

Systematic Review of Delirium Risk Prediction Models in Adult ICU Patients

Yiyuan Qiu¹, Lan Yao^{2*}

¹Department of Thoracic Surgery, The First Affiliated Hospital of Dalian Medical University, Dalian 116011, Liaoning, China

²Department of Nursing, The First Affiliated Hospital of Dalian Medical University, Dalian 116011, Liaoning, China

*Corresponding author: Lan Yao, 561151512@qq.com

Copyright: © 2026 Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), permitting distribution and reproduction in any medium, provided the original work is cited.

Abstract: *Objective:* To systematically retrieve and evaluate delirium risk prediction models in adult intensive care unit (ICU) patients and provide evidence for model optimization and clinical practice. *Methods:* A systematic literature search was conducted in PubMed, Web of Science, Embase, Cochrane Library, CINAHL, CNKI, Wanfang Data, VIP, and CBM databases from inception to August 2025. Two independent researchers performed literature screening, data extraction, and quality assessment. *Results:* Seventeen studies involving 18 prediction models were included, with sample sizes ranging from 100 to 8,878 cases and outcome events from 21 to 4,431 cases. The area under curve (AUC) for model development ranged from 0.732 to 0.941. Twelve models reported calibration, 10 studies performed internal validation, and 6 studies conducted external validation. Overall applicability of the 17 studies was acceptable. The final models contained 3-15 predictors; however, all models exhibited high risk of bias. *Conclusion:* Existing models generally demonstrated acceptable discrimination, but were associated with high risk of bias, with insufficient evidence regarding external validation, calibration assessment, and clinical implementation. Future studies should strictly follow the PROBAST criteria, standardize procedures for model development and validation, improve model stability and clinical application value, and provide reliable evidence for clinical decision-making.

Keywords: Intensive care unit; Delirium; Prediction model; Systematic review

Online publication: October 23, 2026

1. Introduction

Delirium is an acute, reversible cognitive dysfunction syndrome characterized by fluctuating disturbance of consciousness, inattention, disorganized thinking, or altered level of consciousness ^[1]. It is the most common acute brain dysfunction complication in intensive care unit (ICU) patients, with an incidence rate as high as 20–80% ^[2]. Studies have indicated that delirium impairs patients' cognitive function, increases risk of complications, prolongs mechanical ventilation duration and overall hospital stay, and may lead to

disability or death in severe cases, affecting patient prognosis and increasing clinical burden^[3]. Therefore, early identification of high-risk factors and targeted management are of great clinical significance for improving patients' delirium status. As a scientific and effective risk prediction tool, risk prediction models can integrate various independent risk factors, quantify individual disease probability, accurately screen high-risk populations, and provide theoretical support for clinical medical staff to implement personalized interventions^[4]. Currently, scholars worldwide have constructed and validated prediction models for delirium risk in ICU patients; however, differences exist in predictors, modeling methods, and performance indicators, and systematic evaluation of related models is still lacking. Therefore, this study systematically reviews the current status of delirium risk prediction models in ICU patients and comprehensively evaluates the included studies to provide theoretical reference for clinical practice and related research.

2. Materials and methods

2.1. Literature search strategy

A systematic literature search was performed in PubMed, Web of Science, Embase, Cochrane Library, CINAHL, CNKI, Wanfang Data, VIP, and CBM databases for studies on delirium risk prediction models in ICU patients, from database inception to August 2025. The search combined subject terms and free-text words. For Chinese databases, taking CNKI as an example, the search strategy was: Subject = (critical care OR intensive care unit OR ICU) AND Subject = (delirium syndrome OR clinical delirium OR delirium) AND Subject = (early warning OR risk prediction OR prediction model OR diagnostic model OR risk score OR risk assessment OR risk factors OR nomogram). For English databases, taking PubMed as an example, the search strategy was: (((("critical illness"[MeSH]) OR ("intensive care units"[MeSH])) AND ((("delirium"[MeSH Terms]) OR ("delirium"[Title/Abstract])) AND (((((((("risk assessment"[MeSH]) OR ("risk prediction"[Title/Abstract])) OR ("risk forecast model"[Title/Abstract])) OR ("risk score"[Title/Abstract])) OR ("prediction model"[Title/Abstract])) OR ("prediction tool"[Title/Abstract])) OR ("nomograms"[MeSH])) OR ("diagnostic model"[Title/Abstract])) OR ("risk factors"[MeSH])))). This study was registered on the PROSPERO platform (Registration No. CRD42024602398).

