Original Article

Risk predictive models for delirium in the intensive care unit: a systematic review and meta-analysis

Xiangping Chen[1], Yuewen Lao[1], Yi Zhang[1], Lijie Qiao[1], Yiyu Zhuang[2]

[1]Department of Critical Care Medicine, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China; [2]Department of Nursing, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou, China

Contributions: (I) Conception and design: X Chen, Y Lao; (II) Administrative support: Y Zhuang; (III) Provision of study materials or patients: Y Zhang, L Qiao; (IV) Collection and assembly of data: X Chen, Y Lao; (V) Data analysis: X Chen, Y Zhang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Yiyu Zhuang. Department of Nursing, Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, 3#, Qingchun East Road, Hangzhou 310016, China. Email: zhuangyy@zju.edu.cn.

Background: An emerging approach to prevent delirium in an intensive care unit is the use of risk prediction models. At present, there is no scientific comparison of the predictive effect of the prediction model. This systematic review and meta-analysis aimed to compare the performance of available delirium risk prediction models for intensive care units.

Methods: As of June 1st, 2019, articles on delirium prediction models of the intensive care patients were searched in the Cochrane Library, PubMed, Embase, Web of Science, CINAHL, ProQuest, and four Chinese databases. Studies describing the development or validation of risk prediction models for predicting delirium in ICU patients were included. The Prediction model Risk of Bias Assessment Tool (PROBAST) was used to assess the quality of included studies. A meta-analysis of the predictive performance was performed using the forest plot package in R3.6.1.

Results: A total of 21 studies with 14 models were included in this article. PRE-DELIRIC, E-PRE-DELIRIC, and recalibrated PRE-DELIRIC model were the most popular models, which had been externally validated in at least two studies. The pooled area under the receiver operator characteristic curve (AUC) were 0.844 (95% CI: 0.793–0.896), 0.763 (95% CI: 0.680–0.846) and 0.776 (95% CI: 0.738–0.813) respectively. Most of the other models were with C-statistics above 0.7.

Conclusions: The E-PRE-DELIRIC model, PRE-DELIRIC model, or both are recommended to predict ICU delirium risk. However, the recommendation should be considered with caution because of substantial heterogeneity. The protocol was registered with PROSPERO (CRD42019130802).

Keywords: Delirium; intensive care units; meta-analysis; prediction model

Submitted May 28, 2020. Accepted for publication Sep 22, 2020.
doi: 10.21037/apm-20-1183
View this article at: http://dx.doi.org/10.21037/apm-20-1183

Introduction

Delirium is an acute, reversible, and widespread cognitive disorder characterized by fluctuating consciousness, inattention, and disorganized thinking, with a prevalence of 30% to 80% in ICU (1-3). ICU delirium is associated with a lot of severe consequences, such as elevated ICU mortality, prolonged duration of mechanical ventilation and length of hospital stays, increased medical costs, and reduced patients' long-term cognitive function and quality of life (4-6). Besides, ICU staff caring for delirium patients experienced an increased workload and severe psychological stress (7,8). Furthermore, family members will suffer distress when they witness their loved one in this altered state (9).

Evidence shows that delirium can be prevented. The most commonly recommended measure is implementing

delirium prevention clustering measures, such as the ABCDEF bundle (10). Along with universal prophylaxis, tailor preventive measures depending on risk factors and risk levels may be a more effective strategy (11). The risk prediction model was used to predict the independent influence of various risk factors on the disease's occurrence and evaluate the possibility of conditions. Prediction of the ICU delirium risk can help medical personnel effectively identify high-risk patients and develop appropriate clinical decisions. For instance, it can be utilized to better inform family members about the patient's risk of developing delirium and stratify patients in delivering future delirium prevention studies (11-14).

Many scholars have previously established models for predicting delirium of ICU patients based on single-center or multi-center research designs. Additionally, prediction models of delirium for postoperative patients and elderly hospitalized patients were systematically reviewed (15,16), but they were not explicitly targeted at ICU patients. Therefore, the purpose of this study was to comprehensively search studies on prediction models of delirium risk in intensive care unit patients and systematically review them from the aspects such as the basic features, statistical methods, the methodological quality, predictors, and performance of various prediction model, to provide a theoretical basis for clinical practice and scientific research. We present the following article in accordance with the PRISMA reporting checklist (available at http://dx.doi.org/10.21037/apm-20-1183).

Methods

Search strategy

The following databases were searched: Cochrane Library, PubMed, Embase, Web of Science, CINAHL Complete, ProQuest, Chinese Biomedical Literature Database (CBM), China National Knowledge Infrastructure (CNKI), Wanfang and VIP database, from inception to 1st June 2019. Medical Subject Headings (MeSH) terms and free words were combined: (predict* OR prognos* OR risk*) AND (critical care OR critical illness OR intensive car* OR icu* OR critically ill) AND (delirium OR icu psychosis OR icu syndrome OR acute confusional state OR acute brain dysfunction). The computer search was supplemented by hand search of citations. Literature was searched by 2 researchers (Y Zhang and L Qiao) independently, and limited to papers published in English and Chinese. The protocol was registered in PROSPERO (CRD42019130802). Available from: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42019130802. The search strategy for PubMed was outlined below (((((((predict*[Title/Abstract]) OR prognos*[Title/Abstract]) OR risk*[Title/Abstract])) OR "Prognosis"[Mesh])) AND (((((("Critical Care"[Mesh] OR "Critical Care Nursing"[Mesh]) OR "Intensive Care Units"[Mesh]) OR "Critical Illness"[Mesh])) OR ((((((critical care[Title/Abstract]) OR intensive care unit*[Title/Abstract]) OR critical illness[Title/Abstract]) OR icu*[Title/Abstract]) OR intensive care[Title/Abstract]) OR critically ill[Title/Abstract]))) AND (("Delirium"[Mesh]) OR (((((delirium[Title/Abstract]) OR icu psychosis[Title/Abstract]) OR icu syndrome[Title/Abstract]) OR acute confusional state[Title/Abstract]) OR acute brain dysfunction[Title/Abstract])) .

Inclusion and exclusion criteria

Articles were included according to the following criteria: (I) Study participants: ICU patients (age 18 years or older); (II) Study content: a study using multivariable design to describe the development or validation of patient delirium risk prediction model (at least two predictors); (III) Study type: cohort or case-control studies. Meanwhile, the exclusion criteria were as follows: (I) A risk prediction model was not established; (II) The model building process or method was not described; (III) The outcome was evaluated in the absence of the use of tools for reliability and validation; (IV) The original text was not available or the information was incomplete; (V) To avoid extracting the same data. We only included the latest or most complete study.

