

Methodological quality and clinical applicability of obesity-related disease risk prediction models: a systematic review and meta-analysis

Limin Wang

Fudan University

Mengfan Wang

Fudan University

Zhuo Chen

Fudan University

Chuanlin Cao

Fudan University

Yuxiang Ma

Fudan University

Mengyun Sui

smg19916542392@163.com

Shanghai Municipal Center For Disease Control Prevention

Xiaohua Ying

Fudan University

Article

Keywords: Obesity, Overweight, Prediction model, Risk prediction, Chronic disease, External validation, Machine learning, Risk of bias

Posted Date: October 9th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10892305/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background and Objective

Obesity is a major global risk factor for multiple chronic diseases. Prediction models have increasingly been developed to support early identification and prevention; however, evidence regarding their performance, methodological robustness, and applicability across different obesity-related diseases remains fragmented. We conducted a systematic review to synthesise prediction models for obesity-related diseases, focusing on model performance, predictors, validation characteristics, and risk of bias.

Methods

We searched PubMed, Web of Science, Embase, and CNKI from Jan 1, 2020 to Feb 1, 2026. Models developed to estimate the current or future risk of obesity-related diseases were eligible, including cross-sectional risk classification models and longitudinal incident-risk models. Models predicting prognosis, treatment response, or post-diagnosis outcomes were excluded. Two reviewers independently screened studies and extracted data. Model characteristics, performance (discrimination, calibration, and clinical utility), and validation methods were summarised. Risk of bias was assessed using the PROBAST and PROBAST+AI.

Results

We included 24 studies comprising 43 prediction models across multiple outcomes, including metabolic, cardiovascular, and sleep-related diseases. Most models showed favourable discrimination, with area under the curve values ranging from 0.715 to 0.961. However, methodological limitations were common. Half of the models were at high risk of bias, mainly due to deficiencies in the analysis domain. External validation was rare, with only 4.7% of models validated in independent populations. Predictors were predominantly routine clinical variables, such as age, body mass index, and metabolic biomarkers.

Conclusions

Although existing models demonstrate promising discriminatory performance, their methodological limitations and lack of external validation may limit reliability and generalisability. Future research should prioritise methodological rigour and external validation to improve clinical applicability.

Highlight

- First cross-disease systematic review of obesity-related disease prediction models;
- 51% of models were at high risk of bias, driven by analysis domain deficiencies;
- Disease-specific pooled AUCs ranged from 0.766 to 0.885 across seven outcomes;
- External validation was performed in only 4.7% of models included;
- Traditional logistic regression remained dominant; ML showed no consistent advantage.

1 Introduction

High body mass index (BMI) is a major risk factor for disease worldwide.(1, 2)The prevalence of overweight and obesity has risen substantially over the past decades. Nearly 2 billion adults are now classified as overweight, with a growing proportion affected by obesity.(3) This upward trend has been consistently observed across regions and age groups, including children, adolescents, and adults, from 1975 to recent years.(2, 4–6) High BMI contributes substantially to the global burden of disease. It is associated with increased mortality and disability-adjusted life years (DALYs) across a wide range of conditions.(1–3, 5, 7, 8) In the Global Burden of Disease Study, high BMI is recognised as a leading risk factor for multiple chronic diseases. It also accounts for a considerable proportion of cancer burden, with an estimated 481,000 cases attributable to high BMI in 2012(8), and a continuing rise in related deaths and DALYs in recent years.(9)

Obesity is associated with a wide range of chronic conditions, including type 2 diabetes, hypertension, cardiovascular disease, non-alcoholic fatty liver disease, and obstructive sleep apnoea.(10–16) These conditions often cluster and contribute to metabolic syndrome. Multimorbidity is common among individuals with obesity and substantially increases disease burden.(17) The clinical consequences are considerable. Higher BMI is associated with increased morbidity and adverse health outcomes across populations.(10) Obesity is also linked to mental health disorders, although these associations are complex.(12, 16) Together, these conditions place a substantial burden on healthcare systems. The economic impact is also substantial. Obesity-related conditions account for a large proportion of healthcare expenditure and resource use.(18, 19)Delayed identification and management can lead to disease progression and higher costs. Early identification of individuals at high risk is therefore essential.

Risk prediction models are increasingly used in chronic disease prevention. They enable risk stratification by identifying individuals at high risk of disease. This supports early detection and targeted intervention and may improve the efficiency of healthcare resource allocation.(20–24)These models typically integrate multiple predictors to estimate individual risk. They are applied across a range of settings, including primary care and population health management. Recent advances in data availability and analytical methods, including artificial intelligence approaches, have further accelerated their development and use.

However, important methodological limitations are common in prediction model studies. These limitations may compromise model performance, reliability, and generalisability. Frequent concerns include small sample sizes, inappropriate predictor selection, overfitting, and a lack of internal and external validation.(25–29) The risk of bias in these studies is often high. The PROBAST provides a structured framework for assessing methodological quality, and has identified substantial bias across multiple domains in prediction model research.(27, 30–32) These issues are particularly evident in models developed using complex analytical approaches, including machine learning, where inadequate validation and poor study design are common.(27) As a result, many

models show limited transferability and may not perform well in external populations. Methodological shortcomings, including inadequate sample size and suboptimal predictor and outcome definitions, further restrict their clinical applicability.(28, 33)

Several systematic reviews and meta-analyses have evaluated prediction models for specific obesity-related conditions. These include models for childhood obesity and overweight(34), diabetes remission after bariatric surgery(35, 36), cardiovascular and renal outcomes in individuals with type 2 diabetes(37), and disease prediction using artificial intelligence or machine learning approaches(38–41). Other reviews have broadly explored applications of artificial intelligence in obesity research or clinical prediction across selected conditions.(42, 43)However, the current evidence remains fragmented. Most existing reviews focus on single diseases or specific populations, limiting comparability across conditions. In addition, methodological quality assessment is inconsistently reported. Although some studies have applied tools such as the PROBAST to assess risk of bias(34), a comprehensive evaluation of methodological quality across a broad range of obesity-related disease prediction models is lacking. As a result, the overall reliability and applicability of these models remain unclear.

To address these gaps, we conducted a systematic review of risk prediction models for overweight- and obesity-related diseases. We summarised model characteristics and performance, including discrimination (e.g., area under the curve and C statistic, sensitivity, and specificity), calibration (e.g., calibration plots and Hosmer–Lemeshow test), and potential clinical utility (e.g., decision curve analysis). We also examined internal validation methods (e.g., bootstrapping, split-sample validation, and k-fold cross-validation) and external validation performance. In addition, we assessed methodological quality using the PROBAST. We aimed to provide a comprehensive synthesis to inform future model development and facilitate clinical translation.

2 Methods

2.1 Protocol and registration

This systematic review was registered on the International Prospective Register of Systematic Reviews (PROSPERO: CRD420261289018) and was reported based on the PRISMA (Preferred Reporting Items for Systematic Review and Meta-analysis) extension statement for systematic review checklist (Table S1) .

2.2 Eligibility criteria

Inclusion criteria were as follows:(1) studies reporting the development or validation of multivariable risk prediction models for obesity-related diseases, including dysglycaemia, dyslipidaemia, hypertension, non-alcoholic fatty liver disease, obstructive sleep apnoea syndrome, reproductive disorders, cardiovascular diseases, cancers, and mental health disorders(44–46); (2) models aimed at predicting disease onset (incidence), rather than prognosis or diagnosis; (3) models including overweight or obesity (e.g., body mass index, waist circumference) as a predictor.

Exclusion criteria were:(1) studies not developing or validating a multivariable prediction model (e.g., association studies); (2) models predicting prognosis, treatment response, or diagnostic outcomes; (3) reviews, editorials, guidelines, animal studies, or retracted articles; (4) studies with unavailable full text or insufficient data for extraction.