2.2. Inclusion and exclusion criteria

2.2.1. Inclusion criteria

- (1) Study subjects were adult ICU patients;
- (2) Study content was construction or validation of delirium risk prediction models in adult ICU patients;
- (3) Study type was cohort study (prospective or retrospective) or cross-sectional study; (4) Outcome indicator was delirium.

2.2.2. Exclusion criteria

- (1) No clear model reported;
- (2) Conference abstracts, grey literature, and other informally published literature;
- (3) Complete model information unavailable, such as only mentioning model name without providing formula or regression coefficients.

2.3. Literature screening and data extraction

Two researchers independently completed literature screening according to preset inclusion and exclusion criteria. In case of disagreement, a third researcher was consulted. After removing duplicate literature using NoteExpress software, titles and full texts were read to exclude irrelevant literature before inclusion. Two researchers independently extracted the following information using a unified form: basic study characteristics, outcome definition, candidate predictors, modeling methods, sample size and number of events, variable processing methods, validation methods, discrimination indicators, and calibration indicators.

2.4. Risk of bias and applicability assessment

2.4.1. Risk of bias assessment

This study used the Prediction model Risk of Bias ASsessment Tool (PROBAST) to conduct quality evaluation from four domains: participants, predictors, outcome, and analysis, covering 20 items. Each domain included 2-9 questions, judged as “yes/probably yes,” “no/probably no,” or “no information.” Each dimension was individually rated as low risk, high risk, or unclear. A single domain was rated as low risk if all key questions were met, high risk if any key question was not met, and unclear if key information was missing but other items were met. Overall risk of bias was determined by combining results from the four domains: low risk if all domains were low risk, high risk if any domain was high risk, and unclear if only unclear domains existed with others being low risk ^[5,6].

2.4.2. Applicability assessment

The evaluation logic for applicability risk was parallel to that of bias risk, but based on evaluation results from three domains: participants, predictors, and outcome ^[6]. If applicability risk results for all three domains were “low,” overall applicability risk was rated as “low.” If at least one domain had “high” applicability risk, overall applicability risk was rated as “high.” If at least one domain had “unclear” applicability risk and other domains were “low,” overall applicability risk was rated as “unclear.”

3. Results

3.1. Literature search process and results

A total of 3,037 relevant articles were retrieved through the search. After removing duplicates, 2,542 articles were obtained. After screening, 17 articles were finally included. The literature screening process is shown in **Figure 1**.

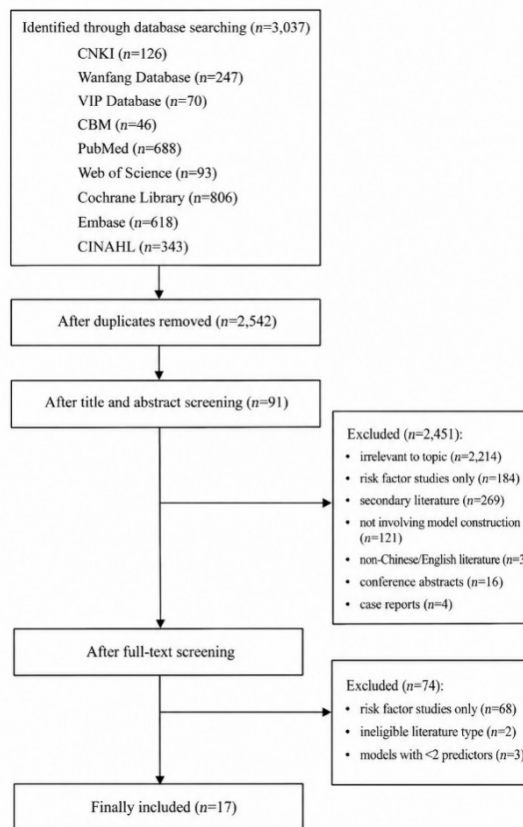


Figure 1. Flow chart of literature search.

3.2. Basic characteristics of included studies

Seventeen studies were finally included, published as follows: 12 from China, 1 from the Netherlands, 1 from South Korea, 1 from Canada, 1 from Argentina, and 1 joint study from 7 countries (Australia, Belgium, United Kingdom, Germany, Spain, Sweden, Netherlands) ^[7–23]. Among them, 8 were Chinese-language articles and 9 were English-language articles ^[7–18, 19–23]; of the 17 studies, 8 were retrospective studies and 9 were prospective studies ^[8–23]. Basic characteristics of included studies are shown in **Table 1**.