Data extraction and study quality

The design of the data extraction table included the year of publication, country, name of the risk prediction model, participants, research design, sample size, measurement of results, the process of testing results, risk factors, development and verification methods (statistical methods used and outcomes, i.e., % delirium), and the performance of the prediction model. For evaluating the model's prediction performance, model discrimination and calibration are two important dimensions (17). The discrimination reflects the ability of the prediction model to distinguish whether the final event will occur or not, and the most generally used evaluation index is AUC value (18).

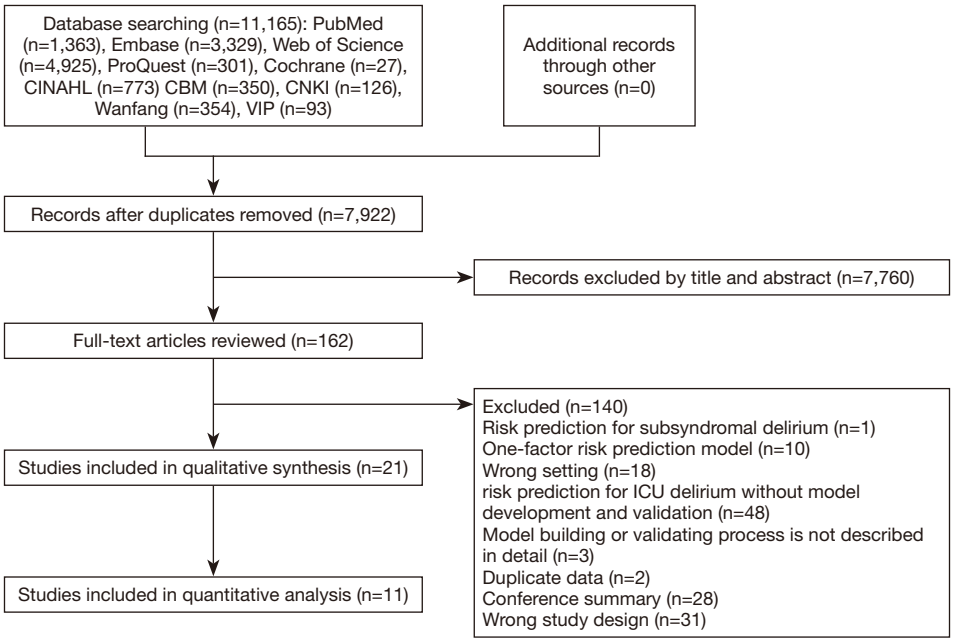

Figure 1 PRISMA Search flow diagram.

The calibration reflects the consistency between the predicted results and the model's observed results (17). Two reviewers (X Chen and Y Lao) collected data independently. The risk of bias (ROB) and clinical applicability was assessed by the PROBAST tool (19,20) from 20 questions in four key domains: participants, predictors, outcome, and analysis. Signaling questions are rated as: "yes/probably yes", "probably no/no", and "no information". According to the signal problem and the author's judgment, each field was divided into "high", "low" and "unclear". All the methodological quality evaluations included in the literature were independently completed by two reviewers (X Chen and Y Zhang). If there were differences, the decision was made by another researcher (Y Lao).

Statistical analysis

All statistical analyses were performed using the forest plot package in R3.6.1. We plotted AUC to summary the predictive performance by forest plot. The Q test was used to determine whether there was heterogeneity between studies. If $I^2 \leq 50\%$, it was considered that the studies did not show statistical homogeneity, and a fixed-effects model was used for analysis. If $I^2 > 50\%$, it was considered that the studies were statistically heterogeneous, and the data were analyzed by using a random-effects model; the sensitivity analysis was used to examine the robustness of the findings. Descriptive analysis was used for studies that data could not be merged.

Results

Eligible articles and study characteristics

According to the search strategy, 11,165 articles were initially obtained, and 21 articles were finally registered (*Figure 1*) (21-41). *Table 1* summarized the primary characteristics of the included study. This review included 14 prediction models. Totally, 22,113 patients were included in our study; most patients were from mixed ICU. Nineteen studies (90.5%) used the Confusion Assessment Method for Intensive Care Unit (CAM-ICU) to evaluate the results, and delirium assessment frequency in 15 studies was two or more times a day. Three models (19,36,41) were dynamic prediction models (daily prediction), and the rest were static prediction models, of which five prediction models used at ICU admission and six models used within 24 hours after admission to ICU. Only two studies followed the TRIPOD statement.

Study quality

The ROB and clinical applicability were assessed with

the PROBAST tool (19) (*Table 2*). Although all 21 studies demonstrated a low applicability risk, only one of the articles (25) had low bias risk, and the remaining studies were associated with a high ROB. High ROB was mostly in the domain of analysis. The reasons for the high risk of bias in the analysis are as follows: (I) Insufficient sample size. Five studies (23-25,30,32) met the sample size standard (events per variable, EPV >10) in model development studies. Whereas validation studies, only six studies (25,30,31,35-38) met the standard of including at least 100 participants with the outcome. (II) Continuous predictors handled unreasonably. There were seven models (21-24,27,29,41) in which collected continuous predictors were converted into ≥ two categories without using a prespecified method. (III) Selection of predictors based on univariable analysis. (IV) Lack of standardized evaluation of model calibration. Model calibration was not assessed or only assessed using the Hosmer-Lemeshow test. Only six studies (23-25,31,36,37) used a calibration plot or belt to assess the calibration. (V) Lack of internal validation techniques. Four studies (24,27,30,40) lacked internal validation in model development. (VI) Predictors and their weights were inconsistent with the report or not reported. The final model calculation formula was not reported after model development in three studies (21,22,36), and the regression coefficients in the final model were adjusted in another three studies (27,39,41). In terms of predictors and outcomes, the risk of bias was unclear in six articles (26,27,30,33-35) because they did not report blind methods in the literature. Details were shown in Appendix 1.