2.3 Search strategy

We systematically searched PubMed, Web of Science, EMBASE, and CNKI from Jan 1, 2020, to Feb 1, 2026. Search strategies combined three key domains: (1) overweight and obesity, (2) obesity-related diseases, and (3) risk prediction models. Controlled vocabulary terms (e.g., MeSH and Emtree) and free-text terms were used in combination with Boolean operators. Search terms included “obesity”, “overweight”, “BMI”, and “waist circumference” for the population domain; a broad range of metabolic, cardiovascular, oncological, and mental health conditions for the outcome domain; and terms such as “risk prediction”, “risk assessment”, “risk evaluation”, “prediction model”, “risk score”, “nomogram”, along with “validation”, “discrimination”, “calibration”, and “AUC” for the model domain. Search strategies were adapted for each database. We also traced the references of included studies and relevant reviews to identify any potentially eligible studies. The full search strategies and results are provided in the Table S2-S5.

2.4 Study selection

EndNote 21 literature management software was used to manage the literature search records. According to the selection criteria, two reviewers independently screened titles and abstracts of all identified articles after excluding duplicates. Any studies with the potential to be included in the study, and any overlapping studies were subjected to full text evaluation. Any disagreements between the two reviewers were resolved by discussion. In cases of persistent disagreement, a final decision was made by a third author.

2.5 Data extraction

Data extraction was guided by the PROBAST and PROBAST + AI frameworks using a standardised form. Extracted information included: (1) study characteristics: author, year, country, study population, design, data source, sample size, number of outcome events, follow-up duration, and number of models; (2) model characteristics: handling of continuous variables, missing data methods, predictor selection methods, number and type of predictors, modelling methods, model presentation, and performance evaluation approaches (apparent performance, internal validation, external validation); (3) model performance: discrimination (e.g., area under the curve [AUC], C statistic, sensitivity, specificity), calibration (e.g., calibration plots, Hosmer–Lemeshow test), potential clinical utility (e.g., decision curve analysis), and results of internal and external validation.

2.6 Risk of bias and applicability assessment

Two reviewers (LM Wang and MF Wang) independently assessed risk of bias and applicability, with disagreements resolved by consensus or a third reviewer. For models developed using conventional statistical methods (e.g., logistic regression, survival analysis), risk of bias was assessed using the PROBAST. For

models developed using machine learning or artificial intelligence approaches (e.g., support vector machines, gradient boosting), both the PROBAST and PROBAST + AI tool was applied.

PROBAST evaluates four domains (participants, predictors, outcomes, and analysis) using 20 signalling questions⁽³⁰⁾. In this review, the PROBAST tool was applied to assess the risk of bias and applicability of the included prediction models using a standardized framework. The risk of bias for each domain was classified as low or high according to the responses to the signalling questions. A domain was considered at high risk of bias if one or more signalling questions indicated potential methodological concerns. Overall risk of bias was classified as high when at least one of the four domains was rated as high risk, whereas models with low risk of bias across all four domains were considered to have a low overall risk of bias.

PROBAST + AI further expands the original framework to identify methodological limitations of artificial intelligence-based models, enabling independent assessment of model development and performance evaluation³¹. PROBAST was applied to all included prediction models, whereas PROBAST + AI was additionally applied to models developed using artificial intelligence or machine learning approaches. PROBAST + AI was used to identify potential methodological limitations related to AI-based model development and validation, enabling a more comprehensive assessment of bias in AI prediction models.

2.7 Statistical analysis

Descriptive analyses were conducted to summarise the characteristics of included studies and prediction models. Study characteristics, target populations, data sources, modelling approaches, predictor selection methods, validation strategies, and reported performance metrics were summarised using frequencies and percentages.

When multiple prediction models were reported within a single study, each model was treated as an independent unit of analysis because the models differed in modelling algorithms, predictor combinations, prediction horizons, outcome definitions, or target populations. Given the objective of comparing individual model performance, separate analyses were conducted for each eligible model.

To quantitatively synthesise model performance, subgroup meta-analyses were performed for discrimination measures. The area under the receiver operating characteristic curve (AUC) was selected as the primary outcome because it was the most consistently reported performance metric across studies and is less dependent on classification thresholds than sensitivity or specificity. Pooling was performed only for outcomes with ≥ 2 models reporting full AUC and CI data; remaining outcomes were described narratively. Forest plots were generated to present pooled estimates, 95% confidence intervals, and prediction intervals. Sensitivity, specificity, and accuracy were not pooled because most studies did not provide sufficient statistical information, such as 95% confidence intervals or standard errors, required for meta-analysis. Random-effects models were applied throughout to account for anticipated clinical and methodological heterogeneity arising from differences in populations, predictor definitions, and modelling strategies.

Statistical heterogeneity was evaluated using the I^2 statistic and Cochran's Q test. An I^2 value greater than 50% was considered indicative of substantial heterogeneity. Prediction intervals were additionally reported to reflect the expected range of model performance across future settings and populations.

Meta-regression analyses and formal assessments of publication bias were not performed because several outcome categories included a limited number of studies and considerable methodological heterogeneity. All statistical analyses were conducted using R software (version 4.4.2). A two-sided p value < 0.05 was considered statistically significant.

3 Results

3.1 Study selection

A total of 6653 records were identified through database searches (PubMed, Web of Science, EMBASE, and CNKI). After removal of duplicates, 4,840 records were screened by title and abstract, of which 4,779 were excluded as irrelevant. Sixty-one reports were sought for full-text retrieval, and 51 were assessed for eligibility. 27 reports were excluded for ineligible model objective or predictors ($n = 11$), no internal validation or no reporting of validation results ($n = 10$), or unavailable full text or insufficient key information ($n = 6$). Ultimately, 24 studies comprising 43 prediction models were included in the systematic review. The study selection process is presented in Figure S1.

3.2 Study characteristics

The 24 included studies were published between 2020 and 2025, predominantly conducted in China ($n = 17$, 70.8%), with the remainder from Japan ($n = 2$), the United States ($n = 2$), South Korea ($n = 1$), and Spain ($n = 1$); one study did not report the country. Participants ranged from 14 to 86 years, with most studies ($n = 21$) focusing on adults (≥ 18 years). Mixed-sex populations were included in 22 studies (91.7%). Cross-sectional designs ($n = 14$, 58.3%) slightly outnumbered cohort designs ($n = 10$, 41.7%); follow-up in cohort studies ranged from 1 to 14 years (mostly 1–4 years). Data sources included hospital-based populations ($n = 11$, 45.8%), public datasets such as NHANES, CHARLS, and NAGALA ($n = 6$, 25%), health examination databases ($n = 4$, 16.7%), population-based cohorts ($n = 2$, 8.3%), and weight management clinics ($n = 1$, 4.2%). Sample sizes ranged from 158 to 82,938 participants, with most studies ($n = 22$, 91.7%) enrolling fewer than 10,000. (Table 1)