Table 1. Basic characteristics of the included studies (n = 17)

Study	Publication year	Country	Study design	Study population	Delirium assessment criteria
Huang et al. ^[23]	2023	China	Retrospective	ICU patients	CAM-ICU (criterion b)
Chang et al. ^[22]	2024	China	Retrospective	Elderly patients with severe pneumonia on mechanical ventilation	CAM-ICU (criterion a)
Zhang et al. ^[21]	2025	China	Prospective	ICU patients	CAM-ICU (criterion a)
Cao et al. ^[20]	2023	China	Prospective	ICU patients	CAM-ICU, ICDSC
Wassenaar et al. ^[19]	2015	Multi-country (7 countries)	Prospective	ICU patients	CAM-ICU positive
Shen et al. ^[18]	2023	China	Retrospective	ICU patients	CAM-ICU (criterion a)
Zhu ^[17]	2017	China	Retrospective	ICU patients	CAM-ICU positive

Dai et al. ^[15]	2024	China	Retrospective	ICU patients with respiratory failure on mechanical ventilation	CAM-ICU (criterion a)
Zhang et al. ^[16]	2024	China	Prospective	ICU patients	CAM-ICU (criterion a) combined with RASS classification
Wang et al. ^[14]	2023	China	Retrospective	ICU patients	ICDSC score ≥ 4
Chen et al. ^[13]	2017	China	Prospective	ICU patients	CAM-ICU (criterion c)
Boogaard et al. ^[12]	2012	Netherlands	Prospective	ICU patients	CAM-ICU (criterion c)
Moon et al. ^[10]	2018	South Korea	Prospective	ICU patients	CAM-ICU (criterion a)
Chen et al. ^[8]	2025	China	Retrospective	ICU patients	CAM-ICU positive
Cherak et al. ^[11]	2020	Canada	Retrospective	ICU patients	ICDSC score ≥ 4
Fan et al. ^[9]	2019	China	Prospective	ICU patients	CAM-ICU (criterion a)
Sosa et al. ^[7]	2018	Argentina	Prospective	ICU patients	CAM-ICU positive

Note: ICDSC: Intensive Care Delirium Screening Checklist; CAM-ICU: Confusion Assessment Method for the ICU; Criterion a: CAM-ICU positive when meeting criteria (1)+(2)+(3) or (4) among four criteria; Criterion b: CAM-ICU single positive when only one positive symptom present; Criterion c: CAM-ICU screening positive when at least one positive screening result during observation period.

3.3. Model development

The included studies had 7 to 4,211 candidate predictors, sample sizes of 100 to 8,878 cases, and outcome events of 21 to 4,431 cases. Regarding modeling methods, 14 studies used logistic regression as the primary modeling method, 1 study combined logistic regression with random forest algorithm, and 1 study applied machine learning (GBT, LR, DL) ^[7–23]. The 17 included studies involved 18 prediction models, of which 13 models constructed risk assessment formulas based on logistic regression coefficients, 2 studies presented prediction models in nomogram form, 1 study presented machine learning models (including models at admission and 24 hours), and 1 study combined logistic regression with random forest algorithm to construct models ^[7–23].

3.4. Model performance and predictors

All included studies reported discrimination assessment indicators during model development or validation, with area under the curve (AUC) ranging from 0.732 to 0.941, indicating overall good model discrimination. Regarding model validation, 10 studies performed internal validation (including cross-validation, Bootstrap method, and random split validation), and 6 studies conducted external validation ^[7,8–14,16,18,19,21,23]. Final models contained 3–15 predictors, with frequently appearing predictors including age, APACHE II score, GCS score, blood urea nitrogen (BUN), mechanical ventilation, hypertension history, diabetes history, and infection. See **Table 2**.

Table 2. Establishment of delirium risk prediction models for adult ICU patients (n = 17)