Statistical methods of models

In this review, 14 studies reported model development (21-25,27-30,32,36,39-41), and 19 studies reported model validation (21-23,25-29,31-41). The method of model development was logistic regression. Six studies (22,27,29,39-41) selected variables for inclusion in the model based on univariable analysis. For 14 included prediction models, ten models were verified internally, of which Bootstrap was used in five studies (21-23,36,39). Random split validation was used in four models (25,28,29,32) internal validation, and the bootstrap validation and random split validation were used in one model (41). In terms of external validation, ten articles reported calibration, four articles (22 27,29,39) carried out the Hosmer-Lemeshow test, and six articles (23-25,31,36,37) reported the calibration plot or belt. Specific statistical methods for the included models were shown in *Table 3*.

Predictors of models

Among the 14 prediction models included, the number of candidate predictors was 11–116, and the number of predictors in the final model was 3–14, with a total of 40 predictors. Furthermore, the predictors included in the prediction models were divided into two categories: predisposing factors and precipitating factors. In this systematic review, the most common risk factor for delirium in the intensive care unit was cognitive dysfunction (including the history of Alzheimer's disease, coma, and cognitive impairment). In terms of precipitating factors, sedation and analgesics were the most common predictors, followed by infection and mechanical ventilation. The predictors in the final model were presented in *Table 4*.

Predictive performance

Of the 14 articles associated with model development, 13 reported discrimination, and the C-statistics ranged from 0.73 to 0.93. For model validation, 14 articles reported AUC, ranging from 0.62 to 0.94, indicating that the included models were quite different. Nine studies were calibrated using the Hosmer-Lemeshow test or a calibration plot, and all of them showed moderate goodness of fit (P>0.05).

In this review, three delirium prediction models were externally validated in at least two studies, including Prediction of Delirium in ICU Patients (PRE-DELIRIC model), Early Prediction of Delirium in ICU Patients (E-PRE-DELIRIC model), and recalibrated PRE-DELIRIC model. Pooled AUC was conducted, and the result was shown in *Figure 2*. For PRE-DELIRIC model and E-PRE-DELIRIC model, heterogeneity among studies were high (I^2=94.66%, P<0.001 and I^2=95.83%, P<0.001). The pooled AUC was calculated using a random-effects model. The results showed the good accuracy of PRE-DELIRIC model and E-PRE-DELIRIC model (pooled AUC =0.84; 95% CI: 79.3–89.6% and pooled AUC =0.76; 95% CI: 68.0–84.6%). The sensitivity analysis demonstrated that a single study could not significantly impact the pooled AUC with 95% CIs. For calibration of the PRE-DELIRIC model, heterogeneity among articles was low (I^2=0%, P=0.32). The pooled AUC was calculated using a fixed-effects model, and the result also indicated good accuracy (pooled AUC =0.78; 95% CI: 73.8–81.3%).

Table 1 Primary characteristics of the prediction model included in the review

Study ID	Country	Participants	Study design	Sample size (n)	Delirium measurement	Frequency of delirium assessment	Timepoint of variables	TRIPOD
Pisani (2007) (21)	America	Age ≥60, patients in medical ICU	Prospective cohort	304	CAM-ICU	–	At ICU admission	No
Katznelon (2009) (22)	Canada	Age ≥18, patients undergoing cardiac surgery with cardiopulmonary bypass in CCU	Prospective cohort	1,059	CAM-ICU	Every 12 h	At ICU admission	No
van den Boogaard (2012) (23)	Netherlands	Age ≥18, patients in ICU	Prospective cohort	3,056	CAM-ICU	At least twice daily	Within 24h after ICU admission	No
van den Boogaard (2014) (24)	6 countries[a]	Age ≥18, patients in ICU	Prospective cohort	1,824	CAM-ICU	At least twice daily	Within 24h after ICU admission	No
Wassenaar (2015) (25)	7 countries[b]	Age ≥18, patients in ICU	Prospective cohort	2,917	CAM-ICU	Once every 8 or 12 h shift	At ICU admission	Yes
Li (2015) (26)	China	Age ≥18, patients in medical ICU	Prospective cohort	306	CAM-ICU	3 times in ICU	Within 24h after ICU admission	No
Yuan (2017) (27)	China	Age ≥18, patients in mixed ICU	Case-control	304	CAM-ICU	Twice daily	Within 24h after ICU admission	No
Chen (2017) (28)	China	Age ≥18, patients in mixed ICU	Prospective cohort	620	CAM-ICU	At both 9 AM and 5 PM	Within 24h after ICU admission	No
Zhu (2017) (29)	China	Age ≥18, patients in mixed ICU	Case-control	300	CAM-ICU	Every 8 h	Within 24h after ICU admission	No
Luo (2017) (30)	China	Age ≥18, patients in mixed ICU	Case-control	1,524	CAM-ICU	–	Within 24h after ICU admission	No
Lee (2017) (31)	China	Age ≥18, patients after cardiac surgery	Prospective cohort	600	CAM-ICU	Every 8 h	–	Yes
Moon (2018) (32)	Korea	Age ≥18, patients in medical or surgical ICU	Prospective cohort	4,303	CAM-ICU	Once daily	Auto-extracted daily	No
Deng (2018) (33)	China	Age ≥18, patients in mixed ICU	Case-control	265	CAM-ICU	Three times daily	Within 24h after ICU admission	No
Linkaitė (2018) (34)	Lithuania	Age ≥18, patients in mixed ICU	Prospective cohort	38	CAM-ICU	–	–	No
Sosa (2018) (35)	Argentina	Age ≥18, patients in mixed ICU	Prospective cohort	178	CAM-ICU	Once daily	Within 24h after ICU admission	No
Marra (2018) (36)	America	Age ≥18, patients in medical or surgical ICU	Prospective cohort	810	CAM-ICU	Twice daily	Auto-extracted daily	No

Table 1 (*Continued*)

Table 1 (*Continued*)

Study ID	Country	Participants	Study design	Sample size (n)	Delirium measurement	Frequency of delirium assessment	Timepoint of variables	TRIPOD
Wassenaar (2018) (37)	7 countries[c]	Age ≥18, patients in ICU	Prospective cohort	2,178	CAM-ICU or ICDSC	Every 12 h	–	No
Green (2019) (38)	Australian	Age ≥18, patients in ICU	Prospective cohort	455	CAM-ICU	Twice daily	–	No
Chaiwat (2019) (39)	Thailand	Age ≥18, patients in surgical ICU	Prospective cohort	250	CAM-ICU	Twice daily	At ICU admission	No
Xing (2019) (40)	China	Age ≥18, patients in surgical ICU	Case-control	265	ICDSC	Twice daily	At ICU admission	No
Fan (2019) (41)	China	Age ≥18, patients in ICU	Prospective cohort	560	CAM-ICU	Twice daily	Daily	No

TRIPOD, Transparent Reporting of a multivariable prediction Model for Individual prognosis or Diagnosis; ICU, Intensive Care Unit; CCU, Cardiac Care Unit; CAM-ICU, Confusion Assessment Method for ICU;ICDSC, Intensive Care Delirium Screening Checklist; a, Australia, Belgium, Germany, Spain, Sweden; b, Australia, Belgium, England, Germany, Spain, Sweden, and the Netherlands; c, Australia, Belgium, Canada, Denmark, Portugal, USA, and the Netherlands; –, no information.

