**A Systematic Review on Prediction Models for Postoperative Delirium in Non-cardiac Surgery Patients.**

**ABSTRACT**

**INTRODUCTION:** Numerous risk prediction models (RPMs) for postoperative delirium (POD) following non-cardiac surgery have been developed and validated recently. However, the robustness and applicability of these models require further investigation.

**METHOD:** Using PRISMA-2020 guidelines and PROBAST checklist, studies on POD RPMs in non-cardiac surgery patients were searched from PubMed and Google Scholar from January 2021 to December 2023. Inclusion criteria were: (a) adults (aged ≥18 years), (b) non-cardiac surgery patients, (c) development and/or validation of delirium RPMs, and (d) full papers in English. Exclusion criteria were studies not meeting these inclusion parameters.

**RESULTS:** Twelve studies included non-cardiac surgery patients with varying rates of POD (3.22% to 38.30%). The Confusion Assessment Method (CAM) was commonly used for assessing POD risk, with logistic regression being the most employed prediction model. Predictors often found were age, intraoperative blood loss, albumin levels, anesthesia duration, and ICU stays. Internal validation was done in 75% of all the models included. The area under the curve (AUC) ranged from 0.68 to 0.94 for internal validation and from 0.630 to 0.880 for external validation sets. Additionally, most of the models showed a minimal risk of bias (83.3%) and were considered to have a low concern regarding their applicability (75%).

**CONCLUSION:** Based on this review, current RPMs for POD among non-cardiac surgery patients exhibit high accuracy, low risk of bias, and minimal concerns regarding their applicability. We recommend that future research prioritize the external validation of existing models to improve their clinical utility.

**Keywords:** non-cardiac surgery; postoperative delirium; risk predictive models; systemic review.

**KEY MESSAGES**

* Comprehensive Analysis: This study systematically reviewed various prediction models for postoperative delirium in non-cardiac surgery patients, assessing model performance and clinical utility.
* Model Diversity: It identified diverse models with varying predictors, such as age, cognitive impairment, and surgery type, reflecting the complexity of postoperative delirium risk factors.
* Clinical Implications: The review emphasizes the need for improved, standardized models to enhance postoperative delirium prediction and management. This is due to differences in patient populations and predictor variables used.
* Research Gaps: Highlighted gaps include inconsistent model validation and limited external validation.
* Future Directions:Review suggests further research to refine, robust, and enhance generalizability of the prediction models, to ensure better patient outcomes post-surgery.
1. **INTRODUCTION**

Postoperative delirium (POD) represents a prevalent neuropsychiatric condition distinguished by sudden alterations in cognition and attention, variable consciousness levels, and disorganized thought processes [1]. The incidence of POD among patients undergoing non-cardiac surgery ranges from 5% to 53% [2, 3]. POD is associated with increased morbidity and mortality rates, delayed recovery, longer hospital stays, escalated healthcare costs, and a greater burden on healthcare systems, patients, and their families [4-6].

 Forecasting the probability of POD has become vital for delivering optimal patient care. Risk prediction models (RPMs) serve as essential tools for healthcare providers to identify patients at high risk and implement early interventions. An RPM is a statistical model that amalgamates data from multiple indicators [7]. These models present numerous benefits and can markedly enhance patient outcomes through prompt intervention.

Numerous RPMs for predicting the risk of POD in patients undergoing non-cardiac surgery have been recently developed [8-19]. Although these models utilize a range of risk factors, there exists uncertainty regarding their proper development and validation, as well as their applicability for screening the general population. Hence, it is critical to assess the methodologies employed in the development and validation of these models, alongside the predictors and performance, to facilitate informed decisions regarding their application and selection.

 To this end, we sought to perform a systematic review (SR) to offer a thorough analysis of the models devised for POD risk prediction, encompassing their methodological traits, utilized predictors, model efficacy, and possible bias risks. This endeavor aims to consolidate existing evidence and furnish crucial insights for subsequent research and clinical decision-making processes.

**2. MATERIALS AND METHODS**

**2.1 Literature Sources and Extraction**

Guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA 2020) guidelines [20], this study was conducted. Searches were performed in the PubMed and Google Scholar databases to identify studies pertaining to prediction models for postoperative delirium in patients who have undergone non-cardiac surgery. The literature was extracted using keywords such as “postoperative delirium,” “older patients,” “non-cardiac surgery,” and “prediction models.” Following the removal of duplicates both automatically and manually, the selected papers underwent evaluation through two screening phases. To optimize inter-rater reliability, frequent communication was maintained among three researchers (anonymous). Discrepancies were resolved through consultation within the study group.