Study	Candidate Variables (n)	Continuous Variable Processing	Modeling Sample Size (n)	Outcome Events (n)	Primary Modeling Method	Model Validation Method	Calibration Method	Development AUC	Validation AUC	Predictors
Huang et al. ^[23]	28	Partially converted to categorical	247	73	Logistic regression	Ten-fold cross-validation	Calibration curve	0.799	—	Age, hypertension history, APACHE II score, benzodiazepine sedatives, mechanical ventilation
Chang et al. ^[22]	—	Kept continuous	397	80	Logistic regression	—	H-L goodness-of-fit test	0.929	—	Critical condition, cerebrovascular disease, diabetes, prolonged mechanical ventilation, prolonged awakening time after drug withdrawal, elevated blood lactate, elevated CRP, elevated PCT
Zhang et al. ^[21]	—	Partially converted to categorical	436	59	Logistic regression, Random forest	Ten-fold cross-validation	Calibration curve	0.832 / 0.871	0.849	Age, APACHE-II, BMI, Barthel index, sleep, combined medication, emergency admission, etc.
Cao et al. ^[20]	9	Kept continuous	212	37	Logistic regression	—	—	0.882	0.882	Age, cognitive impairment history, alcohol abuse history, treatment experience, MAP, respiratory failure, BUN
Wassenaar et al. ^[19]	18	Kept continuous	1,962	481	Logistic regression	Ten-fold cross-validation	Calibration curve	0.76	0.75	Age, cognitive impairment history, alcohol abuse history, BUN level, admission category, emergency admission, MAP, glucocorticoid use, respiratory failure
Shen et al. ^[18]	—	Partially categorical	100	21	Logistic regression	External validation	H-L goodness-of-fit test	0.891	0.838	Age, hypertension history, invasive mechanical ventilation, hospital stay, APACHE II score, acid-base imbalance

Study	Candidate Variables (n)	Continuous Variable Processing	Modeling Sample Size (n)	Outcome Events (n)	Primary Modeling Method	Model Validation Method	Calibration Method	Development AUC	Validation AUC	Predictors
Zhu ^[17]	—	Partially categorical	147	36	Logistic regression	Random split internal validation	H-L goodness-of-fit test	—	0.749	BUN level, GCS score, infection
Dai et al. ^[15]	—	Converted to categorical	140	62	Logistic regression	External validation	—	0.764	—	Age ≥65 years, pain ≥4, APACHE II score >13, ICU environmental stressor score >72, ICU stay >1 week, diabetes, hypertension
Zhang et al. ^[16]	30	Partially converted to categorical	300	93	Logistic regression	Bootstrap internal validation, External validation	H-L goodness-of-fit test	0.941	0.897	Electrolyte imbalance, catheters ≥2, diuretic use, GCS score, CRP concentration, BUN concentration
Wang et al. ^[14]	ADM-8, 24H-15	Kept continuous	ADM-4,999, 24H-2,773	ADM-1,624, 24H-881	Machine learning (GBT, LR, DL)	External validation	Calibration curve	0.858 (ADM), 0.931 (24H)	—	Age, BMI, dementia history, postoperative monitoring, elective surgery, pre-ICU hospital days, GCS score, initial respiratory rate
Chen et al. ^[13]	11	—	620	160	Logistic regression	Random split internal validation	—	0.78	—	Age, APACHE-II score, coma, emergency surgery, mechanical ventilation, multiple trauma, metabolic acidosis, hypertension history, delirium history, dementia history, dexmedetomidine use
Boogaard et al. ^[12]	25	Partially categorical	1,613	411	Logistic regression	External validation	Calibration curve	0.87	0.84	Age, APACHE-II score, coma, admission category, infection, metabolic acidosis, morphine use, sedative use, urea concentration, emergency admission

Study	Candidate Variables (n)	Continuous Variable Processing	Modeling Sample Size (n)	Outcome Events (n)	Primary Modeling Method	Model Validation Method	Calibration Method	Development AUC	Validation AUC	Predictors
Moon et al. ^[10]	4,211	Kept continuous	3,284	688	Logistic regression	Random split internal validation, External validation	—	0.89	0.72	Age, education years, LOC score, pulse, activity level, internal medicine, BUN, infection, number of catheters, restraint, psychotropic drugs
Chen et al. ^[8]	—	Partially continuous	964	186	Logistic regression	Random split internal validation	Calibration curve	0.732	—	Drug abuse, alcohol abuse, male, maximum potassium, minimum chloride, hospital days, maximum BUN, minimum hematocrit
Cherak et al. ^[11]	10	Partially converted to categorical	8,878	4,431	Logistic regression	Ten-fold cross-validation	—	0.76	0.67-0.78	Age, sex, APACHE II, GCS, SOFA, Charlson index, vasoactive drugs, neuropsychiatric history, CRRT, mechanical ventilation
Fan et al. ^[9]	7	Partially converted to categorical	336	68	Logistic regression	Bootstrap internal validation, Random split internal validation	H-L goodness-of-fit test	0.907	0.9	Chronic disease history, hearing impairment, infection, APACHE II >7.5, sedation/analgesia, catheterization, sleep disorder
Sosa et al. ^[7]	10	Kept continuous	178	49	Logistic regression	External validation	—	—	0.83	Age, APACHE II score, coma, emergency admission, admission category, infection, sedative use, morphine use, urea level, metabolic acidosis

