BMC Geriatrics

SYSTEMATIC REVIEW Open Access

The risk prediction models for cognitive frailty in the older people in China: a systematic review and meta-analysis

Minhua Ren[1], Hongtao Guo[2*], Yingjie Guo[1], Wanjun Guo[1] and Liangjin Zhu[3]

Abstract

Background Recently, many risk prediction models for Cognitive Frailty (CF) in older people in China have been developed. However, there is a shortage of large-scale systematic and comprehensive studies of the methods, quality, and predictors involved in model development.

Aims To systematically assess the risk prediction model of CF in older people in China and to conduct a meta-analysis of its predictors.

Methods PubMed, Cochrane Library, EMbase, Web of Science, CNKI, Wanfang, VIP, and SinoMed were searched from the inception to April 30, 2024. Two researchers independently screened the literature and extracted data. The quality of studies was assessed using the PROBAST tool. Additionally, Stata 18.0 software and MedCalc software were employed to perform a meta-analysis of the modeled predictors and area under the curve (AUC).

Results 17 articles were included, encompassing 22 CF risk prediction models, involving 9,614 participants, of which 2488 (25.9%) were diagnosed with CF. 15 models reported discrimination by AUC (0.710 to 0.991). 8 models conducted internal validation, while 7 models performed external validation. PROBAST evaluation results found that 15 articles (15/17, 88.24%) exhibited a high risk of bias (ROB). The most common predictors were advanced age, irregular exercise, malnutrition, depression, Barthel Index score, female gender, and Instrumental Activities of Daily Living (IADL) score.

Conclusion Due to imprecise modeling methods, incomplete presentation, and lack of external validation, the models' usefulness still needs to be determined. Seven predictive factors are established predictors for CF among older people, including advanced age and so on, but the roles of educational level and fall incidents warrant further investigation.

Keywords Older people, Cognitive frailty, Prediction model, System evaluation/Meta-analysis

*Correspondence:
Hongtao Guo
hongtaoguonm@163.com
[1]School of Nursing, Inner Mongolia Medical University, Hohhot, China
[2]Nursing Department, Affiliated Hospital of Inner Mongolia Medical University, Hohhot, China
[3]Department of Neurosurgery, Affiliated Hospital of Inner Mongolia Medical University, Hohhot, China

Introduction

Population aging represents one of the primary challenges confronting numerous countries globally. According to data from the United Nations, by 2050, the older people aged over 65 in China will reach 365 million [1]. "Cognitive frailty" (CF) was initially proposed by the "International Consensus Group" in 2013 as a clinically heterogeneous manifestation characterized by concurrent physical frailty and cognitive impairment among patients without dementia [2]. As the older people demographic continues to expand, the issue of CF among older adults has become increasingly prominent [3], with research interest in this area growing year by year [4]. Multiple studies have indicated that older individuals experiencing CF are at a heightened risk for adverse outcomes such as falls [5], disabilities [6], hospitalization, dementia [7], and mortality [8, 9]. Early identification of their risk for developing CF and implementing targeted interventions can help prevent progression to dementia and mitigate the occurrence of these negative outcomes. Consequently, an accurate and convenient predictive tool is essential for identifying the risk of CF in older people.

In recent years, most of the studies on the construction and verification of CF risk prediction models for older people were conducted by Chinese scholars, and the sites of model construction were mostly concentrated in the community, inpatient department, and outpatient department (no relevant studies have been found in nursing homes), and the diseases constructed were mainly chronic patients such as hypertension, diabetes, coronary heart disease, and hemodialysis.

There is a study carried out a systematic review of the CF prediction model for older people and found that there were methodological defects, incomplete representation, and lack of validation in the process of model construction, but few articles were included and no correlation analysis was conducted on the predictors and AUC [10]. With the application of model quality evaluation tools, researchers continue to correct flaws in the modeling process to reduce the risk of model bias, but the quality of existing models is still uneven, and the prediction accuracy and clinical applicability of these tools are still uncertain. Therefore, this study aims to conduct a large-scale comprehensive comparative study on the CF prediction model construction process, performance evaluation, and potential data sample bias of older people in China while conducting a meta-analysis on their predictors. To provide valuable references and guidance for clinical practitioners in utilizing relevant risk prediction models effectively to mitigate the occurrence and progression of CF among the elderly population.

Methods

This systematic review was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [11]. A study protocol for this research has been registered in PROSPERO (CRD42024588613).

Research problem construction

Construct evidence-based questions based on the PICOTS model [12] recommended by the Cochrane Methodology Group.

P (Population): ≥60 years old who have been living in China for a long time;

I (Intervention model): Risk prediction models of cognitive frailty;

C (Comparator): none;

O (Outcome): Presence of cognitive frailty.

T (Timing): During hospitalization, outpatient visit or at home;

S (Setting): In a hospital, community or nursing home.

Search strategy

We conducted a comprehensive search across PubMed, Embase, The Cochrane Library, Web of Science, China National Knowledge Infrastructure (CNKI), Wanfang Database, China Science and Technology Journal Database (CQVIP), and Chinese Biomedical Literature Service Database (SinoMed) to identify relevant studies published from inception until May 2024. The search terms included MESH terms and keywords; search strategies were tailored to each database. Additionally, we manually searched citations for potential studies. The search terms roughly include the following three aspects (A complete search strategy is provided in Appendix A):

(1) Aged (Mesh), old, elderly, older adults, senior, elder*;
(2) Cognitive Frailty, Cognitive Fragilities, Cognitive Frailness, Cognitive Frailty Syndrome;
(3) Prediction model, risk prediction, risk assessment, risk scor*, predictive value.

Inclusion and exclusion criteria

The inclusion criteria were as follows

(1) The study population comprised individuals aged ≥ 60 years old in China, (2) The research developed and/or validated a prediction model for CF in older people. (3) The research design was observational, including cross-sectional surveys, cohort studies, and case-control studies, (4) The outcome measure was the occurrence of CF in older people, assessed using CF assessment tools that have been validated and demonstrate clear reliability and validity, (5) Both Chinese or English language publications were considered.

The exclusion criteria were as follows

(1) Repeatedly published literature, (2) Reviews, letters, case studies, conference abstracts, and dissertations, (3) Literature lacking a clear definition of outcome indicators, (4) Unable to obtain the original literature.

Study selection

The literature was imported into Endnote X9 and duplicate articles were deleted. Then, according to the predefined inclusion and exclusion criteria, two reviewers (M.R. and Y.G.) independently screened records based on the title, abstract, and full texts. In case of any disagreements, a third reviewer (W.G.) helped reach a consensus.