**2.2 Selection of Studies**

Studies were included in this systematic review (SR) if they focused on: (a) the adult population (aged ≥18 years), (b) patients who underwent non-cardiac surgery, (c) were published between 1 January 2021 and 31 December 2023, (d) involved the development and/or validation of delirium prediction models, and (e) were full papers written in English. Exclusion criteria were: (a) studies on cardiac surgery, (b) studies that did not develop or validate prediction models, (c) papers not reporting on an original patient dataset (including editorials, commentaries, reviews, and meta-analyses), and (d) studies published in languages other than English. All types of study designs were considered, including randomized clinical trials (RCTs), prospective and retrospective studies, cohort studies, among others. Patients were monitored or tested in the postoperative period for over 24 hours using a validated POD scale. The timeframe of delirium development (prevalent vs. incident) was not a limiting factor; however, only prognostic statistics were discussed.

**2.2.1 First Screening Phase**

Through database searching, a total of 258 records were identified. Three researchers (anonymous) independently screened the titles, publication details, and abstracts using EndNote to identify papers that met the inclusion criteria, with POD recording as a primary or secondary outcome. A total of 205 studies were excluded because they did not specifically address delirium in the postoperative setting or due to reasons such as age under 18 years, irrelevance to the study’s focus, review studies, among others.

**2.2.2 Second Screening Phase**

For the second screening phase, qualified studies underwent full-text analysis concerning the type and timing of POD monitoring, conducted by the same researchers (anonymous). Eligibility was conferred only to studies that employed a POD-validated scale for over 24 hours in non-selectively presenting patients with delirium and reported quantitative data on POD incidence. A standardized data extraction form was used to analyze the selected studies. Variables extrapolated included: the validated diagnostic POD scale employed, number of recruited patients, overall incidence of POD, study design and structure (such as RCTs, prospective, retrospective, observational, and secondary analysis based on an observational study), type of surgery (categorized into non-cardiac, general surgery, general abdominal, hip fracture, non-neurology, degenerative spinal, non-brain, laparoscopic gastrectomy, major limb amputation), and the country of study conducted. Upon full-text article assessment for eligibility according to predefined criteria, three studies were excluded for not developing a prediction model. Ultimately, 12 studies specifically focusing on prediction models for postoperative delirium comprised the final count for this systematic review.

**2.3 Risk of Bias**

Using the Prediction model Risk Of Bias Assessment Tool (PROBAST) [21], the three authors evaluated the methodological quality of each included study. This tool encompasses four domains: patient selection, predictors, outcome, and analysis. Featuring 20 signaling questions, it aids in the structured evaluation of bias risk (ROB) and the applicability of diagnostic and prognostic prediction studies. The overall risk of bias (ROB) for each study was categorized as low when all domains were assessed as such, high if one or more domains were deemed high, and unclear if any domains were assessed as unclear. Disagreements were resolved through discussion among the authors.

**2.4 Data Synthesis and Analysis**

The meta-analysis for this SR was not feasible due to several factors. These reasons include: (a) significant variations in study design, patient populations, and predictor variables among the 12 included studies, as evident in **Table 1**. The absence of standardized variables and patient populations posed challenges in drawing meaningful conclusions. (b) Some of the included studies exhibited methodological limitations or biases, such as small sample size [16, 17], inadequate control for confounding factors, or a high risk of bias [12, 17], rendering it inappropriate to aggregate their results in a meta-analysis. Combining flawed studies could lead to misleading conclusions. (c) The heterogeneity in clinical factors, including patient characteristics, surgical procedures, or other clinical variables, made it difficult to interpret and aggregate the pooled results in a meta-analysis. Therefore, we qualitatively synthesized results focusing on the predictors included in each model, model performance, and methodological quality. (**Supplementary material**)

1. **RESULTS**
	1. **Study Selection**

 From the initial literature search across two databases, 258 records were identified, with 38 duplicates subsequently removed. Screening of the remaining studies based on titles and abstracts revealed 15 eligible for full-text evaluation. Due to reasons including irrelevance (n=142), age under 18 years (n=7), non-original research (n=15), and others (n=41), a total of 205 studies were excluded from the review. Upon detailed full-text evaluation, 3 studies were excluded for lacking prediction model development. Ultimately, our systematic review (SR) included 12 studies. **Figure 1** depicts the selection and inclusion process for this review.