Table 1
Basic characteristics of the included studies

Study Design/Outcome	First Author, Year	Country/Region	Population			Sample Source	Sample Size	No. of Events	Follow up (Year)
			Age	Gender	Health Status				
Cohort Study									
New-onset T2DM	Tan C, et al., 2022(47)	Japan	45.42 ± 8.72	Mixed	No T2DM at baseline; with abdominal obesity	DATADRYAD database, NAGALA Cohort	2012	14	3–6
New-onset T2DM	Yang J, et al., 2025(48)	Suzhou, China	< 18	Mixed	No T2DM at baseline	Children's Hospital of Soochow University	292	49	≥ 1
New-onset T2DM	Garcia C, et al., 2020(49)	Spain	≥ 18	Mixed	Baseline without T2DM; all with essential HTN	HTN Dept. Database, Mostoles University Hospital	1647	214	Max
New-onset T2DM	Liu Q, et al., 2020(50)	China	44.99 ± 12.98	Mixed	Some with HTN, dyslipidemia	China Rich Healthcare Group	82938	3069	2–3
SLD (confirmed by ultrasound)	Park J, et al., 2025(51)	South Korea	IQR 43–59	Mixed	NA	SNUH Gangnam Center, Kangbuk Samsung Hospital	20094	9216	NM
OSA	Zhu X, et al., 2025(54)	Shanghai, China	IQR 26–37	Mixed	Some with HTN/T2DM/HLD/NAFLD	Shanghai 9th People's Hospital	453	288	NM
First-time ASCVD	Zhou Y, et al., 2025(52)	USA	44–84	Male	Baseline without ASCVD; some with HTN, T2DM	ARIC (4 US communities), MESA (6 US centers)	2007	219	10
Systemic Thromboembolism	Yu M, et al., 2024(53)	NM	NM	Mixed	Some with HTN/HLD/CKD/T2DM/HF/Stroke	Hospital	952	22	1
Hyperuricemia	Jing J, et al., 2024(55)	Sichuan, China	46190	Mixed	Some with HUA	Affiliated Hospital of North Sichuan Medical College	1410	428	NM
New-onset CKD	Zhang H, et al., 2024(56)	China	58.3 ± 8.5	Mixed	Baseline without CKD; some with comorbidities (HTN, T2DM, etc.)	CHARLS	3246	334	4
Cross-sectional Study									
Hypertension	Yang X, et al., 2025(61)	Gansu, China	Mean 49.82	Mixed	NA	Dongxiang Enclave, Linxia, Gansu Province	1209	336	NA
Dyslipidemia	Liao J, et al., 2023(62)	Shenzhen, China	14(IQR13-15)	Mixed	NA	A weight loss institution in Shenzhen	M: 225, F: 196	M: 103, F: 64	NA
NAFLD	Zhou Y, et al., 2022(52)	Japan	45.05 ± 8.73	Mixed	Some with NAFLD	NAGALA Database	1516	743	NA
MAFLD	Song D, et al., 2022(57)	Shenzhen, China	32.81 ± 11.02	Mixed	NM	Shenzhen People's Hospital (and branches)	535	281	NA
MAFLD	Hu J, et al., 2025(78)	Zhengzhou, China	7–18	Mixed	Some with MAFLD	Children's Hospital Affiliated to Zhengzhou University	2512	1344	NA
Severe OSA (AHI ≥ 30)	Zhao A, et al., 2024(64)	Wuhan, China	28–38	Mixed	No OSA at baseline	Bariatric Surgery Center, Zhongnan	1008	496	NA

Study Design/Outcome	First Author, Year	Country/Region	Population			Sample Source	Sample Size	No. of Events	Follow up (Years)
			Age	Gender	Health Status				
Cohort Study									
						Hospital of Wuhan University			
OHS	Chen W, et al., 2023(79)	China	16–65	Mixed	Some with OHS/HTN/T2DM	Bariatric Surgery Center of a Tertiary University Hospital	577	103	NA
Atherosclerosis Index	Yu H, et al., 2025(66)	China	38–76	Mixed	T2DM	Characteristic Medical Center of the Chinese People's Armed Police Force	158	62	NA
Subclinical Carotid Atherosclerosis	Geng D, et al., 2024(67)	China	53.2 ± 12.7	Male	Some with HTN/T2DM	Physical examination staff, Hebei General Hospital	8745	4175	NA
Diabetic Nephropathy	Shi J, et al., 2025(68)	China	≥ 18	Mixed	T2DM	Dept. of Endocrinology, 3rd Affiliated Hospital of Sun Yat-sen University	370	79	NA
Hyperuricemia	Li L, et al., 2021(58)	China	18–85	Mixed	Some with HTN	Physical examination pop., Affiliated Hospital of Guilin Medical University	5098	1834	NA
Major Depressive Disorder	Yu X, et al., 2024(59)	USA	50.88 ± 16.37	Mixed	Some with depression	NHANES	2674	306	NA
Major Depressive Disorder	Yu C, et al., 2025(60)	China	≥ 18	Mixed	Some with depression/HTN/T2DM	CHARLS	1392	NA	NA
Cognitive Impairment	Lei C, et al., 2024(69)	China	63–75	Mixed	Some with HTN/T2DM/Asthma/Depression	NHANES	1490	376	NA

Abbreviations: T2DM, type 2 diabetes mellitus; SLD, steatotic liver disease; OSA, obstructive sleep apnea; Severe OSA, severe obstructive sleep apnea (AHI≥30); ASCVD, atherosclerotic cardiovascular disease;CKD, chronic kidney disease; NAFLD, non-alcoholic fatty liver disease; MAFLD, metabolic associated fatty liver disease; OHS, obesity-hypoventilation syndrome; HTN, hypertension; HLD, hyperlipidemia; GA, glucose abnormality; CCD, cardiocerebrovascular disease; AI, atherosclerosis index; SCAS, subclinical carotid atherosclerosis; STE, systemic thromboembolism; HUA, hyperuricemia; HF, heart failure; CVD, cardiovascular disease; ARIC, Atherosclerosis Risk in Communities; MESA, Multi-Ethnic Study of Atherosclerosis; CHARLS, China Health and Retirement Longitudinal Study; NHANES, National Health and Nutrition Examination Survey; NA, not available; NM, not mentioned.

Model specifications: ¶Modeling with different machine learning algorithms; ¶Modeling with different predictors and different algorithms; ¶Model predicting 3-year diabetes-free survival vs. 5-year diabetes-free survival; ¶Modeling with different predictors; ¶Modeling with different predictors; ¶Model A: partial predictors converted from continuous to categorical variables; Model B: all variables converted to categorical variables; ¶Male model vs. Female model; ¶Modeling with different predictors; ¶The second model was removed due to a high number of predictors but low performance and a lack of validation information; ¶Classification tree model vs. Nomogram model.

The models covered glucose abnormalities, NAFLD, OSA, cardiocerebrovascular diseases, kidney diseases, hypertension, dyslipidaemia, mental disorders, and cognitive impairment. New-onset type 2 diabetes was the most frequently predicted outcome (4 studies). Fourteen studies (58.3%) developed a single model, whereas the remainder developed multiple models (range: 2–8) using different algorithms, predictors, or subgroups. (Table 1)

3.3 Model characteristics

3.3.1 Data preprocessing

Most models (n = 26, 60.5%) retained continuous variables in their original form; others applied standardisation (n = 11, 25.6%), z-score transformation (n = 3, 7%), categorisation (n = 4, 9.3%), or logarithmic transformation (n = 2, 4.7%). Missing data were most commonly handled by complete-case analysis (n = 30,

69.8%), with simple imputation (n = 8, 18.6%), linear interpolation (n = 3, 7%), and multiple imputation (n = 2, 4.7%) used less frequently.(Table S6)

3.3.2 Predictor selection, predictors, and modelling methods

The most common predictor selection strategy was univariable screening combined with multivariable logistic regression (n = 14, 32.6%), followed by LASSO regression (n = 7, 16.3%) and univariable analysis alone (n = 6, 14%). Most models included 5–8 predictors (> 70%). Frequently selected predictors were age (15 studies), BMI (12 studies), and sex (8 studies); clinical history variables, particularly hypertension and type 2 diabetes, were also common. Biochemical markers related to glucose and lipid metabolism (e.g., fasting glucose, glycated haemoglobin, and triglycerides) appeared frequently, whereas specialised predictors such as body composition indices, insulin resistance measures, and advanced biomarkers were seldom used. (Fig. 1, Fig. 2, Table S6, S8, S9)

Logistic regression was the predominant modelling method (n = 24, 55.8%), followed by Cox proportional hazards models (n = 4, 9.3%). Machine learning approaches—including random forests, XGBoost, and support vector machines—were used in 27.9% of models. (Fig. 1, Table S6)

3.3.3 Model performance and validation

All included models reported apparent performance. Discrimination and calibration were jointly assessed in 18 models (41.9%), while 11 (25.6%) additionally evaluated potential clinical utility using decision curve analysis. Calibration plots were reported in most models (24, 55.8%), and among studies reporting Hosmer-Lemeshow tests (22, 51.2%), all reported p values > 0.05. Internal validation was commonly performed using cross-validation (19, 44.2%), split-sample validation (16, 37.2%), or bootstrapping (11, 25.6%). However, external validation was conducted in only two models (4.7%).