Data extraction

Data extraction was performed by three reviewers (M.R., Y.G., W.G.) who cross-checked each other's results and resolved any discrepancies through discussion. A data collection table was developed based on CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS) [13], encompassing the following information extracted from individual articles: study details (first author, publication date, province, study subjects), participant characteristics (age, sample size, incidence of outcome indicators), modeling situation (evaluation tools, modeling methods, modeling quantity, AUC and calibration methods, model validation methods, predictors, and model presentation style).

Assessment of study quality

Two researchers (M.R. and W.G.) utilized the PROBAST [14] evaluation tool to assess both the ROB and the applicability of studies independently, and if there was disagreement, a third researcher (L.Z.) was consulted to reach a consensus.

This tool contains 20 primary questions from four perspectives and evaluates the model's bias risk from four perspectives, including the research object, predictors, results, and data analysis. Similarly, conduct an applicability evaluation from three perspectives following similar steps as in the bias risk analysis. Each question has five options: "yes", "possible yes", "no", "possible no" and "no information", and the evaluation criteria for different perspectives are categorized into three levels: "low", "high", and "unclear". The research is classified high risk if there are at least one perspective is rated as "no" or "probably no"; If there is uncertainty in one or more perspectives, despite other perspectives being rated as low risk, the overall risk of bias is deemed uncertain.; The research is considered low risk only when all perspectives are unanimously judged to be low risk.

Statistical analysis

Meta-analysis was conducted by Stata 18.0 to examine the predictors (Y.G. and M.R.), with odds ratios (*OR*) and 95% confidence intervals (*95%CI*) as effect statistics. Heterogeneity among multiple studies was assessed using the *Q*-test and I^2 statistics. If Cochran's Q statistic results in $P > 0.1$ or $I^2 < 50\%$, which indicates minimal heterogeneity, the fixed-effects model is selected to provide a more precise estimate. Conversely, $P \leq 0.1$ or $I^2 \geq 50\%$ denoted significant heterogeneity, and the random-effects model is chosen to adjust the weight and modify the confidence interval, thereby providing a more robust combined effect. Additionally, sensitivity analysis identified the potential sources of heterogeneity. The model's predictive performance (AUC) was statistically analyzed using MedCalc software. If a study reported only the *95%CI* of the AUC but no standard error (*SE*), the width of the *95%CI* was estimated by dividing the *CI* width by 3.92 [15]. Funnel plots and Egger's test were utilized to assess publication bias. The symmetric distribution of data points in the funnel plot and the $P>0.05$ of the Egger test indicate no significant publication bias [16]. Statistical significance was defined as two-tailed $P < 0.05$.

Results

Literature search results

By conducting a comprehensive search across relevant databases, we identified a total of 718 pertinent studies. Following a meticulous screening process, we ultimately included 17 articles [17–33], encompassing a substantial cohort of 9617 participants. The detailed procedure and outcomes of the literature screening are visually presented in Fig. 1.

Basic characteristics of included literature

The literature included in this study was exclusively published within the past five years in China, including 11 Chinese publications [17–27] and 6 English publications [28–33]. The incidence rate of CF among the elderly population ranged from 10.0 to 65.3%. Detailed characteristics of the included literature are presented in Table 1.

Basic information on risk prediction models

A total of 22 prediction models were constructed in 17 studies [17–33]. Regarding model construction, one study [24] employed different machine learning algorithms to build the risk prediction models and selected the best model by comparing AUC, while the remaining studies solely utilized logistic regression analysis for model establishment. In terms of processing methods of continuous variables, only a few models [17, 19, 23, 30, 31, 33] retained the original continuous variables ($n = 6$, 35.3%). For variable screening techniques, many articles [17, 19–23, 25, 29, 32] conducted multiple-factor analysis

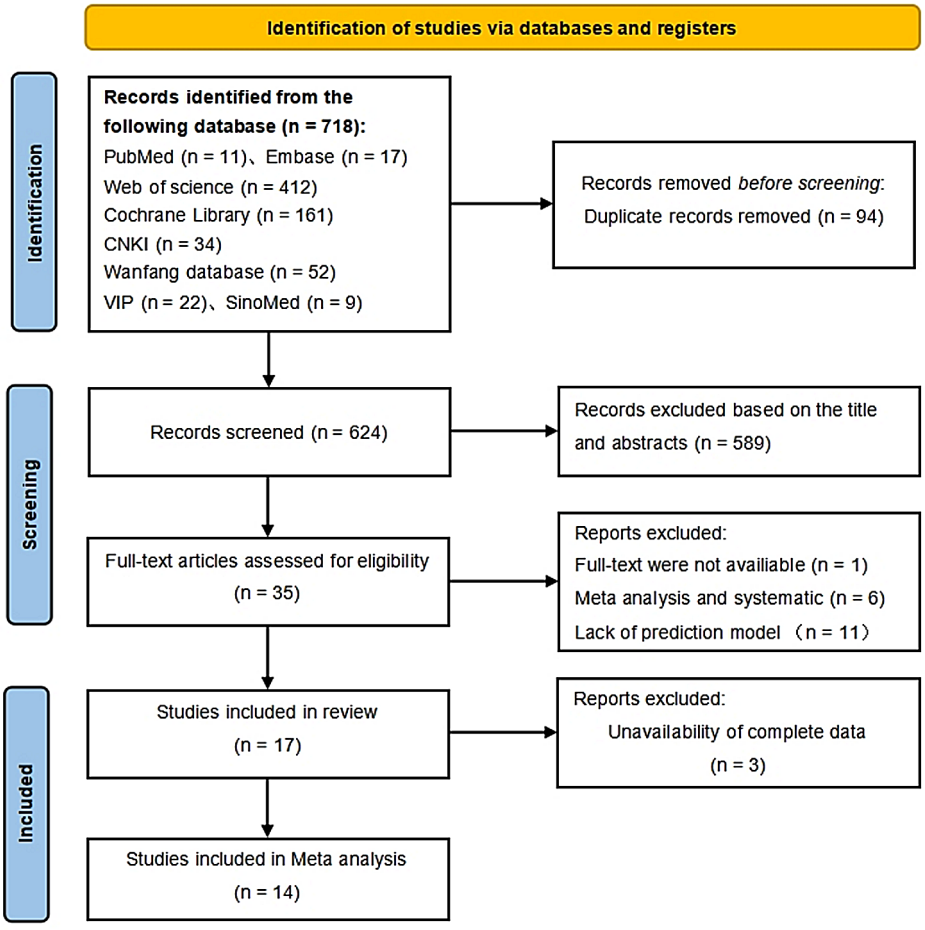

Fig. 1 Flow diagram for literature screening and selection

based on single-factor analysis ($n = 9$, 52.9%). The specific details regarding model establishment are presented in Table 2.