 **3.2 Study Characteristics**

 Twelve models predicting the risk of POD in patients undergoing non-cardiac surgery were identified. Among these, ten were retrospective cohort studies, one was a prospective study, and one utilized secondary analysis of observational study data for prediction model construction. The studies, published between January 2021 and December 2023, had a combined sample size of 47,816 patients, with individual studies ranging from 120 to 29,756 patients (**Table 1**). In these models, Chinese patients constituted 91% of the study population. The type of surgery, study type, and validation type are detailed in **Table 2**.

**3.3 Predictive Factors**

 **Table 3** offers an overview of the predictive factors incorporated into the RPMs. The number of predictors employed in constructing the models varied from three [10] to twelve [14]. The most frequently identified predictors, namely age, intraoperative blood loss, albumin levels, duration of anesthesia, and ICU stays, were featured in twelve, four, six, three, and five models, respectively. Mortality among POD patients were not reported in any of the included studies, while comorbidities among POD patients, with diabetes mellitus being the most frequently cited, were reported in two of the studies [9,11].

**3.4 Incidence of POD**

The majority of the included studies (n=11) provided clear diagnostic criteria for delirium, with the Confusion Assessment Method (CAM) being the most prevalent tool for assessing delirium. To assess POD, seven studies (58.3%) utilized the CAM measurement tool, two (16.7%) employed the DSM-V criteria, another two (16.7%) utilized the DSM-IV criteria, and one study (8.3%) used a Chart-based tool [11]. The CAM scale's advantages include its simplicity and straightforwardness, allowing assessments to be completed in less than 5 minutes and applicable in emergency, postoperative, and mixed settings [22, 23]. Approximately 75% of the studies (n=9) indicated that POD assessments generally occurred between 3–7 days post-surgery, while 25% (n=3) did not specify a time frame. The incidence of POD among non-cardiac surgery patients was reported to vary between 3.22% and 38.30% (**Table 1**).

**3.5 Models Performances and Validations**

Logistic Regression (LR) (n=10) and XGBOOST (n=5) emerged as the most frequently utilized models. Additional common models included LASSO, Nomogram, KNN, MLP, RF, AdaBoost, GNB, CNB, SVM, and the Stacking ensemble model. Every model included in this systematic review has undergone both development and validation. Moreover, nine models underwent only internal validation, whereas three received both internal and external validations. The area under the curve (AUC) values reported ranged from 0.684 to 0.94 for internal validations and from 0.630 to 0.880 for external validations, respectively (**Figure 2**). Good performance was demonstrated through reported calibration for ten models. **Table 4** indicates no significant correlation between the AUC values and the number of variables utilized in the development of the RPMs. This suggests that the quantity of predictors employed does not strongly influence the discriminatory capabilities of the developed models.

**3.6 Risk of Bias Assessment**

Using the PROBAST guide, the twelve included studies underwent risk of bias (ROB) assessment to evaluate their quality. ROB was categorized as high, low, or unclear. Within this SR, a majority of the studies (n=9, 75%) were deemed high quality with a low risk of bias. However, one study [12] was found to have a high ROB, and another study [17] had an unclear ROB. Concerning applicability, three studies were classified with a high risk of bias. Overall, the studies exhibited a low ROB (83.3%, n=10) and low applicability concerns (75%, n=9) (**Table 5**). This demonstrates that the methodologies and designs of the included studies were robust. Such robustness is critical in securing the findings' validity and reliability, thus improving the generated evidence's overall quality.

1. **DISCUSSION**

This study constitutes a comprehensive examination of the most recent evidence on predicting postoperative delirium (POD) risk in patients undergoing non-cardiac surgery. It encompasses a detailed analysis of the methodologies employed in the twelve included risk prediction models (RPMs) developed between 2021 and 2023, of which three models underwent external validation, while the others were validated internally. Our investigation revealed a generally low risk of bias in most RPMs; however, we noted a significant inconsistency in both the number and the types of risk factors incorporated within these models.

Given the proliferation of RPMs in recent years, concerns have primarily centered on their accuracy and clinical utility for predicting outcomes across various diseases. All models evaluated in this systematic review demonstrated commendable accuracy in predicting POD risk following non-cardiac surgery, with most models achieving an Area Under the Curve (AUC) greater than 80%. Nevertheless, relying solely on accuracy might be misleading if the distribution of observations across sets is uneven [24]. To ascertain the clinical utility of these RPMs, it is imperative to calculate their net benefit, which will clarify the accuracy of the classifications, and the nature of any errors made. Out of the twelve included studies only six (50%) [10, 12, 14-16] calculated the net benefits of their developed RPMs, leaving the other 6 RPMs’ clinical usefulness in question.