Across all 43 models, reported AUC values ranged from 0.591 to 0.961 (mean 0.829), with cohort study models showing higher mean AUC (0.857; range 0.758–0.961) than cross-sectional models (0.794; range 0.591–0.920). Fifteen models across seven disease outcomes met the threshold (≥ 2 models with complete CI data) for pooling (Fig. 3, Table S7). Fifteen models across seven disease outcomes met the threshold (≥ 2 models with complete CI data) for pooling.

Among cohort studies, pooled AUCs were 0.861 (95% CI 0.840–0.882; $I^2 = 0.5\%$, $\tau^2 < 0.0001$) for new-onset type 2 diabetes (3 models across 4 studies (47–50)), 0.820 (95% CI 0.813–0.827; $I^2 = 0.0\%$, $\tau^2 < 0.0001$) for sonographically confirmed steatotic liver disease (3 models (51)), 0.788 (95% CI 0.413–1.162; $I^2 = 1.0\%$) for first-time atherosclerotic cardiovascular disease (2 models (52)), and 0.885 (95% CI 0.885–0.885; $I^2 = 0.0\%$) for systemic thromboembolism (2 models (53)). Single-model outcomes not eligible for pooling included OSA (AUC 0.880 (54)), hyperuricemia-related kidney disease (AUC 0.840 (55)), and new-onset CKD (AUC 0.789(56)).

Among cross-sectional studies, pooled AUCs were 0.770 (95% CI 0.637–0.902; $I^2 = 1.0\%$, $\tau^2 = 0.0112$) for metabolic dysfunction-associated fatty liver disease (5 models across 2 studies (57)), 0.789 (95% CI 0.764–0.814; $I^2 = 0.0\%$) for hyperuricemia (2 models (58)), and 0.766 (95% CI 0.118–1.414; $I^2 = 0.9\%$, $\tau^2 = 0.0049$) for major depressive disorder (2 models(59, 60)). Single-model outcomes included hypertension (AUC 0.796 (61)), dyslipidaemia (AUC 0.620–0.700 (62)), NAFLD (AUC 0.863 (63)), severe OSA (AUC 0.845 (64)), OHS (AUC 0.830 (65)), atherosclerosis index (AUC 0.920 (66)), subclinical carotid atherosclerosis (AUC 0.805 (67)), diabetic nephropathy (AUC 0.913 (68)), and cognitive impairment (AUC 0.842 (69)).(Fig. 3, Table S7)

3.4 Risk of bias and applicability

Overall, 22 of 43 models (51.2%) were rated at high risk of bias using PROBAST (Figure S2). The analysis domain was the primary source: 44.2% (n = 19) of models showed high risk, predominantly due to inappropriate predictor selection, inadequate handling or reporting of missing data, and mismatch between evaluation methods and study objectives. In contrast, the participant domain showed uniformly low risk; the predictor and outcome domains also demonstrated low risk (88.4% and 90.7%, respectively). Regarding applicability, all models were judged to have low concern across domains. Using PROBAST + AI to assess 14 machine learning models, only 42.9% (n = 6) were classified as low risk and high quality. Detailed results are in Table S11-12.

4 Discussion

4.1 Principal findings

This review is among the first to synthesise prediction models across a broad range of obesity-related diseases within a common methodological framework, allowing model performance, predictor composition, validation status, and risk of bias to be examined across conditions that are usually evaluated separately. This systematic review included 24 studies comprising 43 prediction models for obesity-related diseases, covering a broad range of outcomes, including metabolic, cardiovascular, and sleep-related conditions.

Overall, most models demonstrated favourable apparent or internally validated discrimination performance. However, model performance varied substantially across disease categories and populations, suggesting limited generalisability across different clinical contexts. Besides, calibration assessment was less consistently reported, and findings based solely on non-significant Hosmer-Lemeshow tests should be interpreted cautiously. A non-significant Hosmer-Lemeshow test should not be interpreted as sufficient evidence of good calibration, because the test is sensitive to sample size and provides limited information regarding the direction and magnitude of miscalibration. Calibration plots, calibration-in-the-large, and calibration slopes generally provide more informative assessments of calibration performance.

However, despite these apparently favourable performance metrics, important methodological limitations were identified. Half of the models were judged to be at high risk of bias, predominantly driven by deficiencies in the analysis domain, including suboptimal predictor selection, inadequate handling of missing data, and inconsistencies in performance evaluation. In addition, external validation was rarely performed, with only a small proportion of models undergoing validation in independent populations. Taken together, these findings suggest that although existing models show promising predictive performance, their methodological robustness, reproducibility, and generalisability remain limited.

4.2 Interpretation of key findings

4.2.1 Apparent Performance versus Real Performance

High reported AUC values do not guarantee real predictive ability; overfitting is a likely contributor. Models developed and evaluated within the same dataset tend to capture noise rather than signal. The substantial heterogeneity observed in several pooled analyses further suggests that performance may vary considerably across populations. Comprehensive external validation combined with thorough calibration assessment is therefore essential before clinical deployment.(70)

4.2.2 Unclear Risk of Bias in the Analysis Domain

The high prevalence of bias in the analysis domain aligns with previous evidence that up to 75% of obesity-related prediction models are at high risk of bias. (70) Recurring issues—inappropriate predictor selection, inadequate missing data handling, and misalignment between model objectives and evaluation methods—reflect a broader tendency to prioritise apparent performance over methodological rigour. Addressing these shortcomings is critical for producing clinically credible models.

4.2.3 Traditional Statistical Models versus Machine Learning (ML) Models

Machine learning methods have been increasingly applied in obesity-related prediction modelling and are often considered promising alternatives to traditional statistical approaches(71, 72). In this review, however, traditional models—particularly logistic regression—remained the dominant approach. Available evidence did not demonstrate a clear and consistent advantage of machine-learning models over conventional regression models. However, direct comparisons between machine-learning and traditional models were limited, and most models lacked independent external validation. This is in line with existing literature, which suggests that although machine learning can handle complex, high-dimensional data and identify nonlinear relationships, its superiority over traditional models is not yet well established(70). Moreover, machine learning models often require larger sample sizes and more rigorous validation, and their interpretability remains a challenge in clinical practice. Therefore, while machine learning holds potential, its advantages in obesity-related disease prediction remain uncertain and context-dependent.

4.2.4 Predictor patterns

The predominance of routinely available variables (age, BMI, sex, and common metabolic biomarkers) reflects a pragmatic approach that enhances clinical usability but may constrain predictive ceiling. More specialised predictors—including advanced body composition indices, inflammatory markers, and imaging-derived measures—remained underutilised and their incremental value is not yet established.

(73–75) Future work should systematically evaluate whether these emerging predictors improve performance beyond traditional risk factors.

4.3 Comparison with existing literature

Previous systematic reviews have examined prediction models for obesity and related diseases, often focusing on specific outcomes such as childhood obesity, diabetes, or cardiovascular risk(34, 76). Some studies have also compared modelling approaches, including traditional statistical methods and machine learning techniques, although findings regarding their relative performance remain inconsistent(71, 77). However, most existing reviews are limited to single diseases or specific populations, and comprehensive assessment of methodological quality has not been consistently undertaken(34).

In contrast, this review provides a cross-disease synthesis of obesity-related prediction models, incorporates a structured methodological quality assessment using PROBAST, and quantitatively synthesises discrimination performance through meta-analysis. Our findings confirm that, despite generally favourable reported and pooled discrimination performance, substantial methodological limitations—particularly within the analysis domain and the limited use of external validation—remain common across studies.

These findings have important implications. Although many models report favourable performance, their limited robustness and generalisability restrict their immediate use in clinical practice. Future research should prioritise methodological quality, including appropriate sample size, transparent handling of predictors and missing data, and, critically, external validation in independent populations. In addition, while machine learning approaches are increasingly used, their added value over traditional models remains uncertain, highlighting the need for more rigorous comparative and validation studies(71, 77).

4.4 Strengths and limitations

This review has several strengths. First, we systematically searched PubMed, Web of Science, EMBASE, and CNKI, ensuring comprehensive and representative evidence coverage. Second, we synthesized evidence across multiple disorders to provide a holistic field overview. Third, we systematically summarised model discrimination, calibration, clinical applicability, and quantitatively synthesised discrimination performance, thereby providing a more comprehensive evaluation of model performance across obesity-related diseases. Fourth, we rigorously assessed methodological quality using PROBAST and PROBAST + AI, and analyzed predictor selection patterns with a robust framework. Fifth, this review covered both traditional statistical and machine learning models, delivering a more complete perspective and informative conclusions.