Performance of risk prediction models

The AUC of the models ranged from 0.710 to 0.991, with sensitivities ranging from 70.00 to 95.12% and specificities ranging from 60.0 to 97.8%. 13 references [17–23, 25, 27, 29–31, 33] simultaneously reported discrimination and calibration measures; Regarding model validation, 7 studies [18, 19, 23, 30–33] conducted external validation while 8 studies [20–22, 24, 25, 27–29] performed internal validation using Bootstrap and sample splitting methods. The model was primarily presented as a nomogram ($n = 9$), while alternative approaches include risk-scoring formulas ($n = 4$) and risk-scoring tables ($n = 2$). The prediction model's performance and presentation format are illustrated in Table 3.

Literature quality evaluation

The quality of the literature was comprehensively evaluated by two researchers (M.R. and W.G.) using the PROBAST scale [14], and the evaluation results were meticulously reviewed for accuracy. When assessing ROB, a retrospective research design in 88.23% (15/17) of studies resulted in high ROB within the participant domain. However, low ROB was observed across all predictors and outcomes. In the analysis field, two articles [30, 33] were classified as low-risk while the others were deemed high-risk for bias. The high ROB in the overall evaluation can be attributed to the following reasons: (1) Insufficient sample size in the literature modeling cohort; (2) Research involving conversion of continuous variables into categorical variables; (3) Research-based on single factor analysis for variable screening purposes; (4) Failure to report whether model calibration verification has been conducted; and (5) The absence of information regarding internal validation or a validation sample size below 100 cases. When assessing the application, all components including participants, predictors, and outcomes are deemed to pose low risk. The original study's incorporation of subjects, clinical design, predictor definition, evaluation process, time frame, and outcome definition align consistently with the systematic review

Table 1 Characteristics of included studies

Reference	Region (Province)	Research object		Age	Participants			Outcome	Number of	EPV	Assessment
		Place	Disease		development (n)	validation (n)	No.of Events (n)	incidence (%)	candidate factors		instrument
Chen et al, 2023	Southern(Zhejiang)	Inpatient Department	Orthopedic surgery	-	268	-	175	65.3	8	21.88	①⑤⑥
Chen et al, 2022	Southern(Guangdong)	Community	Physical examination	68(66，72)	368	158	138	36.31	21	6.57	①④⑤
Deng et al, 2023	Southern(Guangdong)	Inpatient Department	Diabetes and hypertension	-	175	76	53	30.3	13	4.08	①④
Wen et al, 2021	Northern (Shanxi)	Outpatient Department	Stable coronary heart disease	(75.36±8.00)	848	-	101	11.9	9	11.22	①④⑤
Yang et al, 2021	Northern (Liaoning)	Community	Chronic disease	(72.65±8.95)	674	-	226	33.5	9	25.11	①④⑤
Wang et al, 2022	Southern (Jiangsu)	Inpatient Department	Hypertension	-	155	66	41	26.2	5	11.6	①⑥
Li et al, 2023	Southern (Jiangsu)	Inpatient Department	Hemodialysis	-	239	40	54	22.59	13	4.15	④⑤⑦
Zhou et al, 2023	Southern (Jiangsu)	Community	Physical examination	71(67，76)	773	332	224	28.05	20	15.5	②⑥
Wang et al, 2023	Southern (Shanghai)	Inpatient Department	Type 2 diabetes	-	262	-	85	32.44	11	7.73	①④⑤
Shen et al, 2023	Northern (Jilin)	Outpatient Department	Chronic disease	-	207	-	75	36.23	8	9.38	②④
Wei et al, 2023	Northern (Beijing)	Inpatient Department	Hypertension	(71.01±7.93)	379	-	145	38.3	13	11.15	①⑥
Bai et al, 2023	Northern (Beijing)	Community	-	(84.96±5.70)	750	-	326	43%	7	46.57	②④
Deng et al, 2023	Southern(Guangdong)	Inpatient Department	Diabetes	71 (65，77)	221	94	61	27.6	13	4.69	①④
Peng et al, 2023	Southern (Jiangsu)	Community	Multi disease patients	(73.35±7.99)	840	342	293	34.3	9	32.56	①④⑤
Huang et al, 2024	Southern (Hunan)	CLHLS (Community)	-	-	2420	3512	243	10	12	20.25	③⑥
Luo et al, 2022	Southern(Guangdong)	Inpatient Department	Chronic kidney disease	67 (63，73)	311	104	93	21.9	29	3.21	①⑥
Tseng et al, 2019	Southern (Taiwan)	Community	-	(73.1±5.4)	724	547	155	15.8	11	14.09	④⑧⑨⑩

①Frailty Phenotype (FP） ②Fatigue, Resistance, Ambulation, Illness and Loss of Weight Index (FRAIL) ③Revised Fried Standard ④Montreal Cognitive Assessment (MoCA) ⑤Clinical Dementia Rating (CDR) ⑥Minimum Mental State Examination (MMSE) ⑦Tiburg Frailty Index (TFI) ⑧6m walking speed ⑨grip ⑩Multiple neurological assessments

inquiries. Consequently, the assessment of the application is regarded as having minimal risk. In summary, while all studies exhibit a high ROB overall, they demonstrate a low risk in terms of predictive model application. A detailed quality assessment is presented in Appendix B.

Meta-analysis results of predictors

Due to the lack of complete data for 3 articles [24, 29, 31], a meta-analysis was conducted using predictors with identical definitions and measurement methods, including more than three articles from the remaining 14 references [17–23, 25–28, 30, 32, 33]. The results revealed that advanced age, irregular exercise, malnutrition, depression, low Barthel index score, female gender, and impaired IADL were identified as independent risk factors for CF in older people (Table 4).