3.5. Model quality assessment

PROBAST was used to evaluate prediction model studies. Regarding participants, 8 studies were rated as “high risk” because 2 studies excluded patients with “Mini-Mental State Examination (MMSE) score < 10 at admission”, preventing the study population from representing the target population and causing model prediction performance to be overestimated ^[8,11,14,15,17,18,22,23]. One study excluded participants with competing

events such as “postoperative death”, reducing model prediction performance ^[17]. The remaining 5 studies’ main risk source was retrospective study design, which is more prone to selection bias and information bias during modeling and validation phases ^[8,11,15,18,22]. Regarding predictors, 1 study’s bias risk was rated as “unclear” because the article did not clearly state whether predictor assessors were blinded to the outcome (delirium) ^[16]. Regarding outcome, 2 studies’ bias risk was rated as “unclear”, and 10 studies were rated as “high risk” ^[7,8–13,15–18,22]. The main reason for high risk in most studies was using predictor information when determining outcomes; 2 studies did not report whether outcome definitions were consistent across all patients ^[7,18]; outcome definition standards and measurement methods were unclear in 5 studies ^[7,9,11,17,18]; 2 studies did not report whether outcome definition excluded predictor information ^[7,18]; and unreasonable time intervals existed between predictor and outcome assessment in 5 studies ^[7,9,11,17,18]. See **Table 3**.

Table 3. Quality assessment for risk of bias and applicability concern of the models (n = 17)

Study	Risk of Bias				Applicability			Overall Assessment	
	Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome	Risk of Bias	Applicability
Huang et al. ^[23]	High risk	Low risk	Low risk	High risk	Low risk	Low risk	Low risk	High risk	Low risk
Chang et al. ^[22]	High risk	High risk	High risk	High risk	Low risk	Low risk	Low risk	High risk	Low risk
Zhang et al. ^[21]	Low risk	Low risk	Low risk	High risk	Low risk	Low risk	Low risk	High risk	Low risk
Cao et al. ^[20]	Low risk	Low risk	Low risk	High risk	Low risk	Low risk	Low risk	High risk	Low risk
Wassenaar et al. ^[19]	Low risk	Low risk	Low risk	High risk	Low risk	Low risk	Low risk	High risk	Low risk
Shen et al. ^[18]	High risk	Low risk	High risk	High risk	Low risk	Low risk	Low risk	High risk	Low risk
Zhu ^[17]	High risk	High risk	High risk	High risk	Low risk	Low risk	Low risk	High risk	Low risk
Dai et al. ^[15]	High risk	Low risk	High risk	High risk	Low risk	Low risk	Low risk	High risk	Low risk
Zhang et al. ^[16]	Low risk	Unclear	Unclear	High risk	Low risk	Low risk	Low risk	High risk	Low risk
Wang et al. ^[14]	High risk	Low risk	Low risk	High risk	Low risk	Low risk	Low risk	High risk	Low risk
Chen et al. ^[13]	Low risk	Low risk	High risk	High risk	Low risk	Low risk	Low risk	High risk	Low risk
Boogaard et al. ^[12]	Low risk	Low risk	High risk	High risk	Low risk	Low risk	Low risk	High risk	Low risk
Moon et al. ^[10]	Low risk	Low risk	High risk	High risk	Low risk	Low risk	Low risk	High risk	Low risk
Chen et al. ^[8]	High risk	High risk	Unclear	High risk	Low risk	Low risk	Low risk	High risk	Low risk
Cherak et al. ^[11]	High risk	Low risk	High risk	High risk	Low risk	Low risk	Low risk	High risk	Low risk
Fan et al. ^[9]	Low risk	Low risk	High risk	High risk	Low risk	Low risk	Low risk	High risk	Low risk
Sosa et al. ^[7]	Low risk	Low risk	High risk	High risk	Low risk	Low risk	Low risk	High risk	Low risk

4. Discussion

4.1. Overall high risk of bias in delirium risk prediction models for ICU patients, with methodological quality needing improvement