Discussion

In this systematic review, we found 21 studies involving 14 models for predicting ICU delirium. Except for the post-cardiac surgery model (22) and the E-PRE-DELIRIC model (37), whose reported AUC <0.7 in external validation, most models had excellent discrimination (AUC >0.7). However, for the E-PRE-DELIRIC model, the summary AUC score was >0.7. The main reason was that patients received preventive measures according to two models in ref 37, which reduced the incidence of delirium and hence lowered the AUC value of the E-PRE-DELIRIC model. In this systematic review, nine studies reported the calibration and showed that the calibration accuracy was satisfactory overall (P>0.05). However, due to insufficient data, quantitative synthesis cannot be conducted.

At present, the PRE-DELIRIC and E-PRE-DELIRIC model are the most common prediction model of ICU delirium. The meta-analysis indicates that they have a good ability to distinguish delirium. PRE-DELIRIC model has been externally validated by scholars in China (26), Lithuania (34), Argentina (35), Australian (38), and pooled AUC are 0.844, which indicates superior performance. However, the results are partially different from the present one meta-analysis of PRE-DELIRIC by Ho (42), which shows a cumulative AUC is 0.78 (95% CI: 0.74–0.81). The main reason is that we analyze the performance of PRE-DELIRIC without including the recalibrated model, whose accuracy is not as good as the former one. The calibrated model is considered a newly developed model and should not be included. Plus, we have included two Chinese studies whose AUC is nearly 0.93, which attributes to better performance. The PRE-DELIRIC model contains ten predictive factors: age, APACHE II score, admission category, coma, infection, metabolic acidosis, emergency admission, blood urea nitrogen and sedative use, morphine dose within 24 hours. It was designed to predict delirium in surgery, medical, trauma, or neurology adult patients. According to the prediction model, the risks are divided into 4 levels: low-risk group (0–20%), intermediate-risk group (20–40%), high-risk group (40–60%) and extremely high-risk group (>60%).

However, ICU delirium tends to occur in the first day of ICU stay (39), PRE-DELIRIC model only predicts delirium 24 h later after ICU admission. E-PRE-DELIRIC model could predict delirium risk as soon as the patient entered the ICU, which had advantages in a clinical application (43). It includes nine predictors: age, history

Table 2 ROB and clinical applicability of included studies

Study ID	ROB				Applicability			Overall	
	Participants	Predictors	Outcome	Analysis	Participants	Predictors	Outcome	ROB	Applicability
Pisani (2007) (21)	+	+	+	–	+	+	+	–	+
Katznelon (2009) (22)	+	+	+	–	+	+	+	–	+
van den (2012) (23)	+	+	+	–	+	+	+	–	+
van den (2014) (24)	+	+	+	–	+	+	+	–	+
Wassenaar (2015) (25)	+	+	+	+	+	+	+	+	+
Li (2015) (26)	+	?	?	–	+	+	+	–	+
Yuan (2017)(27)	–	?	?	–	+	+	+	–	+
Chen (2017) (28)	+	+	+	–	+	+	+	–	+
Zhu (2017) (29)	+	+	+	–	+	+	+	–	+
Luo (2017) (30)	–	?	?	–	+	+	+	–	+
Lee (2017) (31)	+	+	+	–	+	+	+	–	+
Moon (2018) (32)	+	+	+	–	+	+	+	–	+
Deng (2018) (33)	+	?	?	–	+	+	+	–	+
Linkaitė (2018) (34)	+	?	?	–	+	+	+	–	+
Sosa (2018) (35)	–	?	?	–	+	+	+	–	+
Marra (2018) (36)	+	+	+	–	+	+	+	–	+
Wassenaar (2018) (37)	+	+	+	–	+	+	+	–	+
Green (2019) (38)	+	+	+	–	+	+	+	–	+
Chaiwat (2019) (39)	+	+	+	–	+	+	+	–	+
Xing (2019) (40)	+	+	+	–	+	+	+	–	+
Fan (2019) (41)	+	+	+	–	+	+	+	–	+

ROB, risk of bias; +, low ROB/low concern regarding applicability; –, high ROB/high concern regarding applicability; ?, indicates unclear ROB/unclear concern regarding applicability.

of cognitive dysfunction, history of alcoholism, blood urea nitrogen, admission disease group, emergency admission, mean arterial pressure (MAP) at admission, glucocorticoid use, and respiratory failure at admission. The risks are divided into four levels: extremely low-risk group (0–10%), low-risk group (10–20%), intermediate-risk group (20–35%), and high-risk group (>35%). Both models are suitable for mixed ICU patients. However, the E-PRE-DELIRIC model (the pooled AUC was 0.78) was not as good as that of the PRE-DELIRIC model. Wassenaar (37) conducted a prospective study of 2,178 ICU patients in 11 hospitals in seven countries. It was concluded that the E-PRE-DELIRIC model, in conjunction with the PRE-DELIRIC model, is more sensitive in the prediction of delirium. Hence, it is recommended to use E-PREDELIRIC at ICU admission. If delirium did not occur, PREDELIRIC should be completed within 24 h to improve the detection of low-risk cases.