Sensitivity analysis

Among the 9 studies [17, 18, 20–22, 25–27, 33] that included age as a predictor and the 6 studies [17, 18, 21, 22, 25, 27] that included malnutrition, one study appears to be the primary source of statistical heterogeneity [21] due to its skewed participant distribution (Appendix C: Figures 1 and 3). Excluding this study resulted in acceptable heterogeneity across the remaining studies ($I^2 = 48.5\%$, $P = 0.059$; $I^2 = 47.9\%$, $P = 0.104$). Meta-analysis using fixed effect model ($OR = 3.03$, *95%CI*: 2.54 ~ 3.62; $OR = 3.69$, *95%CI*: 2.90 ~ 4.70). Among the 7 articles that included exercise [18, 20, 21, 23, 27, 30], significant overall heterogeneity remained even after excluding three studies [18, 21, 30] with substantial heterogeneity using the one-by-one elimination method. After three studies were excluded, acceptable statistical heterogeneity was achieved (I^2 = 6.4%, $P = 0.361$), and a fixed-effect model was applied for meta-analysis ($OR = 2.40$, 95%*CI*: 1. 86-3. 10) (Appendix C: Fig. 2). Subgroup analysis was conducted on 9 studies that included age as a predictor, which was divided into inpatient, outpatient, and community groups based on study population, and into southern and northern group based on the study region. Results showed no significant difference in model performance between population or regions, and the source of statistical heterogeneity remained unidentified (Appendix C: Figures 4 and 5).

MedCalc analysis results

In the test set of 17 included studies, two did not report the AUC and *95% CI* [20, 33], and three reported the AUC but omitted the *95%CI* [24, 29, 32]. Therefore, we utilized the MedCalc software to analyze the AUC and *95%CI* for the remaining 12 studies [17–19, 21–23, 25–28, 30, 31]. The results indicated significant heterogeneity

Table 2 Model establishment status

References	Modeling method	Variable selection method	Number of models	Missing values		Continuous variable processing method	Number of predictors	Predictors*
				Quantity	Processing			
Chen et al, 2023	logistic	Single factor analysis multiple factor analysis	1	Unreported	-	-	5	①②③④⑤
Chen et al, 2022	logistic	Single factor analysis Binary logistic regression analysis	1	0	-	categorical variable	7	①②⑥⑦⑧⑨⑩
Deng et al, 2023	logistic	Single factor analysis multiple factor analysis	1	1	Delete	-	5	①④⑤⑪⑫
Wen et al, 2021	logistic	Single factor analysis multiple factor analysis	1	12	Delete	categorical variable	6	①⑩⑬⑭⑮⑯
Yang et al, 2021	logistic	Single factor analysis multiple factor analysis	1	Unreported	-	categorical variable	5	①②⑩⑫⑰
Wang et al, 2022	logistic	Single factor analysis multiple factor analysis	1	0	-	categorical variable	4	①②⑱⑲
Li et al, 2023	logistic	Single factor analysis multiple factor analysis	1	0	-	-	8	①④⑩⑫⑳㉑㉒㉓
Zhou et al, 2023	Different machine learning algorithms	Single factor analysis expert opinion	6	81	Delete	categorical variable	13	①③④⑩⑫⑰㉔㉖㉗㉕㉘㉙㉚
Wang et al, 2023	logistic	Single factor analysis multiple factor analysis	1	8	Delete	categorical variable	8	①②⑥⑩⑫⑰㉔㉕㉛
Shen et al, 2023	logistic	Single factor analysis Binary logistic regression analysis	1	0	-	categorical variable	8	⑧⑨⑬㉜㉝
Wei et al, 2023	logistic	Single factor analysis Binary logistic regression analysis	1	0	-	categorical variable	7	①②⑤⑩㉕㉞㉟
Bai et al, 2023	logistic	multiple factor analysis LASSO regression	1	0	-	categorical variable	7	⑥㉕㊱㊲㊳㊴㊵
Deng et al, 2023	logistic	Single factor analysis multiple factor analysis LASSO regression	1	2	Delete	-	6	①⑪⑫㊶㊷㊸
Peng et al, 2023	logistic	Single factor analysis multiple factor analysis	1	18	Delete	categorical variable	9	①④⑩⑪⑰㉕㊹㊺㊻
Huang et al, 2024	logistic	Binary logistic regression analysis	1	0	-	-	6	①⑩㉔㉜㊼㊽
Luo et al, 2022	logistic	Single factor analysis multiple factor analysis	1	5	Delete	categorical variable	5	⑤⑫㉕㊶㊾
Tseng et al, 2019	logistic	Single factor analysis Binary logistic regression analysis	1	0	-	-	6	①⑲㉔㉛㊷㊿

①Age; ②Nutritional; ③Coexistence of multiple illnesses; ④Falls; ⑤Barthel index; ⑥Instrumental activities of daily living(IADL); ⑦Self assessed healt assessment of daytime mental state; ⑨Number of chronic diseases; ⑩Physical exercise; ⑪Intellectual activities; ⑫Depressive symptoms; ⑬Sleep time ev ⑭History of diabetes; ⑮heart-failure; ⑯History of hypertension; ⑰Living alone; ⑱Hypertension classification; ⑲Insomnia; ⑳Long dialysis time; ㉑High combined disease index; ㉒Low serum albumin level; ㉓Low hemoglobin level; ㉔Gender; ㉕Degree of education; ㉖Multiple drug use; ㉗Chronic pain; impairment; ㉙Hearing impairment; ㉚Sleep disorders; ㉛Memory; ㉜BMI; ㉝Job nature; ㉞Drinking history; ㉟Using angiotensin receptor antagonists; heart disease; ㊲Grip; ㊳Step speed; ㊴Subjective cognitive decline (SCD) ; ㊵Five-times-sit-to-stand test(FTSS); ㊶Albumin; ㊷Calf circumference; ㊸D diabetes; ㊹Marital status; ㊺Social affair; ㊻Sleep quality; ㊼Residence Area; ㊽Physical disability; ㊾Social support; ㊿Centripetal obesity.

(I^2=95.62%, P<0.0001). The random-effects model was selected for effect size combination, yielding a combined AUC of 0.886 (*95%CI*: 0.851–0.920), indicating moderate prediction accuracy for CF using the model (Appendix D: Fig. 6). The funnel plot shows asymmetrical distribution of studies (Appendix D: Fig. 7), with the majority outside the *95%CI* of the weighted aggregate AUC. Furthermore, Egger's test confirms significant publication bias (P=0.0006).

