	32 (11.03)	5 (19.23)	
ADL scale, median (IQR)	100.0 (5.0)	100.0 (10.0)	0.877
FRAIL scale, No. (%)			0.379
Robust	132 (45.52)	12 (46.15)	
Prefrail	143 (49.31)	11 (42.31)	
Frail	15 (5.17)	3 (11.54)	
CCI, median (IQR)	1 (2.0)	1 (1.0)	0.699
Depression scale score, median (IQR)	1 (3.0)	1 (2.0)	0.347
Anxiety scale score, median (IQR)	0 (2.0)	0 (2.0)	0.363
Comorbidities, No. (%)			
Hypertension	163 (56.21)	15 (57.69)	0.884
Diabetes	55 (18.97)	7 (26.92)	0.328
Coronary artery disease	53 (18.28)	8 (30.77)	0.122
Alcohol abuse	9 (3.10)	0*	> 0.999
Smoke	12 (4.14)	0*	0.589
Duration of anesthesia, mean \pm SD, min	145.31 \pm 31.94	153.62 \pm 29.72	0.202
Duration of surgery, mean \pm SD, min	84.95 \pm 27.56	88.92 \pm 25.75	0.479
Estimated volume of infusion, mean \pm SD, ml	1,208.94 \pm 306.54	1,304.96 \pm 336.82	0.130
Estimated blood loss, mean \pm SD, ml	74.52 \pm 166.64	95.38 \pm 172.96	0.542
VAS, median (IQR)			
Preoperative	1 (3.0)	1 (3.0)	0.954
Postoperative D1	3 (0.0)	3 (0.0)	0.458
Postoperative D2	3 (2.0)	2.5 (2.0)	0.336
Postoperative D3	2 (2.0)	2.5 (2.0)	0.961
Neuropsychological tests, median (IQR)			
MMSE	27 (4.0)	25 (6.0)	0.024
MoCA	23 (6.0)	21 (8.0)	0.021

Normal distribution is expressed as mean \pm SD. Skewed distribution is expressed as median (IQR). "Illiteracy" refers to individuals who have not received formal schooling; "primary school" refers to individuals who have received basic education through formal schooling, typically lasting no more than 6 yr; and "secondary school and above" refers to individuals who have received education beyond primary school. "Duration of anesthesia" indicates the time from the initiation of anesthesia procedures to when the patient leaves the monitoring. "Duration of surgery" indicates the time from when the surgical incision is made to when the incision is closed. "Volume of infusion" indicates the total volume of infused fluids administered during the surgery. Bold values indicate statistical significance with $P < 0.05$.

*For cells with a value of 0, the P values were calculated using Fisher's exact test.

ADL, Activities of Daily Living; ASA, American Society of Anesthesiologists; BMI, body mass index; CCI, Charlson Comorbidity Index; FRAIL, Fatigue, Resistance, Ambulation, Illnesses, and Loss of weight; IQR, interquartile range; MMSE, Mini-Mental State Examination; MoCA, Montreal Cognitive Assessment; POD, postoperative delirium; THA, total hip arthroplasty; TKA, total knee arthroplasty; VAS, visual analog scale.

Table 2. Comparison of Preoperation Peripheral Serum and CSF Molecular Biomarkers between Non-POD and POD Patients

Molecular Biomarkers, Mean \pm SD, pg/ml	Peripheral Serum					CSF				
	Non-POD (n = 290, 91.8%)	POD (n = 26, 8.2%)	P Value	*Adjusted P Value	MD (95% CI)	Non-POD (n = 290, 91.8%)	POD (n = 26, 8.2%)	P Value	*Adjusted P Value	MD (95% CI)
IL-6	90.61 \pm 488.08	12.28 \pm 33.80	0.396	0.396	77.96 (14.89 to 141.03)	4.28 \pm 2.77	4.00 \pm 2.69	0.620	0.620	0.27 (-0.51 to 1.06)
CRP	0.45 \pm 0.37	0.41 \pm 0.21	0.673	0.673	0.03 (-0.06 to 0.13)	0.12 \pm 0.05	0.12 \pm 0.02	0.711	0.711	0.01 (-0.01 to 0.02)
NfL	30.51 \pm 11.89	36.76 \pm 16.83	0.042	0.042	-5.90 (-12.60 to 0.81)	1,237.24 \pm 1,328.07	1,534.40 \pm 1,290.51	0.432	0.432	-297.25 (-878.35 to 283.85)
GFAP	116.25 \pm 39.38	113.29 \pm 38.44	0.424	0.424	2.71 (-12.88 to 18.31)	6,643.58 \pm 3,805.16	5,750.29 \pm 4,592.55	0.152	0.152	2,019.22 (-18.36 to 4,056.79)
A β 40	128.69 \pm 93.29	128.53 \pm 92.09	0.926	0.926	-0.29 (-37.64 to 37.05)	8,437.59 \pm 3,050.85	8,561.54 \pm 3,476.88	0.909	0.909	-109.14 (-1,622.71 to 1,404.42)
A β 42	5.05 \pm 3.48	4.86 \pm 2.98	0.752	0.752	0.17 (-1.06 to 1.40)	256.91 \pm 107.46	253.56 \pm 132.11	0.949	0.949	4.10 (-53.14 to 61.35)
pTau217	0.50 \pm 0.25	0.48 \pm 0.13	0.583	0.583	0.02 (-0.04 to 0.08)	9.00 \pm 12.75	8.37 \pm 9.36	0.682	0.682	0.61 (-3.86 to 5.08)