However, there is substantial heterogeneity between different studies. On the one hand, the difference in study characteristics (such as where studies were conducted and what participants were included) (*Table 1*) might lead to significant heterogeneity. On the other hand, the diversity of data collection among studies might affect the model's performance. The resulting variable, delirium, is a fluctuating, sudden acute condition that requires objective and accurate assessment. However, the outcome assessed by clinical staff and the frequency of assessment was quite non-

Table 3 Statistical methods and performance of prediction models

Study ID	Model name	Statistical method (A)	Statistical method (B)	AUC (A)	AUC (B)	Calibration (Statistical method and P value)
Pisani (2007) (21)	Pisani Model	Logistic regression	Internal validation	0.78	–	–
Katznelon (2009) (22)	Katznelon Model	Logistic regression	Internal validation	0.77	0.75	H-L test P=0.3
van den Boogaard (2012) (23)	PRE-DELIRIC Model	Logistic regression	Internal and external validation	0.87	0.89, 0.84	Calibration plot P=0.76
van den Boogaard (2014) (24)	Recalibrated PRE-DELIRIC Model	Logistic regression	–	0.76	–	Calibration plot P=1.09
Wassenaar (2015) (25)	E-PRE-DELIRIC Model	Logistic regression	Internal validation	0.76	0.75	Calibration plot P=0.96
Li (2015) (26)	–	–	External validation	–	PRE-DELIRIC:0.936	–
Yuan (2017) (27)	Yuan Model	Logistic regression	External validation	0.86	0.739	H-L test P>0.05
Chen (2017) (28)	Lanzhou Model	Logistic regression	Internal validation	0.78	–	–
Zhu (2017) (29)	Zhu model	Logistic regression	Internal validation	–	0.749	H-L test P=0.17
Luo (2017) (30)	Luo Model	Logistic regression	–	0.93	–	–
Lee (2017) (31)	–	–	External validation	–	Recalibrated PRE-DELIRIC: 0.75; Katznelon: 0.62	Calibration belt –
Moon (2018) (32)	Auto-Del RAS Model	Logistic regression	Internal and external validation	0.89	0.90, 0.72, 0.93, 0.85	–
Deng (2018) (33)	–	–	External validation	–	PRE-DELIRIC: 0.93; E-PRE-DELIRIC: 0.90	–
Linkaitė (2018) (34)	–	–	External validation	–	PRE-DELIRIC: 0.71	–
Sosa (2018) (35)	–	–	External validation	–	PRE-DELIRIC: 0.84	–
Marra (2018) (36)	ABD-pm Model	Logistic regression	Internal validation	0.73	–	Calibration plot –
Wassenaar (2018) (37)	–	–	External validation	–	PRE-DELIRIC 0.74; E-PRE-DELIRIC: 0.68	Calibration plot –
Green (2019) (38)	–	–	External validation	–	PRE-DELIRIC: 0.79; Recalibrated PRE-DELIRIC: 0.79; E-PRE-DELIRIC: 0.72; Lanzhou: 0.77	–
Chaiwat (2019) (39)	POD Model	Logistic regression	Internal validation	0.84	0.82	H-L test 0.39
Xing (2019) (40)	Xing Model	Logistic regression	Internal validation	0.83	–	–
Fan (2019) (41)	DYNAMIC-ICU Model	Logistic regression	Internal validation	0.91	0.90	–

A, Model Development; B, Model validation; POD, postoperative delirium; H-L test, Hosmer-Lemeshow test; –, no information.

Table 4 Predictors of selected models

Model name	Final predictors (n)	Predictors in final model	Candidate predictors (n)
Pisani Model (21)	4	Dementia, receipt of benzodiazepines before ICU admission, elevated creatinine level, low arterial pH	47
Katznelon Model (22)	6	Older age, preoperative depression, preoperative renal dysfunction, complex cardiac surgery, perioperative intracortical balloon pump support, and massive blood transfusion	18
PRE-DELIRIC Model (23)	10	Age, APACHE-II score, admission group, coma, infection, metabolic acidosis, use of sedatives and morphine, urea concentration, and urgent admission	25
recalibrated PRE-DELIRIC Model (24)	10	Age, APACHE-II, urgent and admission category, infection, coma, sedation, morphine use, urea level, metabolic acidosis	25
E-PRE-DELIRIC Model (25)	9	Age, history of cognitive impairment, history of alcohol abuse, BUN, admission category, urgent admission, mean arterial blood pressure, use of corticosteroids, and respiratory failure	18
Yuan Model (27)	6	History of hypertension, hypoxemia, use of benzodiazepines, deep sedation, mechanical ventilation, infection	14
Lanzhou Model (28)	11	Age, APACHE-II score, coma, emergency operation, mechanical ventilation, multiple trauma, metabolic acidosis, history of hypertension, delirium and dementia, and application of Dexmedetomidine Hydrochloride	11
Zhu Model (29)	3	Urea concentration, infection, level of consciousness	22
Luo Mode l(30)	5	History of alcohol abuse, mechanical ventilation, APACHE-II, Total bilirubin, blood urea nitrogen	24
Auto-DelRAS Model (32)	11	Age, education, level of consciousness score, pulse, activity level, medical department, BUN level, infection, total number of catheters, restraints, and psychopharmacologic medications	116
ABD-pm Model (36)	14	ICU admission included the following: age, ICU type, use of medications to treat Alzheimer's disease, APACHE II and mechanical ventilation. Daily factors included: current brain function status, mechanical ventilation, sepsis, SOFA score, length of stay, administration of benzodiazepines, opiates, propofol and antipsychotic agents	15
POD Model (39)	6	Age, diabetes mellitus, severity of disease (SOFA score), perioperative use of benzodiazepine and mechanical ventilation	25
Xing Model (40)	5	Physiological and Operative Severity Score for the enumeration of Mortality and morbidity, acid-base disturbance, history of coma, diabetes and hypertension	10
DYNAMIC-ICU Model (41)	7	History of chronic diseases, hearing deficits, infection, higher APACHE II scores at admission, the use of sedatives and analgesics, indwelling catheter, and sleep disturbance	27

APACHE-II, Acute Physiology and Chronic Health Evaluation II. SOFA, Sequential Organ Failure Assessment. BUN, blood urea nitrogen.

systematic in some studies (26,35), decreasing consistency. Data loss is also a significant source of heterogeneity. For example, van den Boogaard (23) reported that some values (such as urea, metabolic acidosis, and sodium) were missing, and the mean normal value was imputed. To some extent, the variance and standard deviation of the data will be reduced, and the degree of variation will be underestimated. Therefore, the results of the meta-analysis should be carefully considered.