Subgroup analysis

For the 12 studies that simultaneously reporting AUC and *95%CI*, a subgroup analysis was conducted based on validation status and method. Five studies were unvalidated [17, 21, 23, 25, 26], five underwent internal validation [18, 22, 27, 28, 31], and three underwent external validation [18, 19, 30] (Chen et al. [18] performed both internal and external validation). The AUC values meeting inclusion criteria were analyzed using MedCalc software (Appendix D: Fig. 8). The results indicated significant heterogeneity in the unvalidated and internally validated groups (I^2>50%), requiring a random-effects

Table 3 Model performance and presentation form

References		Predictive performance (modeling set/validation set)					Presentation form
		Discrimination	Calibration	Sensitivity(%)	Specificity(%)	Accuracy (%)	
Chen et al, 2023		AUC：0.852/-	H-L：0.432/-	87.30/-	71.00/-	-	Formula
Chen et al, 2022		AUC：0.920/- CI：0.910/0.850	Brier：0.117/0.145	79.70/-	89.10/-	85.6/-	Nomogram
Deng et al, 2023		AUC：0.893/0.836	Brier：0.116/0.146	86.80/-	77.90/-	-	Nomogram
Wen et al, 2021		CI：0.835/-	H-L：0.103/-	-	-	-	Nomogram
Yang et al, 2021		AUC：0.970/-	H-L：0.985/-	-	-	-	Nomogram
Wang et al, 2022		AUC：0.883/0.770 CI：0.886/0.781	H-L：0.431/0.352	95.12/82.93	72.81/68.42	-	Nomogram
Li et al, 2023		AUC：0.990/-	H-L：0.966/-	92.60/77.78	97.80/87.10	-/85.00	Formula
Zhou et al, 2023	logistic	AUC：0.824/-	-	79.10/-	70.70/-	-	-
	Naive Bayes	AUC：0.834/-	-	72.10/-	79.70/-	-	-
	Random Forest	AUC：0.889/-	-	83.70/-	83.70/-	-	Risk score
	Extreme gradient	AUC：0.859/-	-	84.90/-	75.20/-	-	-
	K-nearest neighbor	AUC：0.824/-	-	81.30/-	70.00/-	-	-
	Support Vector Machine	AUC：0.831/-	-	80.50/-	77.90/-	-	-
Wang et al, 2023		AUC：-/0.897	H-L：0.437/-				-
Shen et al, 2023		AUC：0.870/-	-	70.70/-	89.30/-	-	Formula
Wei et al, 2023		AUC：0.770/0.737	H-L：0.893/-	60.70/-	83.80/-	-	Formula
Bai et al, 2023		AUC：0.845/- CI：0.805/-	-	-	-	-	Nomogram
Deng et al, 2023		AUC：0.866/ 0.821	H-L：0.305/-	-	-	-	Nomogram
Peng et al, 2023		AUC：0.991/ 0.990	H-L：0.870/0.942	-	-	-	Nomogram
Huang et al, 2024		AUC：0.830/0.840	Calibration curve	90.70/-	84.40/-	-	Nomogram
Luo et al, 2022		AUC：0.913	-	80.65/-	88.61/-	86.36/-	Artificial Neural Network (ANN)
Tseng et al, 2019		CI：0.710/0.690	H-L：0.480/-	70.0/-	60.0/-	63.0/72.0	Risk score

model. The pooled AUC and *95%CI* of the unvalidated group was 0.923 (0.873 to 0.972), whereas those for the internal validation group was 0.818 (0.758 to 0.878). In contrast, no significant heterogeneity was observed in the external validation group (I^2 = 0.00%), allowing a fixed-effects model, and the pooled AUC and its *95%CI* for this group were 0.841 (0.822 to 0.859).

Discussion

The unity of evaluation tools should be considered

Currently, the assessment tools for CF lack standardization, so different studies have utilized varying evaluation tools to diagnose CF, leading to discrepancies in diagnosis. Only five studies included in this analysis employed identical evaluation tools, specifically a combination of the Frailty Phenotype scale (FP), Montreal Cognitive Assessment Scale (MoCA), and Clinical Dementia Rating scale (CDR). Thus, future research efforts should

Table 4 Meta-analysis results of Predictors

Predictors	References (number)	Heterogeneity text		Pooling model	Meta analysis		
		I^2 (%)	P		OR (95%CI)	Z	P
Advanced age	9	89.3	< 0.001	Random Effects Model	4.26 (2.51, 7.24)	5.37	< 0.001
Irregular exercise	7	97.7	< 0.001	Random Effects Model	4.07 (1.12, 14.86)	2.12	0.034
Malnutrition	6	95.3	< 0.001	Random Effects Model	6.76 (2.36, 19.38)	3.56	< 0.001
Depression	4	96.3	< 0.001	Random Effects Model	9.53 (1.80, 50.50)	2.65	0.008
Low level of education	4	94	< 0.001	Random Effects Model	0.86 (0.32, 2.33)	0.29	0.773
Low Barthel index score	3	37.5	0.202	Fixed Effects Model	4.43 (2.91, 6.76)	6.91	< 0.001
Female gender	3	0	0.913	Fixed Effects Model	2.36 (1.92, 2.90)	8.16	< 0.001
impaired IADL	3	93.4	< 0.001	Random Effects Model	5.76 (2.04, 16.26)	3.31	0.001
Falls	3	97.1	< 0.001	Random Effects Model	1.25 (0.10, 15.00)	0.18	0.861

prioritize developing standardized CF-evaluation tools to reduce inconsistencies among studies, which may further improve the accuracy of CF diagnostic and model prediction.

The model's accuracy and reliability need to be further verified

The ROB in the included studies is significant, which may compromise the accuracy and reliability of the findings. Future researchers are advised to thoroughly study the PROBAST predictive model evaluation tool [14] prior to conducting relevant studies. This will enable them to identify the primary sources of model bias and mitigate these issues during the implementation process.

Research objects

Regarding the research objects, 90.9% (20/22) of the model data are derived from retrospective studies, introducing a significant ROB. Therefore, future investigations should prioritize prospective or registered datasets as modeling sources to mitigate biases [14]. Furthermore, participants should be strictly included based on the predefined inclusion criteria, and attention should be paid to maintaining the balance of subjects across different stages to minimize sample bias. Additionally, during data collection, diagnoses should be made using a consistent, pre-specified, and standardized outcome definition, while ensuring an appropriate time interval between the assessment of predictors and the determination of outcomes.