Normal distribution is expressed as mean \pm SD.

*Adjusted with age and sex. Bold values indicate statistical significance with $P < 0.05$.

A β 40, amyloid β -protein 40; A β 42, amyloid β -protein 42; CRP, C-reactive protein; CSF, cerebrospinal fluid; GFAP, glial fibrillary acidic protein; IL-6, interleukin-6; MD, mean difference; NfL, neurofilament light chain; POD, postoperative delirium; p-Tau217, phosphorylated Tau protein at threonine 217.

for sex and age ($P = 0.042$). However, no significant differences were detected in IL-6, CRP, GFAP, A β 40, A β 42, and p-Tau217 levels from serum/CSF between the two groups. These results are presented in table 2.

Comparison of Preoperative Saccadic Performance between POD and Non-POD Groups

Overall, the saccadic performance during the three saccadic tasks was poorer in POD patients than in the non-POD patients. As shown in table 3, five saccadic parameters (saccadic reaction time in PS task, saccadic error in PS task, saccadic gain in PS task, peak velocity in anti-saccade task, and peak velocity in MGS task) were significantly poorer in the POD patients than in the non-POD patients ($P < 0.05$).

Comparison of the Predictive Accuracy between Logistic Regression and Machine Learning Models

The results showed that the AUROC was 0.64 (95% CI, 0.53 to 0.76) for neuropsychological tests (MMSE and MoCA), 0.61 (95% CI, 0.50 to 0.72) for the molecular biomarker (NfL), and 0.76 (95% CI, 0.66 to 0.87) for saccade parameters. The saccadic parameters exhibited a nonlinear relationship with POD occurrence (as shown in fig. 3), whereas the other two predictors were linearly correlated with the occurrence of POD. Therefore, we used a logistic regression model with RCS (4 knots) for the saccadic parameters to evaluate their predictive value. The AUROC was 0.81 (95% CI, 0.70 to 0.92), which was higher than

the predictive accuracy obtained using the binary logistic model (AUROC, 0.76; 95% CI, 0.66 to 0.87).

To further examine whether predictive accuracy can be improved by the use of saccadic parameters, we incorporated five saccadic parameters into an MLP machine learning model. The AUROC values of the MLP classifier (trained data: AUROC, 0.91; 95% CI, 0.88 to 0.94; test data: AUROC, 0.89; 95% CI, 0.82 to 0.94) were higher than the AUROC values of logistic regression models (AUROC, 0.81; 95% CI, 0.70 to 0.92). To further evaluate the reliability and stability of these models, we generated positive predictive value sensitivity plot and calibration plot (fig. 4), which showed that the MLP model performed markedly better than the logistic regression model.

Discussion

Owing to the lack of objective tools, early identification and prediction of POD remains challenging.¹⁶⁻¹⁸ The purpose of the current study was to systematically assess and compare the effectiveness of neuropsychological tests (*i.e.*, the MMSE and MoCA), molecular serum and CSF biomarkers, and saccadic parameters in predicting the occurrence of POD in elderly arthroplasty patients. We found that saccadic parameters had the highest accuracy in predicting the occurrence of POD, outperforming both neuropsychological tests and molecular biomarkers. This indicates that saccadic parameters could serve as a complementary behavioral biomarker for preoperatively predicting the occurrence of POD.