We identified a diverse range of RPMs for POD in non-cardiac surgery patients. While these models show promise in identifying patients at risk, several observations were made. First, considerable heterogeneity in predictor variables and model development methods across studies made it challenging to compare model performance. Second, external validation of RPMs was limited, raising concerns about their generalizability to different patient populations and healthcare settings. Third, studies didn’t report the implementation of RPMs in clinical practice, highlighting the gap between research and translation into real-world settings. Future research should focus on external validation and implementation studies to enhance the clinical utility of RPMs for POD. Fourth, some well-known risk factors were not included in all models (chronic medical conditions such as kidney diseases; polypharmacy, especially those with anticholinergic properties; sedative-hypnotics or narcotics, and even sleep disturbances before or after surgery can contribute to the development of delirium) [25]. This may be because they were not found to be significant after analysis. However, Moon et al. [26] have proposed that these established risk factors with clinical credibility should still be included in the models, even if they do not show significance in the analysis. We strongly support this proposal as it would prevent the exclusion of the known predictors and further improve the generalization of the RPMs.

Furthermore, research has shown a strong correlation between the incidence of POD and the type of surgery performed. The twelve studies encompassed in this SR presented data from patients who underwent a single type of surgery in seven cases (58.3%), whereas five instances (41.7%) involved data from multiple surgical procedures; four studies (33.3%) included patients subjected to non-brain and non-cardiac surgeries without detailing the specific type of surgery. POD incidence typically escalates with the complexity of the surgery [17], and the converse is also accurate [12]. Additionally, hip fracture surgery emerged as the most prevalent type of surgery across the twelve studies. Elderly individuals with osteoporosis are particularly susceptible to hip fractures following falls, a condition that is both highly debilitating and potentially lethal [27].

Surgical and physiological factors, such as extended duration of surgery or anesthesia, along with specific comorbidities (notably diabetes), underscore their potential influence on delirium development. Moreover, correlations with low body mass index (BMI), diminished serum albumin levels, and elevated postoperative pain scores indicate that nutritional status and effective pain management are pivotal in POD prevention [28].

**4.1 Strength of this Study**

This study possesses several strengths. Firstly, it conducted a thorough literature review using two predominant scientific databases, PubMed and Google Scholar, focusing on recent studies concerning prediction models for POD, including only those models that underwent internal or external validation. Additionally, data collection was restricted to studies published between 2021 and 2023, reflecting the period during which most recent studies were conducted; this approach aimed to provide a summary, or an overview based on currently published models. Secondly, the literatures were extracted by three independent authors and reviewing all included studies independently to ensure thoroughness and accuracy. Lastly, the recent PRISMA checklist was employed to conduct this review, followed by an assessment of the quality and risk of bias (ROB) of the included studies using the PROBAST checklist, ensuring adherence to current best practices in research methodology.

**4.2 Limitations of this Study**

A primary limitation of our review is its exclusive focus on studies about non-cardiac surgery. The inclusion of a variety of non-cardiac surgeries introduces variability in the findings, potentially constraining the capacity to derive specific and actionable insights for particular surgical interventions. Consequently, a nuanced approach to interpreting and applying the findings is required to ensure their applicability in varied surgical scenarios. However, this temporal limitation implies the possibility of overlooking certain studies that could offer significant insights.

1. **CONCLUSION**

This systematic review (SR) has pinpointed the twelve most effective risk prediction models (RPMs) derived from studies that involved patients undergoing non-cardiac surgery. In surgical patients, postoperative delirium (POD) exhibits a strong association with various modifiable risk factors. The findings of this SR indicate that the incidence of POD following non-cardiac surgery in adult patients is notably high, imposing a substantial burden on the healthcare system. A scant number of models specifically target non-cardiac surgery patients across all age groups. According to the PROBAST criteria, all included studies that developed or validated prediction models were assessed to exhibit a low risk of bias. Future research endeavors should concentrate on enhancing existing prediction models or formulating new models through rigorous methodologies.

**ETHICAL APPROVAL**

It is not applicable as this is a systematic review.