Statistical analysis

In the statistical analysis, the inclusion of 17 studies has revealed several inherent issues, leading to a substantial increase in the risk of bias within the obtained statistical results. First, the number of samples in some studies is insufficient, only 29.4% (5/17) of the studies have an EPV (Event Per Variable, the frequency of events associated with each variable in statistics or data analysis) ≥ 20, and merely 35.3% (6/17) of the studies have a validation sample size ≥ 100. Therefore, future research should establish appropriate inclusion and exclusion criteria and select suitable sample size estimation methods based on the study type to enhance the representativeness and generalizability of the sample [34]; Secondly, the handling of continuous variables was inappropriate. In 64.7% (11/17) of the studies, continuous variables such as age were converted into categorical variables without adhering to standardized definitions, conversion methods, or consistent segmentation points which renders the data unsuitable for secondary analysis. Consequently, future research should aim to preserve the continuity of variables whenever possible. If categorization is necessary, researchers should divide the variables into four or more groups rather than dichotomizing them, thereby minimizing information loss. Finally, the missing values were not adequately addressed, and the 7 studies with missing values were unrepaired. It is crucial to identify whether missing values exist and, if so, to repair them using methods such as multiple interpolation, built-in mechanisms [35], or missing indicator method [36] which can directly handle missing data during the development, validation, or implementation of the prediction model to prevent result bias. In summary, the sample size should be appropriately expanded and properly managed continuous variables and missing values to avoid overfitting the model [37].

Variable screening

In terms of variable screening, the majority of studies have predominantly utilized single-factor analysis without adequately investigating the combined effects of other variables [14], which may result in incorrect predictor selection or variable omission. Therefore, all factors should be directly incorporated into multivariate analysis, or novel variable selection methods, such as LASSO

regression, Ridge regression, and ElasticNet regression, should be adopted. These approaches can effectively mitigate the risk of model overfitting while enhancing screening accuracy.

Modeling approach

In all models, 94.1% (16/17) were developed using conventional logistic regression (LR) analysis methods. Only Zhou et al. [24] utilized six machine learning algorithms to construct multiple CF prediction models for the same sample. The findings demonstrated that the random forest algorithm yielded superior predictive performance in developing the prediction model. However, integrating multiple algorithms also has inherent limitations, such as greater reliance on data, low interpretability, incompatibility, and black-box model [38]. Therefore, future research should endeavor to enhance the predictive capability of models by integrating multiple algorithms that overcome their limitations by using a proper dataset, building a giant model, and explainable artificial intelligence (XAI).

The prediction model's performance is good, but the necessity of model validation

The AUC and H-L tests are the widely used methods for evaluating the performance of predictive models in the included studies. All 22 prediction models exhibited an AUC ≥ 0.7 and a pooled AUC of 0.886, indicating good discrimination and effective identification of high-risk CF among older people. Among these models, 45.4% (10/22) were subjected to the H-L test, with all resulting $P \geq 0.05$, suggesting a favorable agreement between predicted outcomes and actual observations. However, only 18.1% (4/22) of the models were validated using both internal and external verification methods. The limited comprehensiveness of their reported results constrained the completeness of the secondary study. The AUC subgroup analysis results demonstrate that model validation not only effectively reduces heterogeneity within the study but also enhances the model's generalization ability and practical applicability. So it is important to employ methods such as bootstrapping and cross-validation for internal validation during the development of the prediction model, unless the sample size and EPV are sufficiently large [14]. These techniques are crucial for quantifying the degree of overfitting in the developed model and assessing the optimism in prediction performance. Additionally, adjusting or shrinking the model's prediction performance estimates based on internal validation results can avoid prediction bias when the model is applied by others. Therefore, appropriate indicators should be employed to comprehensively evaluate the predictive performance of the model upon its completion. Additionally, rigorous verification processes must be conducted to ensure the accuracy and reliability of the model, which are essential for developing a high-quality and scalable model.

Predictors are objectively quantifiable and readily measurable

The meta-analysis revealed that the predictors of CF in older adults encompassed four domains: demographic characteristics (advanced age, female), behavior and lifestyle factors (irregular exercise), physiological function factors (malnutrition, impairment in IADL, low Barthel index score), and psychosocial factors (depression).

Four studies [25, 27, 28, 32] have identified the level of education as a predictor. Among these, three studies [25, 27, 28] indicate that higher levels of education are associated with a decreased risk of CF. Conversely, Luo et al. [32] report a positive correlation between education level and CF, resulting in non-significant differences findings in the meta-analysis, potentially attributable to an imbalance in the distribution of educational levels among older adults included in the study, as the majority were primary school graduates or below. Therefore, future research should focus on ensuring the balance of predictors when developing models and further investigate the relationship between educational level and CF in older populations.

Falls have been investigated as a predictor in three studies [17, 19, 23]. However, Chen's study [17] reports a lower incidence of falls in the CF group (20.6%) compared to the non-cognitive frailty group (62.4%), indicating that falls may not serve as an independent predictor of CF in this specific context. This could be attributed to the population consisting of older patients following orthopedic surgery, where the fall incidence is higher than in the general older population and directly linked to the underlying disease. Therefore, future researchers should avoid including such factors that are directly related to disease.

The findings indicate that depression is a significant predictor of CF in older people (OR = 9.53). This association may be attributed to the detrimental impact of depression on activities of daily living, social engagement avoidance, and intentional reduction in activities, which subsequently impairs advanced cognitive functions such as memory, reasoning, and spatial ability [39]. Therefore, it can be inferred that depression plays a crucial role in the development of CF among older adults. Future research should prioritize investigating these predictors due to their objectivity or availability of corresponding evaluation scales, thereby facilitating this model's wider adoption and application.

The considerations and prospects for the model update mechanism

With the passage of time and the advancement of research, new predictive factors or assessment tools may emerge, so it is essential to continuously monitor the dynamic changes in the data and the model's performance, and update them timely manner to maintain prediction accuracy and reliability in the practical application of the model. The model can be updated through recalibration, refitting, and dynamic updating to introduce new variables and features. The specific methods include the logistic regression model, Bayesian model updating, and other machine learning algorithms. Several studies have focused on refining existing models; for instance, Hartmann et al. [40] used three methods to update the schizophrenia prediction model and concluded that the dynamic updating method may demonstrate the best predictive performance consistently across the validation period. It is important to highlight that model updating still necessitates the calculation of sample size. The required sample size for this dataset depends on the specific approach used to update the model, as well as whether additional predictors have been incorporated [34]. Furthermore, the updated model must undergo performance evaluation to ensure its accuracy and reliability. A comparison of the performance between the original and updated models is also essential to validate the effectiveness of the update.