Among the 17 included studies, model discrimination was generally good, with all reported development or validation AUC > 0.7; however, overall methodological quality was low, with all identified as high bias risk due to deficiencies in statistical analysis, manifested as insufficient events per variable (EPV), improper continuous variable processing, insufficient explanation or handling of missing data, and lack of external validation. Systematic evaluation of included studies using PROBAST tool revealed bias risk mainly concentrated in the following aspects: ① Regarding participants: most studies used retrospective design,

and some excluded high-risk populations such as those with cognitive impairment, prone to selection bias; ② Regarding predictors: most data came from retrospective medical records, with inconsistent definitions and measurement methods, and some factors had timing bias; ③ Regarding outcome: predictor information was used in outcome determination, with unreasonable assessment time intervals; ④ Regarding analysis: insufficient sample size (most EPV < 20), improper continuous variable processing, unclear missing data handling, variable selection involving only univariate analysis, and insufficient model validation (lack of internal validation or only random split processing). This greatly reduced reliability of model performance assessment and generalization performance. Therefore, future studies should strictly follow PROBAST specifications, adopt prospective multicenter design, ensure EPV $\geq$ 20, implement scientific and effective variable selection methods, and prioritize Bootstrap method or cross-validation for internal validation while conducting multicenter external validation to construct more reliable ICU delirium risk prediction models.

4.2. Certain commonalities exist among predictors across different delirium risk prediction models

Results showed that although variables included in each model had different emphases, they showed obvious commonalities in predictors. Age, APACHE II score, infection/sepsis, metabolic acidosis, and sedative use were common and stable core prediction indicators across multiple model types. Reasons are as follows: with increasing age, elderly patients experience decreased brain reserve function and increased susceptibility to inflammation and metabolic disorders ^[25]; higher APACHE II scores indicate more severe physiological disorders and systemic inflammatory responses, potentially triggering central nervous system inflammatory responses, disrupting neurotransmitter balance, and ultimately causing delirium ^[26]; infection or sepsis can induce delirium by promoting inflammatory responses that damage the blood-brain barrier, activate microglia, induce neuroinflammation, and interfere with neurotransmitter metabolism; metabolic acidosis affects neuronal function and energy metabolism by reducing pH, exacerbating internal environment disorders and promoting delirium occurrence ^[27]; sedative use (especially benzodiazepines) inhibits the central cholinergic system, disrupts sleep cycles, and increases delirium risk ^[28]. Notably, some factors such as GCS score and mechanical ventilation had relatively lower consensus but still held important value in specific clinical contexts. For example, GCS score can reflect consciousness level and has important value for assessing consciousness state in postoperative patients with traumatic brain injury ^[29]; prolonged mechanical ventilation duration can cause acute changes in patients' cognitive and consciousness states ^[30]; cardiopulmonary bypass time has critical value in application to post-cardiac surgery patients. Based on this, during clinical practice, stratified screening strategies combining core factors with scenario-specific factors should be appropriately implemented: for modifiable factors such as sedative use and mechanical ventilation, dynamic dose adjustment and daily spontaneous breathing trials should be managed proactively; for high-risk populations such as elderly patients and those with high APACHE II scores, early warning stratified monitoring programs should be established. Additionally, medical-nursing communication and handover systems should be strengthened to ensure effective transmission of high-risk patient information across shifts; at the nursing level, early mobilization interventions, sensory stimulation regulation, and sleep protection measures should be implemented to achieve organic combination of risk identification and preventive intervention.

4.3. Clinical screening pathways for ICU delirium risk prediction models and future prospects

Currently, the number of ICU delirium risk prediction models is increasing, but clinical generalizability still requires verification. Reasons include: model generalization ability is generally insufficient, most are constructed based on specific subpopulations, and performance significantly decreases when applied to general ICU patients, such as E-PRE-DELIRIC showing AUC of only 0.54 in cardiac surgery validation^[24,31]; differences exist among delirium assessment tools (CAM-ICU, ICDSC, etc.) used in different studies, limiting comparability between models. Additionally, methodologically, 82.4% of studies relied on logistic regression, making it difficult to capture nonlinear relationships, coupled with problems such as insufficient sample size ($EPV < 20$), improper continuous variable processing, and lack of external validation, leading to high bias risk in studies. Therefore, establishing a stratified screening mechanism is recommended: first evaluate compatibility between model and target population, reviewing whether inclusion criteria, assessment tools, and outcome definitions are consistent; second, assess methodological quality according to PROBAST evaluation, prioritizing models with prospective design, sufficient sample size ($EPV \geq 20$), reasonable variable processing, and multicenter validation; finally, balance prediction performance with clinical operability, focusing on calibration and accessibility of predictors in routine monitoring. Future study design should prioritize prospective multicenter large-sample cohorts and follow reporting standards such as TRIPOD and CHARMS; modeling can introduce machine learning algorithms to explore complex associations. Meanwhile, urgent need exists to unify delirium assessment standards and core variable definitions, promote database sharing, and enhance comparability. Ultimately, validated models should be embedded in electronic medical records or mobile nursing systems to achieve real-time dynamic assessment and interface with ABCDEF bundle care, constructing a decision-making closed loop from early warning to intervention to effectively reduce delirium incidence.