Several limitations should be acknowledged. First, the search was restricted to specific databases and a defined time frame, which may have omitted early studies and grey literature. Second, incomplete reporting in some included studies limited the extraction of key data. Third, only incident risk prediction models were included, excluding prognostic and treatment-response models. Fourth, substantial heterogeneity across populations, outcomes, predictor definitions, and modelling methods was observed in the meta-analysis, which may limit the interpretability and generalisability of pooled estimates.

5 Conclusion

This systematic review synthesised evidence on prediction models for obesity-related diseases across multiple conditions. Although existing predictive models for obesity-related diseases generally demonstrate satisfactory discrimination performance; however, most predictive metrics are derived solely from development datasets or internal validation cohorts, accompanied by prevalent risk of bias. Independent external validations remain scarce, which substantially restricts their generalizability and clinical readiness for routine application. Future research should prioritise methodological rigour and external validation to improve the reliability and clinical applicability of prediction models in this field.

Declarations

Contributors

Limin Wang and Mengfan Wang contributed to the conception and design of the study, data acquisition, analysis, and interpretation, and drafted the manuscript. Zhuo Chen and Chuanlin Cao contributed to data curation and manuscript writing. Yuxiang Ma and Mengyun Sui contributed to data visualisation. Xiaohua Ying is the guarantor and accepts full responsibility for the finished work and the conduct of the study, had access to the data, and controlled the decision to publish. All authors reviewed and approved the final manuscript.

Funding

The project was supported by the Noncommunicable Chronic Diseases-National Science and Technology Major Project (2024ZD0543500).

Disclaimer

The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests

None declared.

Patient and public involvement

No patient involved

Ethics approval

This study was based on published data and did not involve human participants or primary data collection. Therefore, ethical approval was not required.

Data availability statement

The data analysed in this systematic review and meta-analysis were extracted from previously published studies cited in this article. The extracted dataset used for the evidence synthesis and meta-analysis is available from the corresponding author upon reasonable request. No new individual participant-level data were collected or generated in this study.

Clinical trial number

Not applicable.

Consent to publish

Not applicable.

Consent to participate

Not applicable.

References

1. Lin, X. et al. Global burden of noncommunicable disease attributable to high body mass index in 195 countries and territories, 1990–2017. *Endocrine* **69** (2), 310–320 (2020).
2. Song, Y., Zhou, Y., Feng, X., Fu, J. & Liu, Y. The global death and disability burden associated with a high BMI in children and adolescents, 1990–2019. *Front. Endocrinol. (Lausanne)*. **15**, 1463002 (2024).
3. Xie, F. et al. Global, regional, and national burden of mortality and DALYs attributable to high body mass index from 1990 to 2021 with projections to 2036. *BMC Public. Health*. **25** (1), 2053 (2025).
4. Height body-mass index trajectories of school-aged children and adolescents from 1985 to 2019 in 200 countries and territories: a pooled analysis of 2181 population-based studies with 65 million participants. *Lancet* **396** (10261), 1511–1524 (2020).
5. Dai, H. et al. The global burden of disease attributable to high body mass index in 195 countries and territories, 1990–2017: An analysis of the Global Burden of Disease Study. *PLoS Med.* **17** (7), e1003198 (2020).

6. Abarca-Gómez, L. et al. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128·9 million children, adolescents, and adults. *Lancet* **390** (10113), 2627–2642 (2017).
7. Zhou, X. D. et al. Burden of disease attributable to high body mass index: an analysis of data from the Global Burden of Disease Study 2021. *EClinicalMedicine* **76**, 102848 (2024).
8. Arnold, M. et al. Global burden of cancer attributable to high body-mass index in 2012: a population-based study. *Lancet Oncol.* **16** (1), 36–46 (2015).
9. Tan, D. J. H. et al. Rising global burden of cancer attributable to high BMI from 2010 to 2019. *Metabolism* **152**, 155744 (2024).
10. Jehan, S. et al. Obesity, obstructive sleep apnea and type 2 diabetes mellitus: Epidemiology and pathophysiologic insights. *Sleep. Med. Disord.* **2** (3), 52–58 (2018).
11. Chu, D. T. et al. An update on physical health and economic consequences of overweight and obesity. *Diabetes Metab. Syndr.* **12** (6), 1095–1100 (2018).
12. Sarma, S., Sockalingam, S. & Dash, S. Obesity as a multisystem disease: Trends in obesity rates and obesity-related complications. *Diabetes Obes. Metabolism.* **23**, 3–16 (2021).
13. Yang, M., Liu, S. & Zhang, C. The Related Metabolic Diseases and Treatments of Obesity. *Healthc. (Basel)* ;**10**(9). (2022).
14. Yao, Z. et al. Associations between Class I, II, or III Obesity and Health Outcomes. *NEJM Evid.* **4** (4), EVIDoa2400229 (2025).
15. Reeves, H. L., Zaki, M. Y. & Day, C. P. Hepatocellular Carcinoma in Obesity, Type 2 Diabetes, and NAFLD. *Dig. Dis. Sci.* **61** (5), 1234–1245 (2016).
16. Tasali, E., Pamidi, S., Covassin, N. & Somers, V. K. Obstructive Sleep Apnea and Cardiometabolic Disease: Obesity, Hypertension, and Diabetes. *Circul. Res.* **137** (5), 764–787 (2025).
17. Shah, S. et al. Assessment of health-related quality of life among patients with obesity, hypertension and type 2 diabetes mellitus and its relationship with multimorbidity. *PLoS One.* **18** (8), e0289502 (2023).
18. Friedman, N. & Fanning, E. L. Overweight and obesity: an overview of prevalence, clinical impact, and economic impact. *Dis. Manag.* **7** (Suppl 1), S1–6 (2004).
19. Chu, D-T. et al. An update on physical health and economic consequences of overweight and obesity. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews.* ;**12**(6):1095–100. (2018).
20. Sharma, A., Adekunle, B. I., Ogeawuchi, J. C., Abayomi, A. A. & Onifade, O. AI-Driven Patient Risk Stratification Models in Public Health: Improving Preventive Care Outcomes through Predictive Analytics. *Int. J. Multidisciplinary Res. Growth Evaluation* (2023).
21. van Daalen, K. R. et al. Risk estimation for the primary prevention of cardiovascular disease: considerations for appropriate risk prediction model selection. *Lancet Glob Health.* **12** (8), e1343–e58 (2024).
22. Oddy, C., Zhang, J., Morley, J. & Ashrafi, H. Promising algorithms to perilous applications: a systematic review of risk stratification tools for predicting healthcare utilisation. *BMJ Health Care Inf.* ;**31**(1). (2024).
23. Ngozi, L. et al. Improving healthcare decision-making with predictive analytics: A conceptual approach to patient risk assessment and care optimization. *Int. J. Sch. Res. Med. Dentistry* :1–010. (2024).
24. Aidoo, E. Advancing precision medicine and health education for chronic disease prevention in vulnerable maternal and child populations. *World J. Adv. Res. Reviews.* **25**, 2355–2376 (2025).
25. Dhiman, P. et al. Methodological conduct of prognostic prediction models developed using machine learning in oncology: a systematic review. *BMC Med. Res. Methodol.* **22** (1), 101 (2022).
26. Meehan, A. J. et al. Clinical prediction models in psychiatry: a systematic review of two decades of progress and challenges. *Mol. Psychiatry.* **27** (6), 2700–2708 (2022).
27. Andaur Navarro, C. L. et al. Risk of bias in studies on prediction models developed using supervised machine learning techniques: systematic review. *Bmj* **375**, n2281 (2021).
28. Efthimiou, O. et al. Developing clinical prediction models: a step-by-step guide. *BMJ* ;**386**. (2024).
29. Van Calster, B., Steyerberg, E. W., Wynants, L. & van Smeden, M. There is no such thing as a validated prediction model. *BMC Med.* **21** (1), 70 (2023).
30. Moons, K. et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. *Ann. Intern. Med.* **170**, W1 (2019).
31. Moons, K. et al. PROBAST + AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. *BMJ* **388**, e082505 (2025).
32. Damen, J. A. A., Moons, K. G. M., van Smeden, M. & Hooft, L. How to conduct a systematic review and meta-analysis of prognostic model studies. *Clin. Microbiol. Infect.* **29** (4), 434–440 (2023).
33. Riley, R. D. et al. Evaluation of clinical prediction models (part 3): calculating the sample size required for an external validation study. *Bmj* **384**, e074821 (2024).
34. Gou, H., Song, H., Tian, Z. & Liu, Y. Prediction models for children/adolescents with obesity/overweight: A systematic review and meta-analysis. *Prev. Med.* **179**, 107823 (2024).
35. Singh, P. et al. Prognostic Models for Predicting Remission of Diabetes Following Bariatric Surgery: A Systematic Review and Meta-analysis. *Diabetes Care.* **44** (11), 2626–2641 (2021).
36. de Abreu Sesconetto, L. et al. Scores for Predicting Diabetes Remission in Bariatric Surgery: a Systematic Review and Meta-analysis. *Obes. Surg.* **33** (2), 600–610 (2023).