Strengths and limitations

Our study offers a systematic evaluation of existing models for predicting the risk of CF in older people in China, along with a comprehensive, objective, and quantitative combing and integration of numerous original studies. It clarifies the performance and existing issues of these models, which provides method guidance for the subsequent construction or updating of the model, and helps to promote the application of this type of risk prediction model in clinical practice. Furthermore, the quantitative synthesis of predictors identified in existing studies can inform the design of effective interventions for CF in older people. Nevertheless, it is imperative to acknowledge the limitations of this study. In addition to the deficiencies identified in the model development process discussed in the article, several inherent constraints warrant attention. Firstly, the literature search for this study was restricted to English and Chinese, with a focus on older people in China, which may have introduced publication bias. Secondly, three studies were excluded from the meta-analysis due to incomplete data reporting, potentially leading to bias. Thirdly, lack of uniformity in evaluation methods, including assessment of CF and predictors. The diverse CF assessment methods used in the included studies led to diagnostic inconsistencies. Meanwhile, some predictors lack consistent evaluation methods, which hinders the possibility of conducting a comprehensive meta-analysis. For instance, sleep-related predictors are assessed using diverse parameters such as insomnia status, sleep duration, and sleep quality. Finally, the meta-analysis revealed a substantial degree of heterogeneity, potentially attributable to variations in study design, participant populations, and assessment tools. Consequently, future researchers are encouraged to unify evaluation methods for various indicators, comprehensively report research data, and actively develop standardized measurement tools for CF. When conducting meta-analyses, it is essential to perform exhaustive searches across literature in multiple languages and databases while defining a more homogeneous study population to minimize heterogeneity and enhance the reliability of findings.

Conclusion

In summary, our systematic review included 17 original studies, with a pooled AUC of 0.886, suggesting that although the current CF prediction model for older adults demonstrates satisfactory performance, its development process lacks sufficient rigor. Additionally, the absence of large-scale, multi-center validation raises concerns about its clinical feasibility and generalizability. Advanced age, female gender, irregular exercise, malnutrition, diminished self-care ability, and depression have been identified as consistent predictors of CF in older adults. However, research findings regarding falls and educational level remain inconsistent. Future studies should strictly follow the PROAST model criteria and validate it through large-scale, multi-center trials to enhance its applicability and adaptability. This will help accurately identify CF risk in older people. Investigating predictors of CF risk is crucial for developing effective interventions. Existing research shows that improving nutrition, promoting physical activity, enhancing mental health, strengthening self-care, and optimizing social support can delay CF onset and progression. Further research is needed to confirm these findings.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12877-025-05961-2.

Supplementary Material 1

Author contributions

M.R, W.G, and L.Z designed the research. M.R, Y.G, and W.G performed the study selection and data extraction. The study quality assessment was completed by M.R and W.G. The meta-analysis was performed by Y.G and M.R. Additionally, M.R and Y.G drafted the manuscript. Finally, M.R, Y.G, W.G, L.Z, and H.G were revised and edited. All authors approved the manuscript and definitive version of the article.

Funding
Funding for open access publishing: Inner Mongolia Medical University. No financial assistance from governmental, corporate, or non-profit entities was used to conduct the study described in the manuscript nor to assist in its preparation.

Data availability
No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
All authors approved the final manuscript and the submission to this journal.

Competing interests
The authors declare no competing interests.

Received: 10 December 2024 / Accepted: 17 April 2025
Published online: 22 May 2025