**ABBREVIATIONS**

POD: Postoperative Delirium

RPM: Risk Prediction Model

SR: Systematic Review

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

RCTs: Randomized Control Trials

LR: Logistic Regression

XGBoost: eXtreme Gradient Boost

LASSO: Least Absolute Shrinkage and Selection Operator

KNN: K-Nearest Neighbor machine

MLP: Multi-layer Perceptron

RF: Random Forest

PROBAST: Prediction Model Risk of Bias Assessment Tool

ROB: Risk of Bias

ICU: Intensive Care Unit

CAM: Confusion Assessment Method

DSM-5: Diagnostic and Statistical Manual of Mental Disorders – 5th edition

GNB: Gaussian Naïve Bayes

CNB: Complement Naïve Bayes

SVM: Support Vector Machine

AUC: Area under the Curve

ASA: American Society Anesthesiology

BMI: Body Mass Index

**REFERENCES**

[1] van Meenen LC, van Meenen DM, de Rooij SE, ter Riet G. Risk prediction models for postoperative delirium: a systematic review and meta-analysis. J Am Geriatr Soc. 2014 Dec;62(12):2383-90. PubMed PMID: 25516034. eng.

[2] Dasgupta M DA-. Preoperative risk assessment for delirium after non-cardiac surgery: a systematic review. J Am Geriatr Soc. 2006.

[3] ​Inouye SK WR, Saczynski JS. . 2014;383(9920):911-22. . Delirium in elderly people. Lancet. 2014.

[4] ​Heyland DK, Muscedere J, Drover J, et al. Persistent organ dysfunction plus death: a novel, composite outcome measure for critical care trials. Crit Care. 2011;15(2):R98. PubMed PMID: 21418560. Pubmed Central PMCID: PMC3219367. Epub 20110318.

[5] ​JR. M. Acute Brain Failure: Pathophysiology, Diagnosis, Management, and Sequelae of Delirium. Crit Care Clin 2017.

[6] ​McCullagh IJ SB, Teodorczuk A, Callaghan M, INCARN.net Psif. 2023;23(1):436. Modifiable risk factors for postoperative delirium in older adults undergoing major non-cardiac elective surgery: a multi-center, trainee-delivered observational cohort feasibility study, and trainee survey. BMC Geriatr. 2023.

[7] ​Janes H, Pepe MS, Gu W. Assessing the value of risk predictions by using risk stratification tables. Ann Intern Med. 2008 Nov 18;149(10):751-60. PubMed PMID: 19017593. Pubmed Central PMCID: PMC3091826. eng.

[8] ​Xu Lin NT YWeaS. Development and validation of a postoperative delirium risk prediction model for non-cardiac surgery in elderly patients: The PNDABLE Study. 2023.

[9] ​Zhi-Hua Huang MKB X-YXeaA. Development and validation of a postoperative delirium prediction model for patients undergoing abdominal surgery: A retrospective, observational, single-center study. 2023.

[10] ​Yang Y, Wang T, Guo H, et al. Development and Validation of a Nomogram for Predicting Postoperative Delirium in Patients With Elderly Hip Fracture Based on Data Collected on Admission. Front Aging Neurosci. 2022;14:914002. PubMed PMID: 35783136. Pubmed Central PMCID: PMC9243358. Epub 20220616. eng.

[11] Shen J, An Y, Jiang B, Zhang P. Derivation and validation of a prediction score for postoperative delirium in geriatric patients undergoing hip fracture surgery or hip arthroplasty. Front Surg. 2022;9:919886. PubMed PMID: 36061065. Pubmed Central PMCID: PMC9437918. Epub 20220819. eng.

[12] ​Song YX, Yang XD, Luo YG, et al. Comparison of logistic regression and machine learning methods for predicting postoperative delirium in elderly patients: A retrospective study. CNS Neurosci Ther. 2023 Jan;29(1):158-67. PubMed PMID: 36217732. Pubmed Central PMCID: PMC9804041. Epub 20221011. eng.

[13] ​Hu XY LH, Zhao X, Sun X, Zhou J, Gao X, et al. 2022;28(4):608-18. . An automated machine learning-based model predicts postoperative delirium using readily extractable perioperative-collected electronic data. . CNS Neurosci Ther. 2022.

[14] ​Zhang Y WD, Chen M, Li YL, Ying H, Yao GL, et al. . 2023;29(1):282-95. . Automated machine learning-based model for predicting delirium in patients after surgery for degenerative spinal disease. . CNS Neurosci Ther. 2023.

[15] ​Liu Y, Shen W, Tian Z. Using Machine Learning Algorithms to Predict High-Risk Factors for Postoperative Delirium in Elderly Patients. Clin Interv Aging. 2023;18:157-68. PubMed PMID: 36789284. Pubmed Central PMCID: PMC9922512. Epub 20230208. eng.

[16] ​Chen J, Ji X, Xing H. Risk factors and a nomogram model for postoperative delirium in elderly gastric cancer patients after laparoscopic gastrectomy. World J Surg Oncol. 2022 Sep 29;20(1):319. PubMed PMID: 36171580. Pubmed Central PMCID: PMC9520878. Epub 20220929. eng.