Table 3. Comparison of Saccadic Parameters between Non-POD and POD Patients

Parameters, Mean ± SD	PS				AS				MGS			
	Non-POD (n = 290, 91.8%)	POD (n = 26, 8.2%)	P	MD (95% CI)	Non-POD (n = 290, 91.8%)	POD (n = 26, 8.2%)	P	MD (95% CI)	Non-POD (n = 290, 91.8%)	POD (n = 26, 8.2%)	P	MD (95% CI)
Correct rate	0.76±0.31	0.64±0.37	0.181	0.06 (-0.07 to 0.18)	0.32±0.24	0.21±0.28	0.110	0.09 (-0.02 to 0.19)	0.57±0.29	0.44±0.30	0.146	0.09 (-0.04 to 0.21)
Reaction time, ms	239.66±43.07	269.61±45.44	0.030	-16.88 (-35.17 to 1.40)	424.77±98.85	444.80±88.77	0.330	-20.04 (-71.52 to 31.44)	395.40±108.54	438.46±117.82	0.127	-43.38 (-90.65 to 3.89)
Primary saccade error, degree	2.34±1.47	2.92±1.43	0.035	-0.44 (-1.06 to 0.19)	7.95±9.18	10.69±12.55	>0.999	-2.26 (-7.31 to 2.78)	6.81±8.44	8.84±8.70	0.122	-1.25 (-4.38 to 1.89)
Saccadic gain	0.81±0.10	0.74±0.15	0.046	0.05 (0.01 to 0.10)	0.55±0.52	0.72±0.89	0.912	-0.18 (-0.65 to 0.30)	0.63±0.28	0.52±0.31	0.152	0.11 (-0.03 to 0.25)
Peak velocity, degrees/s	380.38±65.28	340.96±51.08	0.052	27.09 (0.33 to 53.85)	379.93±94.08	331.43±122.74	0.041	44.09 (-5.78 to 93.97)	316.41±75.31	275.52±71.74	0.032	29.60 (-4.46 to 63.67)

Normal distribution is expressed as mean ± SD. Bold values indicate statistical significance with $P < 0.05$. AS, anti-saccade; MD, mean difference; MGS, memory-guided saccade; POD, postoperative delirium; PS, pro-saccade.

Limitations of Neuropsychological Tests and Molecular Biomarkers

Previous research has demonstrated that the development of POD is associated with various factors.^{17,19,20} Our findings are consistent with previous studies,^{3–5} and indicate that the preoperative cognitive impairment is a key risk factor for POD. Therefore, screening and identifying patients with preoperative cognitive impairment is crucial for predicting the occurrence of POD. However, preoperative cognition assessment is limited by the lack of accurate and objective assessment tools. For example, some patients with cognitive impairment may be assessed as “normal” on currently available neuropsychological tests. Furthermore, education level has a significant impact on performance of neuropsychological tests, such as MoCA.^{21,22}

Several studies have established a link between molecular biomarkers and neurodegenerative diseases, such as Alzheimer’s disease.^{23,24} However, whether these molecular biomarkers can accurately predict the occurrence of POD remains unknown.^{25–27} In the current study, we analyze seven molecules (*i.e.*, NfL, IL-6, CRP, GFAP, Aβ40, Aβ42, and p-Tau217) from both serum and CSF that have been reported to correlate strongly with neurodegenerative diseases. Serum NfL levels were significantly higher in POD patients than in non-POD patients ($P = 0.042$). The NfL level in the CSF of POD patients also tended to be elevated ($1,534.40 \pm 1,290.51$ vs. $1,237.24 \pm 1,328.07$). Vasunilashorn *et al.*²⁸ examined the temporal variation of IL-6 and CRP levels in TKA/THA patients and found that IL-6 and CRP were increased in the month after surgery but found neither of these molecular biomarkers that could predict the occurrence of POD. Although GFAP has been extensively studied as an Alzheimer’s disease biomarker and therapeutic target,^{29,30} there is currently no consensus on whether it can be applied to POD. Recent studies have revealed that reduced preoperative CSF Aβ42 level is an independent risk factor for POD.³¹ However, a systematic review⁹ found no significant link between amyloid-β, Tau, and POD. Furthermore, Ballweg *et al.*³² found no significant difference in preoperative Tau levels between patients with and without POD. We found no significant differences in Aβ40, Aβ42, or p-Tau217 in the serum and CSF between patients with and without POD. Overall, it remains uncertain whether Alzheimer’s disease-related molecular biomarkers can predict POD occurrence, and further studies are needed to confirm this issue.

In the current study, only serum NfL showed a significant elevation in POD patients, which suggests the limited value of the other six molecular biomarkers in identifying and predicting POD occurrence, which is consistent with the conclusions of a recent systematic review of 28 studies.⁹ They found no sufficient evidence to endorse the use of any single molecule as a definitive risk marker for delirium. Given that NfL originates from the axon of neuron, the increased serum NfL level suggests axonal injury or neurodegeneration, which may contribute to POD pathogenesis.