References

1. Nations U. World population prospects 2022[M]. United Nations; 2022.
2. Kelaiditi E, Cesari M, Canevelli M, et al. Cognitive frailty: rational and definition from an (I.A.N.A./I.A.G.G.) international consensus group[J]. J Nutr Health Aging. 2013;17(9):726–34. https://doi.org/10.1007/s12603-013-0367-2.
3. Qiu Y, Li G, Wang X, et al. Prevalence of cognitive frailty among community-dwelling older adults: A systematic review and meta-analysis[J]. Int J Nurs Stud. 2022;125:104112. https://doi.org/10.1016/j.ijnurstu.2021.104112.
4. Zhu M, Li Z. Analysis of research hotspots and development trends of cognitive frailty. J Nurses Train. 2022;37(22):2102–6. https://doi.org/10.16821/j.cnki.hsjx.2022.22.017.
5. Guo X, Pei J, Ma Y, et al. Cognitive frailty as a predictor of future falls in older adults: A systematic review and Meta-Analysis[J]. J Am Med Dir Assoc. 2023;24(1):38–47. https://doi.org/10.1016/j.jamda.2022.10.011.
6. Choi K, Ko Y. Cross sectional association between cognitive frailty and disability among community-dwelling older adults: focus on the role of social factors[J]. Front Public Health. 2023;11:1048103. https://doi.org/10.3389/fpubh.2023.1048103.
7. Ward DD, Wallace LMK, Rockwood K. Frailty and risk of dementia in mild cognitive impairment Subtypes[J]. Ann Neurol. 2021;89(6):1221–5. https://doi.org/10.1002/ana.26064.
8. Qiu Y, Li G, Zheng L, et al. Relationship between cognitive frailty and mortality in older adults: A systematic review and Meta-Analysis[J]. J Am Med Dir Assoc. 2023;24(11):1637–e16448. https://doi.org/10.1016/j.jamda.2023.08.001.
9. Zhang XM, Wu XJ, Cao J, et al. Association between cognitive frailty and adverse outcomes among older adults: A Meta-Analysis[J]. J Nutr Health Aging. 2022;26(9):817–25. https://doi.org/10.1007/s12603-022-1833-5.
10. Huang J, Zeng X, Hu M, et al. Prediction model for cognitive frailty in older adults: A systematic review and critical appraisal[J]. Front Aging Neurosci. 2023;15:1119194. https://doi.org/10.3389/fnagi.2023.1119194.
11. Page MJ, McKenzie JE, Bossuyt PM, The PRISMA 2020 statement: an updated guideline for reporting systematic reviews[J]. BMJ (, Clinical et al. research ed.), 2021, 372: n71. https://doi.org/10.1136/bmj.n71
12. Moons KG, Hooft L, Williams K, et al. Implementing systematic reviews of prognosis studies in Cochrane[J]. Cochrane Database Syst Rev. 2018;10(10):ED000129. https://doi.org/10.1002/14651858.ED000129.
13. Moons KGM, de Groot JAH, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist[J]. PLoS Med. 2014;11(10):e1001744. https://doi.org/10.1371/journal.pmed.1001744
14. 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[J]. Ann Intern Med. 2019;170(1):W1–33. https://doi.org/10.7326/M18-1377.
15. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve[J]. Radiology. 1982;143(1):29–36. https://doi.org/10.1148/radiology.143.1.7063747.
16. Egger M, Davey Smith G, Schneider M, et al. Bias in meta-analysis detected by a simple, graphical test[J]. BMJ. 1997;315(7109):629–34. https://doi.org/10.1136/bmj.315.7109.629.
17. Chen C, Wu D, Sun B, et al. Construction of a risk prediction model for postoperative cognitive weakness in older adult patients subjected to orthopedic surgery. Chin J Prim Med Pharm. 2023;30(2):287–91. https://doi.org/10.3760/cma.j.cn341190-20220802-00627.
18. Chen Y, Zhang Z, Zuo Q, et al. Construction and validation of a prediction model for the risk of cognitive frailty among the elderly in a community. Chin J Nurs. 2022;57(2):197–203. https://doi.org/10.3761/j.issn.0254-1769.2022.02.012.
19. Deng Y, Li N, Wang Y, et al. Influencing factors and construction of risk prediction nomogram model of cognitive frailty in elderly inpatients with diabetes mellitus and hypertension. Practical J Cardiac Cereb Pneumal Vascular Disease. 2023;31(12):60–5. https://doi.org/10.12114/j.issn.1008-5971.2023.00.311.
20. Wen F, Cheng M, Zhao C, et al. Development of a cognitive frailty prediction model for elderly patients with stable coronary artery disease. J Nurs Sci. 2021;36(10):21–6. https://doi.org/10.3870/j.issn.1001-4152.2021.10.021.
21. Yang Z, Zhang H. A nomogram for predicting the risk of cognitive frailty in community-dwelling elderly people with chronic diseases. J Nurs Sci. 2021;36(12):86–9. https://doi.org/10.3870/j.issn.1001-4152.2021.12.086.
22. Wang Y, Liu Y. Influencing factors of cognitive frailty in hospitalized elderly hypertensive patients and construction of nomogram model. Practical J Cardiac Cereb Pneumal Vascular Disease. 2022;30(7):54–9. https://doi.org/10.12114/j.issn.1008-5971.2022.00.167.
23. Li M. Establishment of a risk prediction model for cognitive frailty in elderly patients undergoing maintenance Hemodialysis. Chin Gen Pract Nurs. 2023;21(10):1392–6. https://doi.org/10.12104/j.issn.1674-4748.2023.10.025.
24. Zhou C, Jin X, Guo Z, et al. Comparison of cognitive frailty risk prediction models for community older adults based on machine learning algorithms. J Nurs Sci. 2023;38(19):1–5. https://doi.org/10.3870/j.issn.1001-4152.2023.19.001.
25. Wang X, Xu Y. Prediction of cognitive decline among elderly patients with type 2 diabetes mellitus. CHINA Prev Med J. 2023;35(12):1037–42. https://doi.org/10.19485/j.cnki.issn2096-5087.2023.12.006.
26. Shen T, Wang Y, Jin W, et al. Risk factors analysis and risk model construction of cognitive frailty in elderly patients with chronic diseases. J Jilin University: Med Ed. 2023;49(5):1304–9. https://doi.org/10.13481/j.1671-587X.20230525.
27. Wei M, Li M, Xu L, et al. Construction and verification of cognitive frailty risk prediction model in elderly hospitalized hypertensive patients. Chin J Mod Nurs. 2023;29(36):4952–8. https://doi.org/10.3760/cma.j.cn115682-20230731-00286.
28. Bai A, Zhao M, Zhang T, et al. Development and validation of a nomogram-assisted tool to predict potentially reversible cognitive frailty in Chinese community-living older adults[J]. Aging Clin Exp Res. 2023;35(10):2145–55. https://doi.org/10.1007/s40520-023-02494-9.
29. Peng S, Zhou J, Xiong S, et al. Construction and validation of cognitive frailty risk prediction model for elderly patients with Multimorbidity in Chinese community based on non-traditional factors[J]. BMC Psychiatry. 2023;23:266. https://doi.org/10.1186/s12888-023-04736-6.
30. Huang J, Zeng X, Ning H, et al. Development and validation of prediction model for older adults with cognitive frailty[J]. Aging Clin Exp Res. 2024;36(1):8. https://doi.org/10.1007/s40520-023-02647-w.
31. Deng Y, Li N, Wang Y et al. Risk Factors and Prediction Nomogram of Cognitive Frailty with Diabetes in the Elderly[J]. Diabetes, Metabolic Syndrome and Obesity, 2023, 16: 3175–3185. https://doi.org/10.2147/DMSOS426315
32. Luo B, Luo Z, Zhang X, et al. Status of cognitive frailty in elderly patients with chronic kidney disease and construction of a risk prediction model: a cross-sectional study[J]. BMJ Open. 2022;12(12):e060633. https://doi.org/10.1136/bmjopen-2021-060633.
33. Tseng SH, Liu LK, Peng LN, et al. Development and validation of a tool to screen for cognitive frailty among Community-Dwelling Elders[J]. J Nutr Health Aging. 2019;23(9):904–9. https://doi.org/10.1007/s12603-019-1235-5.
34. Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model[J]. BMJ. 2020;368:m441. https://doi.org/10.1136/bmj.m441.
35. Nijman S, Leeuwenberg AM, Beekers I, et al. Missing data is poorly handled and reported in prediction model studies using machine learning: a literature

review[J]. J Clin Epidemiol. 2022;142:218–29. https://doi.org/10.1016/j.jclinepi.2021.11.023.
36. Ehrig M, Bullock GS, Leng XI, et al. Imputation and missing indicators for handling missing longitudinal data: data simulation analysis based on electronic health record data[J]. JMIR Med Inf. 2025;13:e64354–64354. https://doi.org/10.2196/64354
37. Zhang Q, Yuan KH, Wang L. Asymptotic bias of normal-distribution-based maximum likelihood estimates of moderation effects with data missing at random[J]. Br J Math Stat Psychol. 2019;72(2):334–54. https://doi.org/10.1111/bmsp.12151.
38. Lim SJ, Son M, Ki SJ, et al. Opportunities and challenges of machine learning in bioprocesses: categorization from different perspectives and future direction[J]. Bioresour Technol. 2023;370:128518. https://doi.org/10.1016/j.biortech.2022.128518.
39. Yuan M, Du J, Wang WC, et al. The mediating effect of depression on the ability of daily living and cognitive function in the elderly[J]. Mod Prev Med. 2022;49(24):4500–4. https://doi.org/10.20043/j.cnki.MPM.202206101.
40. Hartmann S, Dwyer D, Scott I, et al. Dynamic updating of psychosis prediction models in individuals at ultra high-risk of psychosis[J]. Biol Psychiatry: Cogn Neurosci Neuroimaging. 2025. https://doi.org/10.1016/j.bpsc.2025.03.006. S2451-9022(25)00119-3.

Publisher's note