37. Rayner, D. G. et al. Prognostic models for cardiovascular and kidney outcomes in people with type 2 diabetes: living systematic review and meta-analysis of observational studies. *BMJ Med.* **4** (1), e001369 (2025).
38. Niakan Kalhori, S. R., Najafi, F., Hasannejadasl, H. & Heydari, S. Artificial intelligence-enabled obesity prediction: A systematic review of cohort data analysis. *Int. J. Med. Inf.* **196**, 105804 (2025).
39. Kee, O. T. et al. Cardiovascular complications in a diabetes prediction model using machine learning: a systematic review. *Cardiovasc. Diabetol.* **22** (1), 13 (2023).
40. Nazirun, N. N. N. et al. Prediction Models for Type 2 Diabetes Progression: A Systematic Review. *IEEE Access.* **12**, 161595–161619 (2024).
41. Fregoso-Aparicio, L., Noguez, J., Montesinos, L. & García-García, J. A. Machine learning and deep learning predictive models for type 2 diabetes: a systematic review. *Diabetol. Metab. Syndr.* **13** (1), 148 (2021).
42. Abdulazeem, H., Whitelaw, S., Schauburger, G. & Klug, S. J. A systematic review of clinical health conditions predicted by machine learning diagnostic and prognostic models trained or validated using real-world primary health care data. *PLoS One.* **18** (9), e0274276 (2023).
43. An, R., Shen, J. & Xiao, Y. Applications of Artificial Intelligence to Obesity Research: Scoping Review of Methodologies. *J. Med. Internet Res.* **24** (12), e40589 (2022).
44. Garvey, W. T. et al. American Association of Clinical Endocrinologists and American College of Endocrinology Comprehensive Clinical Practice Guidelines for Medical Care of Patients with Obesity. *Endocr. Pract.* **22** (7), 842–884 (2016).
45. National Health Commission of the People's Republic of China. Clinical Guidelines for Obesity (2024 Version): Document No. Guo Wei Ban Yi Zheng Han 2024382[A].
46. Forouzanfar MH, Alexander L, Anderson HR, Bachman VF, Biryukov S, Brauer M, et al. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013. *The Lancet.* 2015;386(10010):2287–323.
47. Tan, C. et al. A Prediction Model of the Incidence of Type 2 Diabetes in Individuals with Abdominal Obesity: Insights from the General Population. *Diabetes Metab. Syndr. Obes.* **15**, 3555–3564 (2022).
48. Yang, J. X. et al. Development and internal validation of a machine learning algorithm for the risk of type 2 diabetes mellitus in children with obesity. *Front. Endocrinol. (Lausanne).* **16**, 1649988 (2025).
49. Garcia-Carretero, R. et al. Use of a K-nearest neighbors model to predict the development of type 2 diabetes within 2 years in an obese, hypertensive population. *Med. Biol. Eng. Comput.* **58** (5), 991–1002 (2020).
50. Liu, Q. et al. Development and Validation of a Nomogram to Predict Type 2 Diabetes Mellitus in Overweight and Obese Adults: A Prospective Cohort Study from 82938 Adults in China. *Int. J. Endocrinol.* **2020**, 8899556 (2020).
51. Park, J. et al. A Novel Point-of-Care Prediction Model for Steatotic Liver Disease: Expected Role of Mass Screening in the Global Obesity Crisis. *Gut Liver.* **19** (1), 126–135 (2025).
52. Zhou, Y. et al. Equations for Predicting the 10-Year Risk of Atherosclerotic Cardiovascular Disease in Population With Obesity. *JACC Adv.* **4** (2), 101542 (2025).
53. Yu, M., Li, X., Zong, L., Wang, Z. & Lv, Q. A Novel Body Mass Index-Based Thromboembolic Risk Score for Overweight Patients with Nonvalvular Atrial Fibrillation. *Anatol. J. Cardiol.* **28** (1), 35–43 (2024).
54. Zhu, X. et al. A Prediction Model for Obstructive Sleep Apnea in Bariatric Surgery Candidates with Obesity. *Obes. Surg.* **35** (10), 4205–4215 (2025).
55. JING Jianying, X. N., Xuemei, G. U. O., Xueming, J. I. N. G., Rong, X. I. E. & Yonglong, H. E. Prediction and evaluation of nomogram model on risk of hyperuricemia in overweight and obese children and adolescents. *Chongqing Med. J.* **53** (2), 220–225 (2024).
56. Zhang, H., Zhang, Y., Gao, W. & Mu, Y. Identification of risk factors and development of a predictive model for chronic kidney disease in patients with obesity: a four-year cohort study. *Lipids Health Dis.* **23** (1), 57 (2024).
57. Song, D. et al. Development and Validation of a Nomogram for Prediction of the Risk of MAFLD in an Overweight and Obese Population. *J. Clin. Transl Hepatol.* **10** (6), 1027–1033 (2022).
58. Li Luwei, H. Q. et al. Construction and evaluation of hyperuricemia prediction model for overweight and obese population. *J. Practical Med.* **37** (18), 2401–2406 (2021).
59. Yu, X. et al. A nomogram and online calculator for predicting depression risk in obese Americans. *Heliyon* **10** (13), e33825 (2024).
60. Yu, C., Cao, J., Chen, W. & Hong, E. Development and validation of a predictive model for depression risk in Chinese obese adults. *Front. Public. Health.* **13**, 1574386 (2025).
61. YANG Xiulin, C. P., Bin, M. A., Weihong, M. A. & Xiangjun, H. A. I. Construction and evaluation of a nomogram model for predicting the risk of hypertension in Dongxiang adults based on obesity and muscle mass indices. *Acta Anat. Sinica.* **56** (3), 336–341 (2025).
62. Jing, Z. L. & LIAO Nomograph model for predicting the risk of dyslipidemia with anthropometric indices in obese adolescents aged 13–16 years. *Chin. J. Child. Health Care.* **31** (4), 379–384 (2023).
63. Zhou, Y. et al. A Prediction Model of the Incidence of Nonalcoholic Fatty Liver Disease With Visceral Fatty Obesity: A General Population-Based Study. *Front. Public. Health.* **10**, 895045 (2022).
64. Zhao, A. et al. A Prediction Nomogram of Severe Obstructive Sleep Apnea in Patients with Obesity Based on the Liver Stiffness and Abdominal Visceral Adipose Tissue Quantification. *Nat. Sci. Sleep.* **16**, 1515–1527 (2024).
65. Chen, W. et al. A Novel Nomogram and Online Calculator for Predicting the Risk of Obesity Hypoventilation Syndrome in Bariatric Surgery Candidates. *Obes. Surg.* ;**33**. (2022).