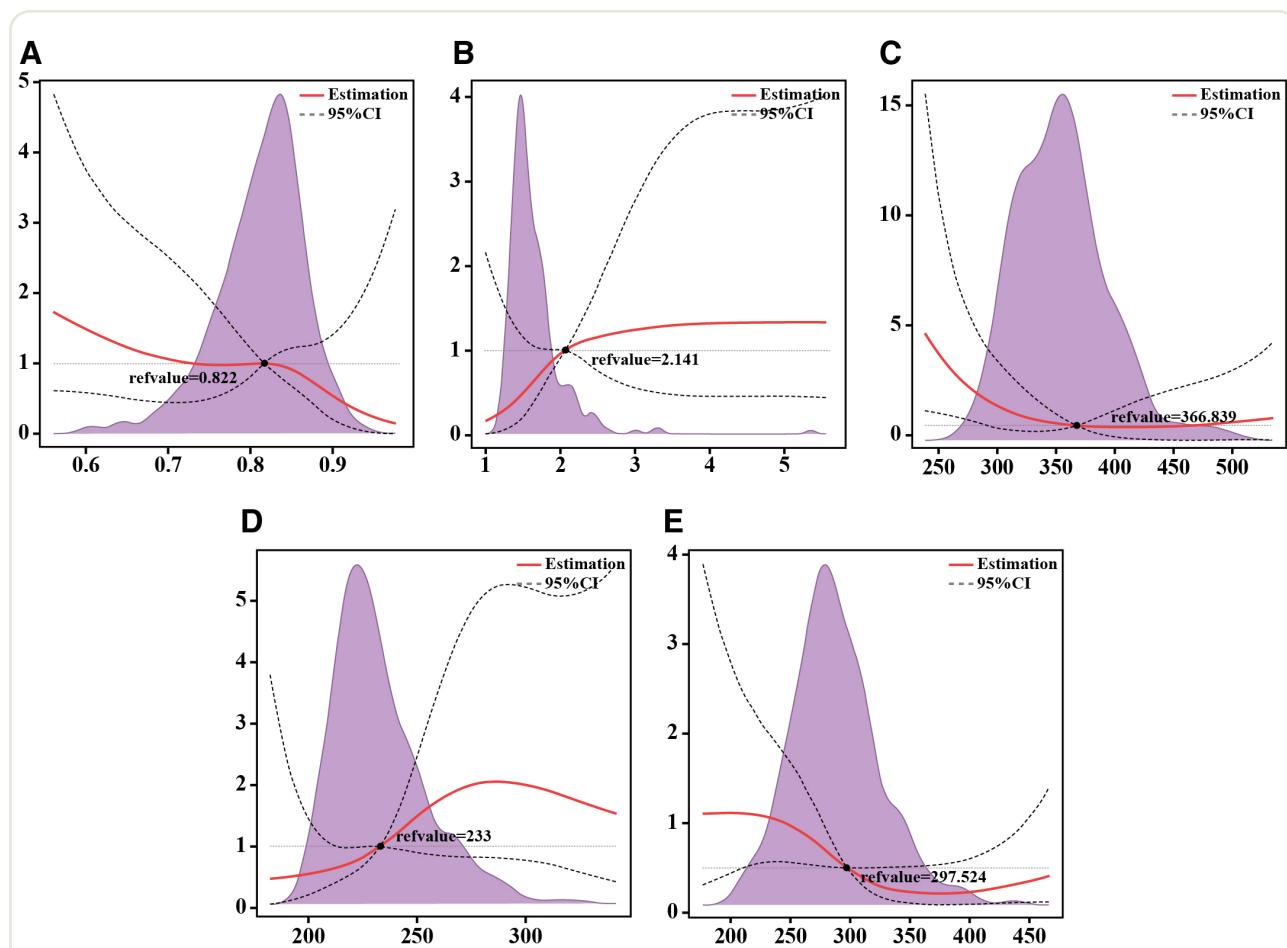


Fig. 3. Restricted cubic splines for the saccadic parameters. (A) Reaction time during the pro-saccade (PS) task. (B) Primary saccade errors during the PS task. (C) Saccadic gains during the PS task. (D) Peak velocity during the anti-saccade task. (E) Peak velocity during the memory-guided saccade task.