[17] ​Chen Y MJ, Zeng J, Pan Z, Nie T.2022. . Preoperative risk factors for delirium after major amputation: establishment of a nomogram, 2. 2022.

[18] ​Chen D, Wang W, Wang S, et al. Predicting postoperative delirium after hip arthroplasty for elderly patients using machine learning. Aging Clin Exp Res. 2023 Jun;35(6):1241-51. PubMed PMID: 37052817. Epub 20230413. eng.

[19] ​Matsumoto K, Nohara Y, Sakaguchi M, et al. Temporal Generalizability of Machine Learning Models for Predicting Postoperative Delirium Using Electronic Health Record Data: Model Development and Validation Study. JMIR Perioper Med. 2023 Oct 26;6:e50895. PubMed PMID: 37883164. Pubmed Central PMCID: PMC10636625. Epub 20231026. eng.

[20] ​Page MJ MJ, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. . 2021;372:n71. . The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021.

[21] Wolff RF, Moons KGM, Riley RD, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann Intern Med. 2019 Jan 1;170(1):51-8. PubMed PMID: 30596875.

[22] ​De J, Wand AP. Delirium Screening: A Systematic Review of Delirium Screening Tools in Hospitalized Patients. Gerontologist. 2015 Dec;55(6):1079-99. PubMed PMID: 26543179. Epub 20151105. eng.

[23] ​Grover S, Kate N. Assessment scales for delirium: A review. World J Psychiatry. 2012 Aug 22;2(4):58-70. PubMed PMID: 24175169. Pubmed Central PMCID: PMC3782167. eng.

[24] ​Van Calster B, Wynants L, Verbeek JFM, et al. Reporting and Interpreting Decision Curve Analysis: A Guide for Investigators. Eur Urol. 2018 Dec;74(6):796-804. PubMed PMID: 30241973. Pubmed Central PMCID: PMC6261531. Epub 20180919. eng.

[25] ​Ormseth CH, LaHue SC, Oldham MA, et al. Predisposing and Precipitating Factors Associated With Delirium: A Systematic Review. JAMA Netw Open. 2023 Jan 3;6(1):e2249950. PubMed PMID: 36607634. Pubmed Central PMCID: PMC9856673. Epub 20230103. eng.

[26] ​Moons KGM, Wolff RF, Riley RD, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. Ann Intern Med. 2019 Jan 1;170(1):W1-w33. PubMed PMID: 30596876. eng.

[27] ​Koso RE, Sheets C, Richardson WJ, Galanos AN. Hip Fracture in the Elderly Patients: A Sentinel Event. Am J Hosp Palliat Care. 2018 Apr;35(4):612-9. PubMed PMID: 28823174. Epub 20170821. eng.

[28] ​Collard RM, Boter H, Schoevers RA, Oude Voshaar RC. Prevalence of frailty in community-dwelling older persons: a systematic review. J Am Geriatr Soc. 2012 Aug;60(8):1487-92. PubMed PMID: 22881367. Epub 20120806. eng.

**Figure 1** Shows the selection and inclusion procedures for this review according to the PRISMA (2020) checklist.



**Figure 2** Shows the AUC and validation type of the included studies. Abbreviations: AUC; Area Under the Curve, CI; Confidence Interval.



Abbreviations: AUC; Area Under the Curve, CI; Confidence Interval.