In this systematic review, the number of predictors in the final model ranged from 3 to 14. Prior cognitive impairment was the most common predisposing factor in 14 included models. Moreover, sedative and analgesics were the most precipitating factors, followed by infection and

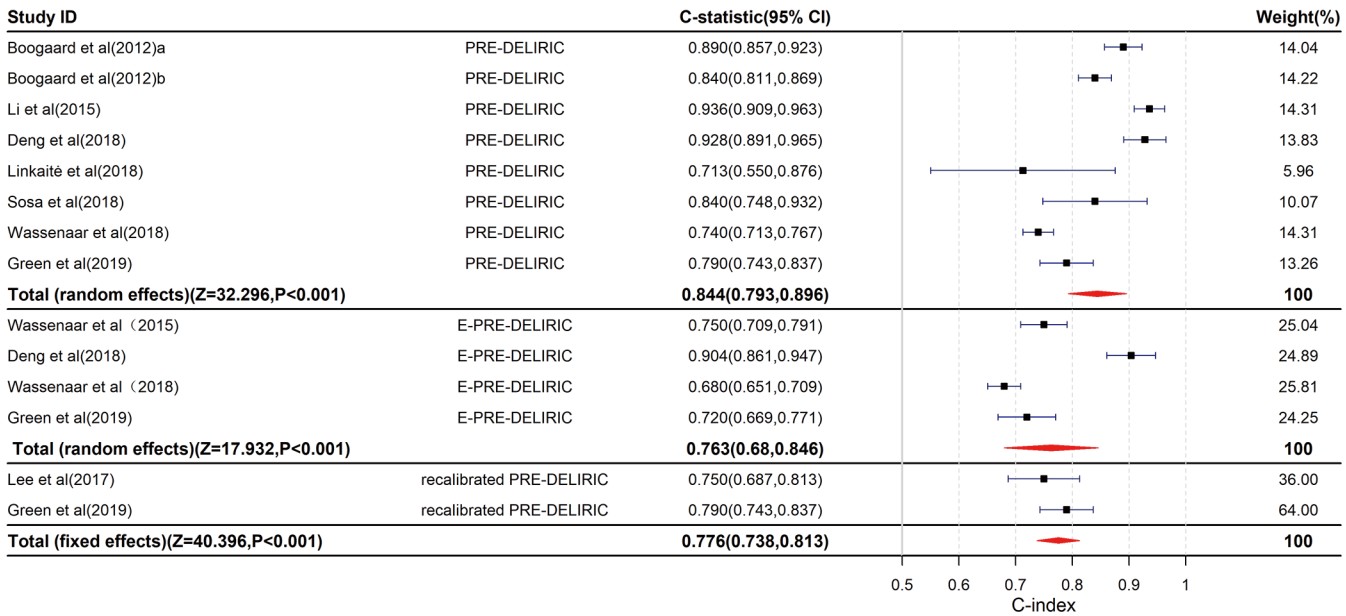

Figure 2 Meta-analysis of delirium prediction models.

mechanical ventilation. These risk factors were included in the PRE-DELIRIC and E-PRE-DELIRIC model, the most extensively used models. However, the definition of prior cognitive impairment was inconsistent, and the drugs used for sedative and analgesics were not the same among studies, which would make predictors' clinical application challenging.

In this systematic evaluation, the quality of evidence was low because of the overall high ROB. The main reason for high ROB was in the analysis field, such as insufficient sample size, lack of internal or external verification, and selection of prediction factors based on univariate analysis. It is suggested that scholars should follow the rigorous research scheme designed by the PROBAST tool when developing or verifying the delirium risk prediction model in ICU in later research. Besides, the TRIPOD list should be submitted as an attachment when submitting articles on predictive models, so that academic editors, peer-reviewed experts, and researchers can evaluate the methodological quality of the articles. However, in this review, only two articles followed the TRIPOD reporting standard, which might prevent readers from assessing the quality of a prognostic model (44).

In this review, the probability for delirium development was divided into groups, like very low, low, moderate, and high risk of delirium. Sensitivity and specificity were calculated for these different groups. We cannot derive the model's sensitivity and specificity based on the available data, which is regarded as suitable measures to express discriminative abilities. In future research, the overall specificity and sensitivity of the model can be reported. Moreover, it is of considerable significance to promote the external verification and clinical application of the model. Most models have been developed in the past five years, and the number of studies is limited. Specifically, there is a lack of prospective cohort validation studies on population models of intensive care units in different hospitals and regions and comparative studies on the prediction efficiency of different models in the same population. Many models have not been verified externally, so it is difficult to compare their accuracy. At present, few scholars have discussed the application effect of the model in clinical practice (45), which may be related to the increase of staff workload and the tedious calculation of the model. Some scholars believe that intelligent risk prediction is conducive to promoting the model's clinical application of (46). It is necessary to rely on information systems to automatically collect prediction factor information and calculate prediction results in future experiments. At the same time, RCTs should include delirium risk score as inclusion/stratification criteria if a risk cut-off must be chosen to deliver delirium preventive intervention. The stratified research design will make the groups more comparable. In the clinic, we can inform the family caregivers whether the patient is at high risk and

what we can do together to prevent it according to the ICU delirium prediction model.

Conclusions

This meta-analysis indicates that PRE-DELIRIC and E-PRE-DELIRIC model yield a good prediction accuracy, and we suggest they could be the consideration to predict the risk of ICU delirium. However, heterogeneity does exist among studies, so the suggestion should be carefully considered. As for the future research direction, it is suggested that researchers should combine advanced statistical techniques and strictly follow the statement of the TRIPOD statistical performance report.

Acknowledgments

We would like to thank Prof. Qi-Sheng Gao for statistical assistance.
Funding: This work was supported by the Chinese Nursing Association Annual Research Project (Grant number: ZHKY201820) and Zhejiang Medicine Health Science and Technology Project (Grant number: 2018KY472).

Footnote

Reporting Checklist: The authors have completed the PRISMA reporting checklist. Available at http://dx.doi.org/10.21037/apm-20-1183