Contribution of Saccadic Parameters in Predicting POD

One of the clinical symptoms of POD is cognitive impairment, particularly attention. Saccades are influenced by multiple cognitive functions, such as attention, memory, inhibition, and motor execution.^{11,33} Saccades can be accurately captured using eye-tracking devices and analyzed.^{34,35} We previously reported a strong correlation between saccadic parameters and the scores of MMSE and MoCA.¹¹ In addition, we found that the accuracy of cognitive impairment diagnosis was higher using saccadic parameters than using MMSE and MoCA scores, which is consistent with previous studies.^{11,35–37} This demonstrates the diagnostic utility of saccadic parameters for cognitive impairment. We identified five saccadic parameters across three saccadic tasks that showed a significant difference between patients with and without POD, which suggested that saccadic parameters can serve as a behavioral biomarker to predict the development of POD.

Studies using an eye-tracking technique to study delirium remain scarce.^{38–40} Moreover, the primary purpose of

using eye-tracking techniques varies among studies. Our focus was on assessing the utility of saccadic parameters in predicting the occurrence of POD, whereas previous studies investigated the features of delirium patients using eye-tracking techniques.

Reasons for Poorer Saccadic Parameters in the PS Task than in the Anti-saccade and MGS Tasks

In the current study, patients with POD exhibited significantly longer saccadic reaction times, larger saccadic spatial errors, and smaller saccadic gains during the PS task than those without POD, which indicated reduced visual stimuli response, saccadic planning, and saccadic execution ability. However, these saccadic parameters did not differ significantly between the POD and non-POD patients during the anti-saccade and MGS task. In terms of saccadic peak velocity, POD patients exhibited significantly lower velocity across all three saccadic tasks than the non-POD patients, which indicated a decline in motor control ability that was not influenced by task complexity. Collectively, our findings suggest that POD patients

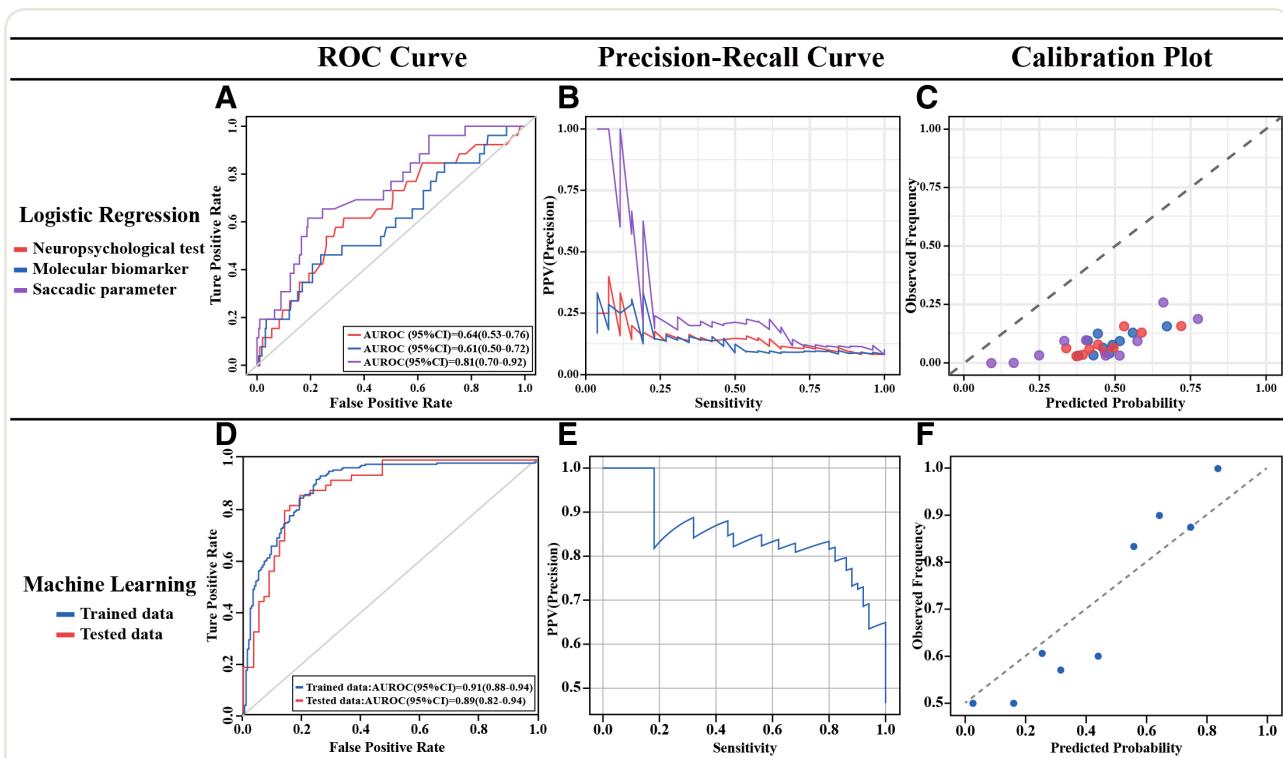


Fig. 4. Receiver operating characteristic (ROC) curves, positive predictive value (PPV) sensitivity plots, and calibration plots of logistic regression (*A–C*) and the machine learning models (*D–F*). (*A, D*) ROC curves. (*B, E*) PPV sensitivity plots. (*C, F*) Calibration plots. AUROC, area under the receiver operating characteristic curve.