Table 1 Characteristic of Included Studies.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Author, Year | Study Design | Population | Sample Size | AgeGroup | Number of Risk Factors | Incidence of POD | Model Used | Model Performance(AUC) | POD Measurement |
| Lin X. et al, 2023 | Prospective study | Chinese | 633 patients | > 65 years | 8 | 19.76% | Multivariate LR |  0.939 | CAM |
| Huang Z. et al, 2023 | Retrospective study | Chinese | 838 patients | > 18 years | 4 | 10.90% | Multivariate LR |  0.684 and 0.630 | CAM |
| Yang Y. et al, 2022 | Retrospective study | Chinese | 404 patients | > 60 years | 3 | 13.60% | Multivariate LR |  0.820 and 0.841 | CAM |
| Shen J. et al, 2022 | Retrospective study | Chinese | 1312 patients | > 60 years | 9 | 14.33% | LASSO and LR |  0.833 | Chart based tool.  |
| Song Y. et al, 2022 | Retrospective study | Chinese  | 29756 patients | ≥ 65 years  | 8 | 3.22% | LR, RF, GBM, AdaBoost, XGBoost, and a stacking ensemble model | 0.783 (0.765–­0.8) for the LR method, 0.78 for RF, 0.76 for GBM, 0.74 for AdaBoost, 0.73 for XGBoost, and 0.77 for the stacking ensemble model. | DSM-IV |
| Hu X. et al, 2021 | Secondary analysis based on an observational study.  | Chinese | 531 patients | ≥ 65 years  | 8 | 23.54% | LR, RF, XGBoost, SVM |  80.44% (95%CI, 72.24% - 88.64%)  | CAM and CAM-ICU |
| Zhang Y. et al, 2022 | Retrospective study | Chinese  | 663 patients | ≥ 49 years  | 12 | 27.45% | XGBoost, LR, RF, AdaBoost, GNB, CNB, MLP, SVM, KNN | 87.0%, 95% CI: 80.7%–93.3%) | DSM-IV, CAM, DRS, 16-digit span, family/nursing staffinterviews |
| Author, Year | **Study Design** | **Population** | **Sample Size** | **Age****Group** | **Number of Risk Factors** | **Incidence of POD** | **Model Used** | **Model Performance****(AUC)** | **POD Measurement** |
| Liu Y. et al, 2023 | Retrospective study | Chinese  | 950 patients | > 60 years  | 7 | 13.79%  | XGBoost, KNNMLP | Internal validation 0.924 (0.841–0.998)External validation 0.880 | DSM-5 |
| Chen J. et al, 2022 | Retrospective study | Chinese | 270 patients | 65-85 years | 5 | 27.40% | Nomogram | 0.860  | DSM-5 |
| Chen Y. et al, 2022 | Retrospective study | Chinese | 120 patients | > 18 years | 4 | 38.30% | Nomogram | 0.918 | CAM |
| Chen D. et al, 2023 | Retrospective study | Chinese | 476 patients | > 60 years | 10 | 18.10% | Combined MI and LR | 0.94 | CAM |
| Matsumoto K. et al, 2023 | Retrospective study | Japanese | 11863 patients | > 18 years | 8 | 9.10% | XGBoost, LASSO | 0.86 - 0.89 | CAM |

**Abbreviations:** LR; Logistic Regression, AUC; Area Under the Curve, CAM; Confusion Assessment Method, LASSO;  Least Absolute Shrinkage and Selection Operator, DSM-5; Diagnostic and Statistical Manual of Mental Disorders – 5th edition, POD; Postoperative Delirium, MI; Mutual Information, XGBoost; eXtreme Gradient Boost tree, RF; Random Forest, SVM; Support Vector Machine, CI; Confidence Interval, ICU; Intensive Care Unit, GNB; Gaussian NB, CNB; Complement NB, MLP; Multi-layer Perceptron, KNN; K-Nearest Neighbor machine, DRS; Delirium Rating Scale-Revised-98.

**Table 2** Evaluation of Surgery Type, Study Type, POD Assessment Time, and Validation Type.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Author, Year  | Type of Surgery  | Study Type  | POD Diagnosis Time  | Validation Type  |
| Lin X. et al, 2023 | Non-cardiac surgery | DV | 1-7 days post-surgery | Internal |
| Huang Z. et al, 2023 | Abdominal surgery | DV | 1-3 days post-surgery | Internal and External |
| Yang Y. et al, 2022 | Hip fracture surgery. | DV | Not reported | Internal and External |
| Shen J. et al, 2022 | Non-cardiac surgery | DV | Not reported | Internal |
| Song Y. et al, 2022 | Non-cardiac surgery and non-neurology surgery  | DV | 1-7 days post-surgery | Internal |
| Hu X. et al, 2021 | Major non-cardiac or non-neurological surgery | DV | 2 hours post-surgery, then twice a day for 3 days | Internal |
| Zhang Y. et al, 2022 | Degenerative spinal surgery | DV | 1-5 days post-surgery | Internal |
| Liu, Y. et al, 2023 | Non-brain surgery | DV | Not reported | Internal and External |
| Chen, J. et al, 2022 | Laparoscopic gastrectomy | DV | 7 days post-surgery | Internal |
| Chen, Y. et al, 2022 | Major limb amputation | DV | Postoperatively on a daily basis | Internal |
| Chen, D. et al, 2023 | Hip arthroplasty. | DV | Previous day and the 1-3 postoperative days | Internal |
| Matsumoto, K. et al, 2023 | General surgery  | DV | First day after surgery | Internal |