Peer Review File: Available at http://dx.doi.org/10.21037/apm-20-1183

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/apm-20-1183). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. Mehta S, Cook D, Devlin JW, et al. Prevalence, risk factors, and outcomes of delirium in mechanically ventilated adults. Crit Care Med 2015;43:557-66.
2. Krewulak KD, Stelfox HT, Leigh JP, et al. Incidence and Prevalence of Delirium Subtypes in an Adult ICU: A Systematic Review and Meta-Analysis. Crit Care Med 2018;46:2029-35.
3. Ely EW, Shintani A, Truman B, et al. Delirium as a predictor of mortality in mechanically ventilated patients in the intensive care unit. JAMA 2004;291:1753-62.
4. Salluh JI, Wang H, Schneider EB, et al. Outcome of delirium in critically ill patients: systematic review and meta-analysis. BMJ 2015;350:h2538.
5. Abelha FJ, Luis C, Veiga D, et al. Outcome and quality of life in patients with postoperative delirium during an ICU stay following major surgery. Crit Care 2013;17:R257.
6. Vasilevskis EE, Chandrasekhar R, Holtze CH, et al. The Cost of ICU Delirium and Coma in the Intensive Care Unit Patient. Med Care 2018;56:890-7.
7. Öztürk Birge A, Beduk T. The relationship of delirium and risk factors for cardiology intensive care unit patients with the nursing workload. J Clin Nurs 2018;27:2109-19.
8. LeBlanc A, Bourbonnais FF, Harrison D, et al. The experience of intensive care nurses caring for patients with delirium: A phenomenological study. Intensive Crit Care Nurs 2018;44:92-8.
9. Day J, Higgins I. Adult family member experiences during an older loved one's delirium:a narrative literature review. J Clin Nurs 2015;24:1447-56.
10. Devlin JW, Skrobik Y, Gelinas C, et al. Clinical Practice Guidelines for the Prevention and Management of Pain, Agitation/Sedation, Delirium, Immobility, and Sleep Disruption in Adult Patients in the ICU. Crit Care Med 2018;46:e825-73.
11. Paton L, Elliott S, Chohan S. Utility of the PRE-DELIRIC delirium prediction model in a Scottish ICU cohort. J Intensive Care Soc 2016;17:202-6.
12. Wassenaar A, Rood P, Schoonhoven L, et al. The impact of nUrsiNg DEliRium Preventive INnterventions in the Intensive Care Unit (UNDERPIN-ICU): A study protocol for a multi-centre, stepped wedge randomized controlled

trial. Int J Nurs Stud 2017;68:1-8.
13. Smithburger PL, Korenoski AS, Alexander SA, et al. Perceptions of Families of Intensive Care Unit Patients Regarding Involvement in Delirium-Prevention Activities: A Qualitative Study. Crit Care Nurse 2017;37:e1-9.
14. Fiest KM, Krewulak KD, Sept BG, et al. A study protocol for a randomized controlled trial of family-partnered delirium prevention, detection, and management in critically ill adults: the ACTIVATE study. BMC Health Serv Res 2020;20:453.
15. van Meenen LCC, van Meenen DMP, de Rooij SE, et al. Risk Prediction Models for Postoperative Delirium: A Systematic Review and Meta-Analysis. J Am Geriatr Soc 2014;62:2383-90.
16. Lindroth H, Bratzke L, Purvis S, et al. Systematic review of prediction models for delirium in the older adult inpatient. BMJ Open 2018;8:e019223.
17. Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. Eur J Clin Invest 2015;45:204-14.
18. Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162:W1-73.
19. 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;170:W1-W33.
20. Wolff RF, Moons K, Riley RD, et al. PROBAST:A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. Ann Intern Med 2019;170:51-8.
21. Pisani MA, Murphy TE, Van Ness PH, et al. Characteristics Associated With Delirium in Older Patients in a Medical Intensive Care Unit. Arch Intern Med 2007;167:1629-34.
22. Katznelson R, Djaiani GN, Borger MA, et al. Preoperative Use of Statins Is Associated with Reduced Early Delirium Rates after Cardiac Surgery. Anesthesiology 2009;110:67-73.
23. van den Boogaard M, Pickkers P, Slooter AJC, et al. Development and validation of PRE-DELIRIC (PREdiction of DELIRium in ICu patients) delirium prediction model for intensive care patients: observational multicentre study. BMJ 2012;344:e420.
24. van den Boogaard M, Schoonhoven L, Maseda E, et al. Recalibration of the delirium prediction model for ICU patients (PRE-DELIRIC): a multinational observational study. Intensive Care Med 2014;40:361-9.
25. Wassenaar A, van den Boogaard M, van Achterberg T, et al. Multinational development and validation of an early prediction model for delirium in ICU patients. Intensive Care Med 2015;41:1048-56.
26. Li Y, He T, Wu W. Study for the reliability and validity of Chinese version PRE-DELIRIC delirium prediction model. J Nurs Training 2015;30:1356-9.
27. Yuan J. Development of mathematical model and metabonomic study of blood plasma for ICU patients with delirium[dissertation]. Anhui: Wan Nan Medical College; 2017.
28. Chen Y, Du H, Wei B, et al. Development and validation of risk-stratification delirium prediction model for critically ill patients. Medicine 2017;96:e7543.
29. Zhu X. Analysis and development of a delirium prediction model for intensive care patients. [dissertation]. Chongqing: Third Military Medical University; 2017.
30. Luo X, Li J, Luo M, et al. To Establish and Evaluate a Prediction Model for Delirium. Neural Injury and Functional Reconstruction 2017;12:532-5.
31. Lee A, Mu JL, Joynt GM, et al. Risk prediction models for delirium in the intensive care unit after cardiac surgery: a systematic review and independent external validation. Br J Anaesth 2017;118:391-9.
32. Moon KJ, Jin Y, Jin T, et al. Development and validation of an automated delirium risk assessment system (Auto-DelRAS) implemented in the electronic health record system. Int J Nurs Stud 2018;77:46-53.
33. Deng X, Cao L, Huang Y, et al. Comparison of PRE-DELIRIC and E-PRE-DELIRIC delirium prediction models in ICU patients. Chinese Journal of Practical Nursing 2018;34:1172-6.
34. Linkaitė G, Riauka M, Buneviciute I, et al. Evaluation of PRE-DELIRIC (PREdiction of DELIRium in ICu patients) delirium prediction model for the patients in the intensive care unit. Acta Med Litu 2018;25:14-22.
35. Sosa FA, Roberti J, Franco MT, et al. Assessment of delirium using the PRE-DELIRIC model in an intensive care unit in Argentina. Revista Brasileira de Terapia Intensiva 2018;30:50-6.
36. Marra A, Pandharipande PP, Shotwell MS, et al. Development and Validation of a Daily Prediction Model. Chest 2018;154:293-301.
37. Wassenaar A, Schoonhoven L, Devlin JW, et al. Delirium prediction in the intensive care unit: comparison of two delirium prediction models. Crit Care 2018;22:114.
38. Green C, Bonavia W, Toh C, et al. Prediction of ICU