exhibited early motor control deficits, including greater saccade initiation time (*i.e.*, increased saccadic reaction times), lower saccadic accuracy (*i.e.*, increased saccadic errors), and slower saccadic speed (*i.e.*, reduced peak velocity) but did not manifest symptoms in comprehension and cognitive impairment. Despite their infrequent use in POD-related studies, saccade tasks offer a promising avenue for POD assessment owing to their simplicity and noninvasive nature and hold significant potential for clinical application.

Unexpectedly, we found greater impairment of saccadic parameters during the PS task than during the anti-saccade and MGS task in POD patients. Accurately performing the anti-saccade and MGS tasks likely requires more cognitive functions than accurately performing the PS tasks. Therefore, when comparing patients with and without POD, we would expect to observe poorer saccadic parameters during the anti-saccade and MGS tasks than during the PS tasks. However, we found opposite results, which may be because all of our participants had chronic osteoarthritis, and chronic pain is well established as a critical causal factor for cognitive impairment.⁴¹ Therefore, both patients with POD and patients without POD may find it equally challenging to perform high-load cognitive tasks, such as the anti-saccade and MGS tasks. This phenomenon is called a “ceiling effect.” In contrast, simple tasks, such as the PS task, may be more

sensitive to detecting saccadic impairments in POD patients than in non-POD patients.

Study Limitations

There are several limitations in the current study. First, the limited number of POD cases ($n = 26$) and the single-center design may have introduced data bias, which may limit the generalizability of our findings. Second, although the MMSE and MoCA are widely used cognitive assessment, they are not the gold standard for cognitive assessment. Third, the first POD assessment was done at 8:00 AM the day after surgery rather than on the day of surgery. These limitations should be considered and addressed in future studies.

Conclusions

Compared with neuropsychological tests and molecular biomarkers, saccadic parameters exhibited higher accuracy in predicting the occurrence of POD. Therefore, saccadic parameter may serve as a complementary behavioral biomarker for predicting the development of POD.

Acknowledgments

The authors thank all the physicians in the Departments of Anesthesiology, Orthopedics, Geriatrics, and Laboratory

Medicine at Peking University Third Hospital (Beijing, China) for support.

Research Support

Supported by grant No. 2021ZD0204300 from the Scientific and Technological Innovation 2030 and grant No. CNLZD2202 from the Open Research Fund of the State Key Laboratory of Cognitive Neuroscience and Learning.

Competing Interests

The authors declare no competing interests.

Reproducible Science

Full protocol available at: hanyongzheng@bjmu.edu.cn. Raw data available at: hanyongzheng@bjmu.edu.cn.

Correspondence

Address correspondence to Dr. Han: Peking University Third Hospital, No. 49 North Garden Road, Haidian District, 100191 Beijing, China. hanyongzheng@bjmu.edu.cn