DV: Development with Validation; POD: Postoperative Delirium

**Table 3** Predictors of Included Studies.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| References | Age | Intraoperative Blood Loss  | Albumin  | Anesthesia Duration  | ICU Stay  | Other Risk Factors  |
| Lin X. et al | √ |  |  |  |  | Surgery duration, diabetes history, low years of education, high sleep quality index, high ASA classification, and high NRS score. |
| Huang Z. et al | √  |  |  |  |  | Diazepam usage history, and intraoperative positive fluid balance. |
| Yang Y. et al | √  |  | √  |  |  | Dementia and COPD. |
| Shen J. et al | √  |  |  |  |  | Preoperative delirium, diabetes, CCI, application of benzodiazepines in surgery, surgical delay, ≥2 days, creatine ≥90 μmol/L, active smoking, and application of benzodiazepines in surgery. |
| Song Y. et al | √  | √  | √  | √  |  | ASA grade, depression, emergency surgery, duration of anesthesia, WBC count, and antipsychotic drugs. |
| Hu X. et al | √  | √  |  | √  | √  | Extubating time, MMSE, CCI, and postoperative NLR. |
| Zhang Y. et al | √  | √  | √  |  |  | Admission-to-surgery time interval, C-­reactive protein level, smoking, alcohol consumption, pulmonary disease, and admission-­intraoperative maximum BP difference hypertension, and intraoperative minimum BP. |
| Liu Y. et al  | √  |  |  |  | √  | Surgery duration, history of smoking, history of alcoholism, history of hypertension and history of COPD. |
| Chen J. et al | √  |  | √  |  | √  | AFR, NLR, and sleeping pills. |
| References | **Age** | **Intraoperative Blood Loss** | **Albumin** | **Anesthesia Duration** | **ICU Stay** | **Other Risk Factors** |
| Chen Y. et al | √  |  | √  |  | √  | Barthel classification. |
| Chen D. et al | √  |  | √  |  |  | Cystatin C, GFR, CHE, CRP, LDH, Mg2+, MONO, and History of mental illness or psychotropic drug use. |
| Matsumoto K. et al | √  | √  |  | √  | √  | BMI, emergency admission, use of ambulance, neurosurgery, thoracic cavity and mediastinum surgery, and history of delirium. |

**Abbreviations:** ICU; Intensive Care Unit, ASA; American Society Anesthesiologists, NRS; Numerical Pain Score, COPD; Chronic Obstructive Pulmonary Disease, CCI; Charlson Comorbidity Index, WBC; White Blood Cells, MMSE; Mini-Mental State Examination score, NLR; Neutrophil-to-Lymphocyte Ratio, BP; Blood Pressure, AFR; Albumin/Fibrinogen Ratio, GFR; Glomerular Filtration Rate, CHE; Cholinesterase, CRP; C-Reactive Protein, LDH; Lactate Dehydrogenase, Mg2+; Magnesium ions, MONO; Monocyte Count, BMI; Body Mass Index.

**Table 4** Shows the Correlation between the AUC and the Number of Variables used in the Development of the RPMs.

|  |  |  |  |
| --- | --- | --- | --- |
| r | r2 | 95% CI  | p-value (two-tailed) |
| 0.2989  | 0.08937  | 0.20655 (-0.3319-0.3452)  | 0.3452  |

**Abbreviations:** r; Correlation Coefficient, r2; Correlation of Determination, CI; Confidence Interval, p-value; Probability value.

**Table 5** PROBAST Results of Included Studies.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| References | Participants | Predictors | Outcome | Analysis | Overall |
| **RoB** | **DA** | **RoB** | **DA** | **RoB** | **DA** | **RoB** | **DA** | **RoB** | **Applicability** |
| Lin X. et al  | low | low | low | low | low | low | low | low | low | low |
| Huang Z. et al  | low | low | low | low | low | low | low | low | low | low |
| Yang Y. et al  | low | low | low | low | low | low | low | low | low | low |
| Shen J. et al  | low | low | low | low | low | low | low | low | low | low |
| Song Y. et al  | low | low | low | low | low | low | high | high | high | high  |
| Hu X. et al  | low | low | low | low | low | low | low | low | low | low |
| Zhang Y. et al  | low | low | low | low | low | low | low | low | low | low |
| Liu Y. et al  | low | low | low | low | low | low | low | low | low | low |
| Chen J. et al  | high | low | low | low | low | low | low | low | low | high |
| Chen Y. et al  | high | low | low | low | low | low | low | unclear | high  | high |
| Chen D. et al  | low | low | low | low | low | low | low | low | low | low |
| Matsumoto K. et al  | low | low | low | low | low | low | low | low | low | low |

**Abbreviations:** RoB; Risk of Bias, DA; Domain Applicability.