66. YU Hui, S. G. & Yue, D. I. N. G. Risk factors and early warning risk model for elevated atherogenic index of plasma in overweight/obese type 2 diabetes patients. *Anhui Med. Pharm. J.* **29** (10), 2066–2070 (2025).

67. Dandan, G. *Risk Prediction Model for Subclinical Carotid Atherosclerosis in Centrally Obese Population: Establishment and Evaluation* (Hebei Medical University, 2024).

68. SHI Jie, Z. Y., Jiarui, L. I. N., Xiaodi, G. U. O., Shuhong, L. I. U. & Xiling, H. U. Development and validation of a nomogram-based assessment model for the risk of diabetic nephropathy in obese type 2 diabetes mellitus patients. *J. New. Med.* **56** (10), 986–994 (2025).

69. Lei, C. et al. Development and validation of a cognitive dysfunction risk prediction model for the abdominal obesity population. *Front. Endocrinol. (Lausanne)*. **15**, 1290286 (2024).

70. Boakye, N. F., O'Toole, C. C., Jalali, A. & Hannigan, A. Comparing logistic regression and machine learning for obesity risk prediction: A systematic review and meta-analysis. *Int. J. Med. Inf.* **199**, 105887 (2025).

71. Delpino, F. M. et al. Does machine learning have a high performance to predict obesity among adults and older adults? A systematic review and meta-analysis. *Nutr. Metab. Cardiovasc. Dis.* **34** (9), 2034–2045 (2024).

72. Ferdowsy, F., Rahi, K. S. A., Jabiullah, M. I. & Habib, M. T. A machine learning approach for obesity risk prediction. *Curr. Res. Behav. Sci.* **2**, 100053 (2021).

73. Lin, W. et al. Identification of influence factors in overweight population through an interpretable risk model based on machine learning: a large retrospective cohort. *Endocrine* **83** (3), 604–614 (2024).

74. Colmenarejo, G. Machine Learning Models to Predict Childhood and Adolescent Obesity: A Review. *Nutrients* ;**12**(8). (2020).

75. Helforoush, Z. & Sayyad, H. Prediction and classification of obesity risk based on a hybrid metaheuristic machine learning approach. *Front. Big Data.* **7**, 1469981 (2024).

76. Simmonds, M. et al. The use of measures of obesity in childhood for predicting obesity and the development of obesity-related diseases in adulthood: a systematic review and meta-analysis. *Health Technol. Assess.* **19** (43), 1–336 (2015).

77. Ferreras, A. et al. Systematic Review of Machine Learning applied to the Prediction of Obesity and Overweight. *J. Med. Syst.* **47** (1), 8 (2023).

78. Hu, J. et al. Development and validation of an online nomogram for screening metabolic-associated fatty liver disease in obese children. *Ital. J. Pediatr.* **51** (1), 307 (2025).

79. Chen, W. et al. A Novel Nomogram and Online Calculator for Predicting the Risk of Obesity Hypoventilation Syndrome in Bariatric Surgery Candidates. *Obes. Surg.* **33** (1), 68–77 (2023).

Figures

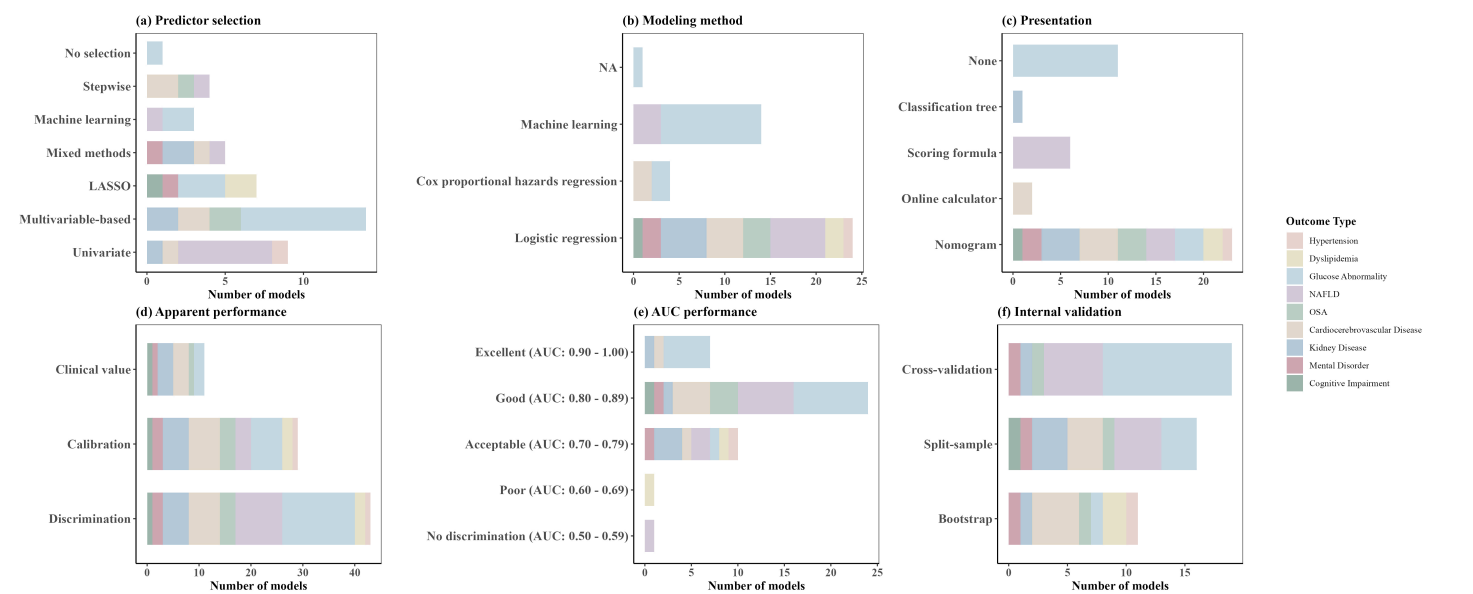


Figure 1

Methodological characteristics of included prediction models stratified by outcome type.

(a) Predictor selection methods; (b) modeling approaches; (c) model presentation formats; (d) apparent performance reporting; (e) discrimination performance based on AUC categories; (f) internal validation methods. Bars represent the number of models, and colors indicate different outcome types.

Abbreviations: AUC, area under the receiver operating characteristic curve.

Notes:(1) Predictor selection methods include univariate analysis, multivariable-based approaches, LASSO, machine learning–based selection, stepwise procedures, mixed methods, and no selection. (2) Modeling approaches are categorized as logistic regression, Cox proportional hazards regression, or machine learning methods. (3) Model presentation includes nomogram, online calculator, scoring formula, classification tree, or none. (4) Apparent

performance reporting includes discrimination, calibration, and clinical value. (5)AUC categories are defined as: no discrimination (0.50–0.59), poor (0.60–0.69), acceptable (0.70–0.79), good (0.80–0.89), and excellent (0.90–1.00). (6)Internal validation methods include bootstrap, split-sample validation, and cross-validation. Models using multiple validation methods are counted in each applicable category.