Supplemental Digital Content

Supplementary material 1, <https://links.lww.com/ALN/E314>

Supplementary material 2, <https://links.lww.com/ALN/E315>

References

- Kitsis P, Zisimou T, Gkiatas I, et al.: Postoperative delirium and postoperative cognitive dysfunction in patients with elective hip or knee arthroplasty: A narrative review of the literature. *Life (Basel)* 2022; 12:314. doi:10.3390/life12020314
- Sprung J, Roberts RO, Weingarten TN, et al.: Postoperative delirium in elderly patients is associated with subsequent cognitive impairment. *Br J Anaesth* 2017; 119:316–23. doi:10.1093/bja/aex130
- Li F, Miao M, Li N, Zhou J, Sun M, Zhang J: Prevalence of preoperative cognitive impairment among elderly thoracic surgery patients and association with postoperative delirium: A prospective observational study. *Front Hum Neurosci* 2023; 17:1234018. doi:10.3389/fnhum.2023.1234018
- Varpaei HA, Robbins LB, Farhadi K, Bender CM: Preoperative cognitive function as a risk factor of postoperative delirium in cancer surgeries: A systematic review and meta-analysis. *J Surg Oncol* 2024; 130:222–40. doi:10.1002/jso.27730
- Folstein MF, Folstein SE, McHugh PR: “Mini-mental state.” A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* 1975; 12:189–98. doi:10.1016/0022-3956(75)90026-6
- Nasreddine ZS, Phillips NA, Bédirian V, et al.: The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *J Am Geriatr Soc* 2005; 53:695–9. doi:10.1111/j.1532-5415.2005.53221.x
- Jia X, Wang Z, Huang F, et al.: A comparison of the Mini-Mental State Examination (MMSE) with the Montreal Cognitive Assessment (MoCA) for mild cognitive impairment screening in Chinese middle-aged and older population: A cross-sectional study. *BMC Psychiatry* 2021; 21:485. doi:10.1186/s12888-021-03495-6
- Taylor J, Parker M, Casey CP, et al.: Postoperative delirium and changes in the blood-brain barrier, neuroinflammation, and cerebrospinal fluid lactate: A prospective cohort study. *Br J Anaesth* 2022; 129:219–30. doi:10.1016/j.bja.2022.01.005
- Wang S, Greene R, Song Y, et al.: Postoperative delirium and its relationship with biomarkers for dementia: A meta-analysis. *Int Psychogeriatr* 2022; 34:377–90. doi:10.1017/S104161022100274X
- Leung JM, Rojas JC, Tang C, et al.; Perioperative Medicine Research Group: Presence of preoperative neurodegeneration biofluid markers in patients with postoperative delirium. *ANESTHESIOLOGY* 2023; 139:432–43. doi:10.1097/ALN.0000000000004666
- Wu J, Li M, Ma W, Zhang Z, Zhang M, Li X: The performance of saccade tasks correlates with cognitive test scores in elderly population: Evidence for the usefulness of oculomotor tests in cognitive assessment. *Gerontology* 2023; 69:321–35. doi:10.1159/000526802
- Yang L, Lu L: Anti-saccade can be used as a screening tool for early cognitive impairment: A correlation study based on anti-saccade parameters and cognitive function. *Psychogeriatrics* 2025; 25:e13215. doi:10.1111/psyg.13215
- Klein C, Rauth R, Biscaldi M: Cognitive correlates of anti-saccade task performance. *Exp Brain Res* 2010; 203:759–64. doi:10.1007/s00221-010-2276-5
- Zhang S, Huang X, An R, Xiao W, Wan Q: The application of saccades to assess cognitive impairment among older adults: A systematic review and meta-analysis. *Aging Clin Exp Res* 2023; 35:2307–21. doi:10.1007/s40520-023-02546-0
- Banerjee J, Taroni JN, Allaway RJ, Prasad DV, Guinney J, Greene C: Machine learning in rare disease. *Nat Methods* 2023; 20:803–14. doi:10.1038/s41592-023-01886-z
- Cai S, Li J, Gao J, Pan W, Zhang Y: Prediction models for postoperative delirium after cardiac surgery: Systematic review and critical appraisal. *Int J Nurs Stud* 2022; 136:104340. doi:10.1016/j.ijnurstu.2022.104340
- Jin Z, Hu J, Ma D: Postoperative delirium: Perioperative assessment, risk reduction, and management. *Br J Anaesth* 2020; 125:492–504. doi:10.1016/j.bja.2020.06.063