Delirium. Crit Care Med 2019;47:428-35.
39. Chaiwat O, Chanidnuan M, Pancharoen W, et al. Postoperative delirium in critically ill surgical patients: incidence, risk factors, and predictive scores. BMC Anesthesiol 2019;19:39.
40. Xing H, Lu D, Wang X, et al. The development and application of a risk prediction model for postoperative delirium in ICU patients. Chinese Journal of Nursing 2019;54:8-13.
41. Fan H, Ji M, Huang J, et al. Development and validation of a dynamic delirium prediction rule in patients admitted to the Intensive Care Units (DYNAMIC-ICU): A prospective cohort study. Int J Nurs Stud 2019;93:64-73.
42. Ho MH, Chen K, Montayre J, et al. Diagnostic test accuracy meta-analysis of PRE-DELIRIC (PREdiction of DELIRium in ICu patients):A delirium prediction model in intensive care practice. Intensive Crit Care Nurs 2020;57:102784.
43. Han QY, Hu Z. Clinical prediction model construction and evaluation:a challenge for clinical researchers. Ann Transl Med 2019;8:74.
44. Logullo P, MacCarthy A, Kirtley S, et al. Reporting guideline checklists are not quality evaluation forms: they are guidance for writing. Health Sci Rep 2020;3:e165.
45. van den Boogaard M, Slooter AJC, Brüggemann RJM, et al. Effect of Haloperidol on Survival Among Critically Ill Adults With a High Risk of Delirium. JAMA 2018;319:680-90.
46. Duceppe MA, Williamson DR, Elliott A, et al. Modifiable Risk Factors for Delirium in Critically Ill Trauma Patients: A Multicenter Prospective Study. J Intensive Care Med 2019;34:330-6.

Cite this article as: Chen X, Lao Y, Zhang Y, Qiao L, Zhuang Y. Risk predictive models for delirium in the intensive care unit: a systematic review and meta-analysis. Ann Palliat Med 2021;10(2):1467-1479. doi: 10.21037/apm-20-1183

Supplementary

Appendix 1 Details of the ROB and clinical applicability of included studies

Study ID	Participants		Predictors			Outcome						Analysis								
	1.1	1.2	2.1	2.2	2.3	3.1	3.2	3.3	3.4	3.5	3.6	4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9
Pisani (2007) (15)	Y	Y	Y	PY	Y	Y	Y	Y	Y	PY	Y	N	N	Y	Y	Y	Y	N	Y	NI
Katznelon (2009) (16)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	Y	Y	N	Y	N	Y	NI
Boogaard (2012) (17)	Y	Y	Y	PY	Y	Y	Y	Y	Y	PY	Y	N	N	Y	N	Y	Y	Y	Y	Y
Boogaard (2014) (18)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	Y	N	Y	Y	Y	N	Y
Wassenaar (2015) (19)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Li (2015) (20)	Y	Y	Y	NI	Y	Y	Y	Y	Y	NI	Y	N	Y	Y	Y	—	PY	N	—	—
Yuan (2017) (21)	N	Y	Y	NI	Y	Y	Y	Y	Y	NI	Y	N	N	Y	Y	N	Y	N	N	N
Chen (2017) (22)	Y	Y	Y	PY	Y	Y	Y	Y	Y	PY	Y	N	Y	Y	Y	Y	Y	N	N	Y
Zhu (2017) (23)	Y	Y	Y	PY	Y	Y	Y	Y	Y	PY	Y	N	N	Y	Y	N	Y	N	N	Y
Luo (2017) (24)	N	Y	Y	NI	Y	Y	Y	Y	Y	NI	Y	Y	Y	Y	Y	Y	N	N	N	Y
Lee (2017) (25)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	—	Y	Y	—	—
Moon (2018) (26)	Y	Y	Y	PY	Y	Y	Y	Y	Y	PY	Y	N	Y	Y	Y	Y	Y	N	Y	Y
Deng (2018) (27)	Y	Y	Y	NI	Y	Y	Y	Y	Y	NI	Y	Y	Y	Y	Y	—	N	N	—	—
Linkaitė (2018) (28)	Y	Y	Y	NI	Y	Y	Y	Y	Y	NI	Y	N	Y	Y	Y	—	N	N	—	—
Sosa (2018) (29)	Y	N	Y	NI	Y	Y	Y	Y	Y	NI	Y	Y	Y	Y	Y	—	N	N	—	—
Marra (2018) (30)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	PY	Y	Y	N	Y	Y	Y	Y	NI
Wassenaar (2018) (31)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	—	Y	N	—	—
Green (2019) (32)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	—	Y	Y	—	—
Chaiwat (2019) (33)	Y	Y	Y	PY	Y	Y	Y	Y	Y	PY	Y	N	Y	Y	Y	N	Y	N	Y	N
Xing (2019) (34)	Y	Y	Y	PY	Y	Y	Y	Y	Y	PY	Y	N	Y	Y	Y	N	Y	N	N	Y
Fan (2019) (35)	Y	Y	Y	PY	Y	Y	Y	Y	Y	PY	Y	N	N	Y	Y	N	Y	N	Y	N

ROB, Risk of bias; Y/PY, yes/ probably yes; N/PN, no/probably no; —, no information; NI, not appliable.

1.1 Were appropriate data sources used, e.g., cohort, RCT, or nested case-control study data?
1.2 Were all inclusions and exclusions of participants appropriate?
2.1. Were predictors defined and assessed in a similar way for all participants?
2.2. Were predictor assessments made without knowledge of outcome data?
2.3 Are all predictors available at the time the model is intended to be used?
3.1. Was the outcome determined appropriately?
3.2. Was a prespecified or standard outcome definition used?
3.3. Were predictors excluded from the outcome definition?
3.4. Was the outcome defined and determined in a similar way for all participants?
3.5. Was the outcome determined without knowledge of predictor information?
3.6. Was the time interval between predictor assessment and outcome determination appropriate?
4.1. Were there a reasonable number of participants with the outcome?
4.2. Were continuous and categorical predictors handled appropriately?
4.3. Were all enrolled participants included in the analysis?
4.4. Were participants with missing data handled appropriately?
4.5. Was selection of predictors based on univariable analysis avoided?*
4.6. Were complexities in the data (e.g., censoring, competing risks, sampling of control participants) accounted for appropriately?
4.7. Were relevant model performance measures evaluated appropriately?
4.8. Were model overfitting, underfitting, and optimism in model performance accounted for?*
4.9. Do predictors and their assigned weights in the final model correspond to the results from the reported multivariable analysis?*
"*": Development studies only.
Signaling questions are answered as yes, probably yes, probably no, no, or no information. ROB and concerns for applicability are rated as low, high, or unclear.

http://dx.doi.org/10.21037/apm-20-1183