5. Conclusion

This study included 17 studies involving 18 prediction models for adult ICU delirium risk, systematically evaluating participants, predictors, outcomes, and analysis aspects of each model. Results showed that existing adult ICU delirium risk prediction models generally demonstrated good discrimination (AUC range 0.732–0.941), but most studies had high bias risk due to: incomplete model validation system, inconsistent validation methods, improper continuous variable processing, and lack of multicenter large-sample data. Therefore, in future research, referring to PROBAST guidelines, procedures for prediction model construction and validation should be standardized, using combined internal-external validation, cross-validation, and other methods to improve model validation system. Meanwhile, multicenter, large-sample prospective studies should be conducted to enhance model stability and clinical application value. Moreover, during clinical nursing application, models should be carefully selected based on target population, delirium assessment tools, prediction time points, and variable availability, and risk warning should be linked with standardized delirium screening and ABCDEF bundle care interventions to promote translation of prediction models into clinical practice.

Funding

Chinese Nursing Association Scientific Research Project in 2022 (Project No.: CMAPH-NRD2021023)

Disclosure statement

The authors declare no conflict of interest.

References

- [1] American Geriatrics Society, 2015, Abstracted Clinical Practice Guideline for Postoperative Delirium in Older Adults. *Journal of the American Geriatrics Society*, 63(1): 142–150.
- [2] Stollings J, Kotfis K, Chanques G, et al., 2021, Delirium in Critical Illness: Clinical Manifestations, Outcomes, and Management. *Intensive Care Medicine*, 47(10): 1089–1103.
- [3] Wilcox M, Girard T, Hough C, 2021, Delirium and Long Term Cognition in Critically Ill Patients. *BMJ (Clinical Research Ed)*, 373: n1007.
- [4] Vickers A, Holland F, 2021, Decision Curve Analysis to Evaluate the Clinical Benefit of Prediction Models. *The Spine Journal: Official Journal of the North American Spine Society*, 21(10): 1643–1648.
- [5] Wolff R, Moons K, Riley R, et al., 2019, PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Annals of Internal Medicine*, 170(1): 51–58.
- [6] Chen X, Zhang Y, Zhuang Y, et al., 2020, PROBAST: An Assessment Tool for Risk of Bias in Diagnostic or Prognostic Multivariable Prediction Model Studies. *Chinese Journal of Evidence-Based Medicine*, 20(6): 737–744.
- [7] Sosa F, Roberti J, Franco M, et al., 2018, Assessment of Delirium Using the PRE-DELIRIC Model in an Intensive Care Unit in Argentina. *Revista Brasileira de Terapia Intensiva*, 30(1): 50–56.
- [8] Chen D, Yang X, 2025, Developing a Nomogram to Predict the Risk of Delirium in ICU Patients: A Retrospective Cohort Study. *Risk Management and Healthcare Policy*, 18: 3221–3233.
- [9] Fan H, Ji M, Huang J, et al., 2019, Development and Validation of a Dynamic Delirium Prediction Rule in Patients Admitted to the Intensive Care Units (DYNAMIC-ICU): A Prospective Cohort Study. *International Journal of Nursing Studies*, 93: 64–73.
- [10] Moon K, Jin Y, Jin T, et al., 2018, Development and Validation of an Automated Delirium Risk Assessment System (Auto-DelRAS) Implemented in the Electronic Health Record System. *International Journal of Nursing Studies*, 77: 46–53.
- [11] Cherak S, Soo A, Brown K, et al., 2020, Development and Validation of Delirium Prediction Model for Critically Ill Adults Parameterized to ICU Admission Acuity. *PLoS One*, 15(8): e0237639.
- [12] Van Den Boogaard M, Pickkers P, Slooter A, et al., 2012, Development and Validation of PRE-DELIRIC (PREdiction of DELIRium in ICu Patients) Delirium Prediction Model for Intensive Care Patients: Observational Multicentre Study. *BMJ (Clinical Research Ed)*, 344: e420.
- [13] Chen Y, Du H, Wei B, et al., 2017, Development and Validation of Risk-Stratification Delirium Prediction Model for Critically Ill Patients: A Prospective, Observational, Single-Center Study. *Medicine*, 96(29): e7543.
- [14] Wang M, Kuo Y, Kuo L, et al., 2023, Early Prediction of Delirium Upon Intensive Care Unit Admission: Model Development, Validation, and Deployment. *Journal of Clinical Anesthesia*, 88: 111121.
- [15] Dai X, Xing Y, Guo D, et al., 2024, Construction and Application of Delirium Risk Prediction Model for ICU