18. Arias F, Wiggins M, Urman RD, et al.: Rapid in-person cognitive screening in the preoperative setting: Test considerations and recommendations from the Society for Perioperative Assessment and Quality Improvement (SPAQI). *Perioper Care Oper Room Manag* 2020; 19:100089. doi:10.1016/j.pcorm.2020.100089
19. Aldecoa C, Bettelli G, Bilotta F, et al.: European Society of Anaesthesiology evidence-based and consensus-based guideline on postoperative delirium. *Eur J Anaesthesiol* 2017; 34:192–214. doi:10.1097/EJA.0000000000000594
20. Swarbrick CJ, Partridge JS: Evidence-based strategies to reduce the incidence of postoperative delirium: A narrative review. *Anaesthesia* 2022; 77:92–101. doi:10.1111/anae.15607
21. Trzepacz PT, Hochstetler H, Wang S, Walker B, Saykin AJ; Alzheimer's Disease Neuroimaging Initiative: Relationship between the Montreal Cognitive Assessment and Mini-mental State Examination for assessment of mild cognitive impairment in older adults. *BMC Geriatr* 2015; 15:107. doi:10.1186/s12877-015-0103-3
22. Siqueira GSA, Hagemann PMS, Coelho DS, Santos FHD, Bertolucci PHF: Can MoCA and MMSE be interchangeable cognitive screening tools? A systematic review. *Gerontologist* 2019; 59:e743–63. doi:10.1093/geront/gny126
23. Therriault J, Schindler SE, Salvadó G, et al.: Biomarker-based staging of Alzheimer disease: Rationale and clinical applications. *Nat Rev Neurol* 2024; 20:232–44. doi:10.1038/s41582-024-00942-2
24. Blennow K, Zetterberg H: Biomarkers for Alzheimer's disease: Current status and prospects for the future. *J Intern Med* 2018;284:643–63. doi:10.1111/joim.12816
25. Schaefer ST, Koenigsperger S, Olotu C, Saller T: Biomarkers and postoperative cognitive function: Could it be that easy? *Curr Opin Anaesthesiol* 2019; 32:92–100. doi:10.1097/ACO.0000000000000676
26. 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* 2019; 19:25. doi:10.1186/s12871-019-0693-y
27. Lin Y, Yu N, Lin X, et al.: Preoperative cerebrospinal fluid biomarkers may be associated with postoperative delirium in patients undergoing knee/hip arthroplasty: The PNDABLE study. *BMC Geriatr* 2023; 23:282. doi:10.1186/s12877-023-03943-w
28. Vasunilashorn SM, Ngo LH, Dillon ST, et al.; RISE Study Group: Plasma and cerebrospinal fluid inflammation and the blood-brain barrier in older surgical patients: The Role of Inflammation after Surgery for Elders (RISE) study. *J Neuroinflammation* 2021; 18:103. doi:10.1186/s12974-021-02145-8
29. Pereira JB, Janelidze S, Smith R, et al.: Plasma GFAP is an early marker of amyloid- β but not Tau pathology in Alzheimer's disease. *Brain* 2021; 144:3505–16. doi:10.1093/brain/awab223
30. Asken BM, Elahi FM, La Joie R, et al.: Plasma glial fibrillary acidic protein levels differ along the spectra of amyloid burden and clinical disease stage. *J Alzheimers Dis* 2020; 78:265–76. doi:10.3233/JAD-200755
31. Wang B, Sun X, Wang J, et al.: Potential value of cerebrospinal fluid progranulin in the identification of postoperative delirium in geriatrics patients undergoing knee replacement: The perioperative neurocognitive disorder and biomarker LifestyE study. *Front Aging Neurosci* 2022; 13:772795. doi:10.3389/fnagi.2021.772795
32. Ballweg T, White M, Parker M, et al.: Association between plasma Tau and postoperative delirium incidence and severity: A prospective observational study. *Br J Anaesth* 2021; 126:458–66. doi:10.1016/j.bja.2020.08.061
33. Carpenter RH: Frontal cortex. Choosing where to look. *Curr Biol* 1994; 4:341–3. doi:10.1016/s0960-9822(00)00074-9
34. Wilcockson TDW, Mardanbegi D, Xia B, et al.: Abnormalities of saccadic eye movements in dementia due to Alzheimer's disease and mild cognitive impairment. *Aging (Albany NY)* 2019; 11:5389–98. doi:10.18632/aging.102118
35. Lin J, Xu T, Yang X, et al.: A detection model of cognitive impairment via the integrated gait and eye movement analysis from a large Chinese community cohort. *Alzheimers Dement* 2024; 20:1089–101. doi:10.1002/alz.13517
36. Yang LW, Lu LM: Anti-saccade can be used as a screening tool for early cognitive impairment: A correlation study based on anti-saccade parameters and cognitive function. *Psychogeriatrics* 2025; 25:e13215. doi:10.1111/psyg.13215
37. Chehrehnegar N, Shati M, Esmaeli M, Foroughan M: Executive function deficits in mild cognitive impairment: Evidence from saccade tasks. *Aging Ment Health* 2022; 26:1001–9. doi:10.1080/13607863.2021.1913471
38. Al-Hindawi A, Vizcaychipi M, Demiris Y: A prospective multi-center study quantifying visual inattention in delirium using generative models of the visual processing stream. *Sci Rep* 2024; 14:15698. doi:10.1038/s41598-024-66368-4
39. Garry J, Casey K, Cole TK, et al.: A pilot study of eye-tracking devices in intensive care. *Surgery* 2016; 159:938–44. doi:10.1016/j.surg.2015.08.012
40. Exton C, Leonard M: Eye tracking technology: A fresh approach in delirium assessment? *Int Rev Psychiatry* 2009; 21:8–14. doi:10.1080/09540260802675106
41. He YY, Ding KR, Tan WY, et al.: The role of depression and anxiety in the relationship between arthritis and cognitive impairment in Chinese older adults. *Am J Geriatr Psychiatry* 2024; 32:856–66. doi:10.1016/j.jagp.2024.01.228