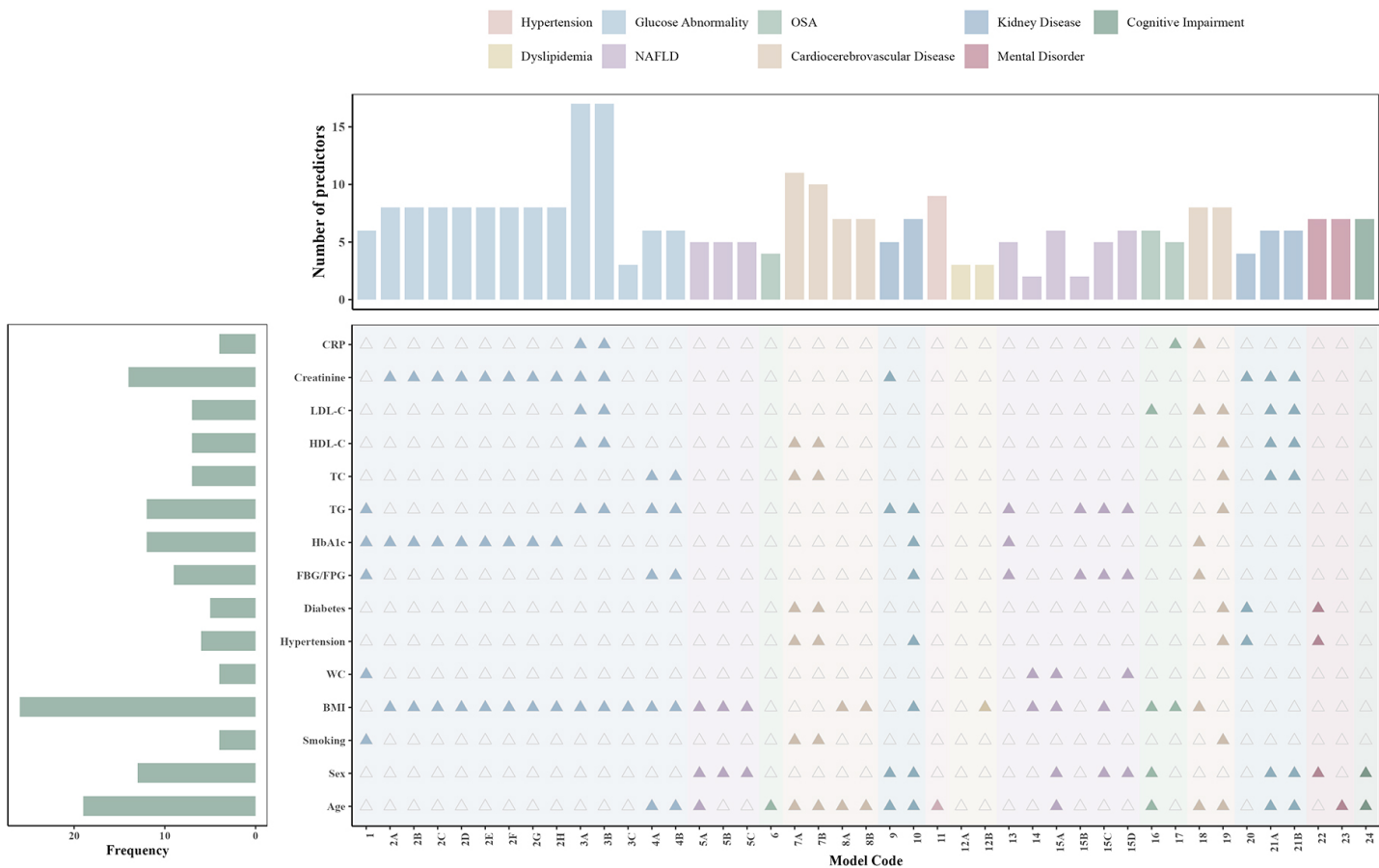


Figure 2

Distribution and combination patterns of frequently used predictors across included risk prediction models.

The upper panel shows the number of predictors included in each model, stratified by outcome type. The central matrix illustrates the presence of individual predictors across models, where filled triangles indicate inclusion of a predictor and hollow triangles indicate absence. The left panel displays the frequency of each predictor across all models. Only predictors with a frequency greater than four were included for visualization.

Abbreviations: BMI, body mass index; WC, waist circumference; FBG/FPG, fasting blood glucose/fasting plasma glucose; HbA1c, glycated hemoglobin; TG, triglycerides; TC, total cholesterol; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; Cr, creatinine; CRP, C-reactive protein.

Notes: (1) Each column represents an individual prediction model; (2)Models are grouped by outcome type; (3)Predictor frequencies were calculated as the number of models in which each predictor was included; (4) Predictors with a frequency ≤ 4 were excluded from the figure for clarity.

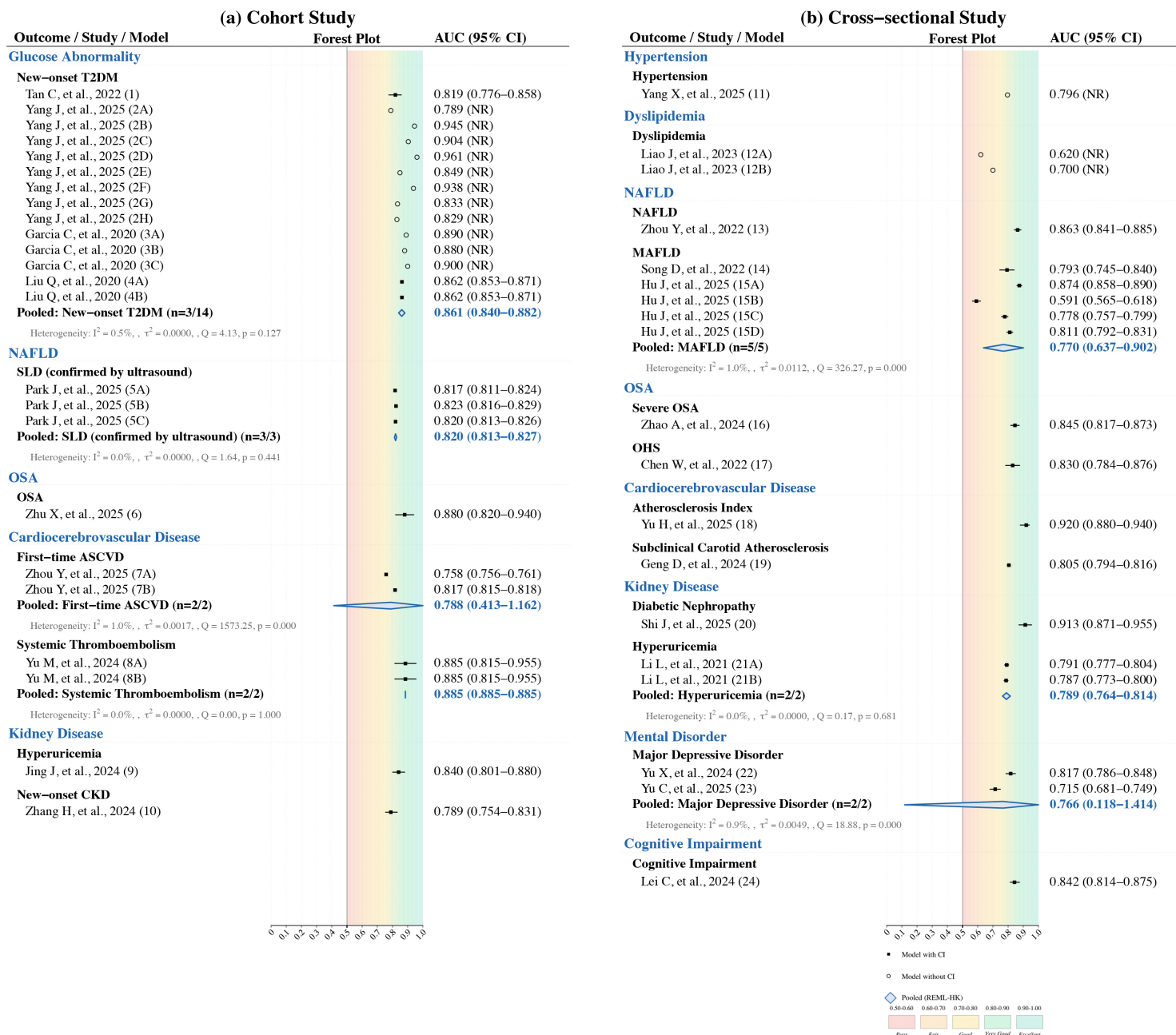


Figure 3

Forest plot of AUC estimates for obesity-related disease risk prediction models.

Random-effects meta-analysis with Hartung-Knapp adjustment was performed to synthesize model discrimination across outcome categories. Individual model estimates are presented as squares with 95% confidence intervals (CIs), while pooled subgroup estimates are shown as diamonds. Prediction intervals were calculated for the overall analysis to reflect the expected range of AUC values in future studies. Dashed vertical reference lines indicate conventional thresholds for model discrimination performance.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryAppendixReviewofriskpredictionmodels.docx](#)