- Patients with Respiratory Failure on Mechanical Ventilation. *Heilongjiang Medical Journal*, 48(5): 624–626.
- [16] Zhang H, Gan X, Zhou W, et al., 2024, Construction and Validation of Dynamic Nomogram Prediction Model for Hypoactive Delirium Risk in ICU Patients. *Chinese Journal of Critical Care & Emergency Nursing*, 5(6): 485–492.
 - [17] Zhu X, 2017, Analysis of Risk Factors for ICU Delirium and Construction of Risk Prediction Model, thesis, Third Military Medical University.
 - [18] Shen X, Ma P, Feng G, et al., 2023, Analysis of Risk Factors and Prediction Model for Delirium in Critically Ill ICU Patients. *China Modern Doctor*, 61(4): 45–49.
 - [19] Wassenaar A, Van Den Boogaard M, Van Achterberg T, et al., 2015, Multinational Development and Validation of an Early Prediction Model for Delirium in ICU Patients. *Intensive Care Medicine*, 41(6): 1048–1056.
 - [20] Cao Y, Zhang M, Pan G, et al., 2023, Effectiveness and Predictive Performance of Delirium Management Model Based on E-PRE-DELIRIC Risk Stratification in Intensive Care Unit. *Academic Journal of Chinese PLA Medical School*, 44(5): 501–507 + 524.
 - [21] Zhang S, Liu L, Wu Y, 2025, Construction of Delirium Prediction Model for Middle-Aged and Elderly ICU Patients Based on Logistic Regression and Random Forest Algorithm. *Journal of Nurses Training*, 40(24): 2663–2670.
 - [22] Chang J, Ding S, Feng B, 2024, Impact of Delirium During Mechanical Ventilation on Prognosis in Elderly Patients with Severe Pneumonia and Construction of Prediction Model. *Shandong Medical Journal*, 64(31): 30–34.
 - [23] Huang B, Chen Y, Yang L, 2023, Construction and Validation of Nomogram Prediction Model for Delirium in Intensive Care Unit Patients. *Chinese Evidence-Based Nursing*, 9(14): 2601–2605.
 - [24] Song Q, Xu H, Cui H, et al., 2025, Current Status and Influencing Factors of ICU Delirium in Patients After Cardiac Surgery. *Smart Healthcare*, 11(12): 29–36.
 - [25] Shi Q, Mu X, Zhang C, et al., 2019, Risk Factors for Postoperative Delirium in Type A Aortic Dissection Patients: A Retrospective Study. *Medical Science Monitor: International Medical Journal of Experimental and Clinical Research*, 25: 3692–3699.
 - [26] Li C, Li D, Bai S, et al., 2019, Progress in Pathophysiological Mechanisms and Risk Factors Prevention and Treatment of Postoperative Delirium. *Chinese Journal of Gerontology*, 39(5): 1273–1277.
 - [27] Cortese G, Burger C, 2017, Neuroinflammatory Challenges Compromise Neuronal Function in the Aging Brain: Postoperative Cognitive Delirium and Alzheimer's Disease. *Behavioural Brain Research*, 322(Pt B): 269–279.
 - [28] Jayaswal A, Sampath H, Soohinda G, et al., 2019, Delirium in Medical Intensive Care Units: Incidence, Subtypes, Risk Factors, and Outcome. *Indian Journal of Psychiatry*, 61(4): 352–358.
 - [29] Reznik M, Kalagara R, Moody S, et al., 2021, Common Biomarkers of Physiologic Stress and Associations with Delirium in Patients with Intracerebral Hemorrhage. *Journal of Critical Care*, 64: 62–67.
 - [30] Bulic D, Bennett M, Rodgers H, et al., 2017, Delirium After Mechanical Ventilation in Intensive Care Units: The Cognitive and Psychosocial Assessment (CAPA) Study Protocol. *JMIR Research Protocols*, 6(2): e31.
 - [31] Gao W, Zhang Y, Jin J, 2022, Validation of E-PRE-DELIRIC in Cardiac Surgical ICU Delirium: A Retrospective Cohort Study. *Nursing in Critical Care*, 27(2): 233–239.

Publisher's note

Bio-Byword Scientific Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.