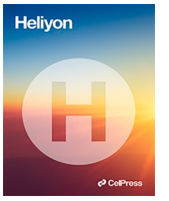
Heliyon 9 (2023) e15529

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Review article

Performance of a prediabetes risk prediction model: A systematic review

Yujin Liu [a], Wenming Feng [b], Jianlin Lou [d], Wei Qiu [c], Jiantong Shen [a,d,*], Zhichao Zhu [a,e], Yuting Hua [a], Mei Zhang [a], Laura Flavorta Billong [a]

[a] Schools of Nursing and Medicine, Huzhou University, Huzhou, 313000, China
[b] Huzhou First People's Hospital, Huzhou, 313000, China
[c] Department of Endocrinology, Huzhou Central Hospital, Huzhou, 313000, China
[d] Huzhou Key Laboratory of Precise Prevention and Control of Major Chronic Diseases, Huzhou University, Huzhou, 313000, China
[e] Internal Medicine General Ward, Jinhua Municipal Central Hospital Medical Group, Jinhua, 321200, China

ARTICLE INFO

Keywords:
Prediabetes
Risk prediction model
Screening tool
Systematic review

ABSTRACT

Backgrounds: The prediabetes population is large and easily overlooked because of the lack of obvious symptoms, which can progress to diabetes. Early screening and targeted interventions can substantially reduce the rate of conversion of prediabetes to diabetes. Therefore, this study systematically reviewed prediabetes risk prediction models, performed a summary and quality evaluation, and aimed to recommend the optimal model.

Methods: We systematically searched five databases (Cochrane, PubMed, Embase, Web Of Science, and CNKI) for published literature related to prediabetes risk prediction models and excluded preprints, duplicate publications, reviews, editorials, and other studies, with a search time frame of March 01, 2023. Data were categorized and summarized using a standardized data extraction form that extracted data including author; publication date; study design; country; demographic characteristics; assessment tool name; sample size; study type; and model-related indicators. The PROBAST tool was used to assess the risk of bias profile of included studies.

Findings: 14 studies with a total of 15 models were eventually included in the systematic review. We found that the most common predictors of models were age, family history of diabetes, gender, history of hypertension, and BMI. Most of the studies (83.3%) had a high risk of bias, mainly related to under-reporting of outcome information and poor methodological design during the development and validation of models. Due to the low quality of included studies, the evidence for predictive validity of the available models is unclear.

Interpretation: We should pay attention to the early screening of prediabetes patients and give timely pharmacological and lifestyle interventions. The predictive performance of the existing model is not satisfactory, and the model building process can be standardized and external validation can be added to improve the accuracy of the model in the future.

1. Introduction

Prediabetes (PDM) is a large reservoir of diabetic patients. According to the latest data from the International Diabetes Federation

* Corresponding author.
E-mail address: sjiantong@163.com (J. Shen).

https://doi.org/10.1016/j.heliyon.2023.e15529
Received 6 September 2022; Received in revised form 11 April 2023; Accepted 12 April 2023
Available online 6 May 2023

(IDF) [1], 541 million adults (10.6%) worldwide have impaired glucose tolerance and 319 million adults (6.2%) have impaired fasting glucose [2], a total that far exceeds the number of people with diabetes. Studies have shown that about 70% of prediabetes patients will progress to diabetes within 10 years, and the incidence of diabetes exceeds 90% within 20 years [3], and if prediabetes patients are not detected in time and take reasonable interventions, individuals are at increased risk of developing diabetes, chronic kidney disease, cardiovascular disease, cancer and dementia, which become a direct cause of reduced quality of life, increased medical and social expenditures and even direct cause of death [4].

Prediabetes mostly has no obvious clinical symptoms, is not easily detected, and is more likely to be overlooked, resulting in low patient attention, awareness, detection rates, and adherence to treatment [5]. Studies of diabetes prevention programs have shown that intensive lifestyle interventions and use of glucose-lowering medications can substantially reduce the incidence of diabetes [6,7], and that early screening, identification, prevention and management of patients with prediabetes [8], with little potential harm, can improve patient health outcomes and substantially reduce the incidence, mortality and healthcare costs of type 2 diabetes in the prediabetes population [9]. In clinical practice, the assessment of prediabetes continues to rely on clinical tests. The oral glucose tolerance test (OGTT), the traditional gold standard for clinical screening, requires repeated multiple blood collections and is limited by time, space, medical materials, and other conditions [10], and blood glucose values fluctuate with climate, medication use, constipation, and inadequate drinking water. Clinical screening criteria for fasting glucose and glycated hemoglobin are controversial [11], and fasting glucose has a certain underdiagnosis rate; standardized glycated hemoglobin testing has low clinical penetration, high cost, and complex operation. These invasive operations are not suitable for screening a wide range of people due to poor patient experience, time-consuming, and expensive. Therefore, there is an urgent need for easy-to-apply ancillary screening tools for identifying and screening individuals at high risk for prediabetes, delaying disease progression and preventing adverse outcomes.

The risk of progression from prediabetes to diabetes varies widely, and simple and reliable predictive models are far superior to invasive laboratory tests for initial screening, are suitable for large sample populations, and have a more reasonable cost-benefit ratio and higher social benefit. In addition, the use of validated predictive models to screen asymptomatic patients with prediabetes and regularly assess risk allows for flexible modification of intervention decisions to reduce the conversion rate of prediabetes and reduce the occurrence of long-term complications. However, most existing models have been developed for people with type 2 diabetes or abnormal glucose metabolism, and when applied to the prediabetes population, model performance is often poor, resulting in no model yet being the best model for screening patients with prediabetes.

Therefore, this study provides a systematic review of models specific to prediabetes, describing the characteristics of model development, including predictors, predicted outcomes, presentation, and validation methods. The applicability, specificity, and accuracy of the models were also comprehensively evaluated, with the aim of identifying the best performing screening tool for clinical decision making [12].

2. Methods

We followed the Cochrane guidelines for systematic review evaluation and data extraction for predictive modeling studies (CHARMS checklist) [13] and performed a systematic evaluation based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [14].

2.1. Search strategy and selection criteria

We systematically searched five databases, Cochrane, PubMed, Embase, Web of Science, and CNKI. The search time frame was from database creation to March 01, 2023. A combination of subject terms and free words was used to identify key articles on the development and/or validation of prediabetes prediction models. In addition, we manually searched for original articles describing model development, while retrospectively incorporating references to the literature to supplement the relevant literature (Fig. 1).

Two investigators (LYJ and ZZC) searched the full text of selected literature after screening based on article titles and abstracts to determine eligibility. When there is disagreement, a third researcher will be asked. Included studies could report multiple predictive models, and all articles reporting predictive models (i.e., containing at least two and more predictors), instruments, or scores for one or more variables were included. We included cohort studies, cross-sectional studies, case-control studies, and randomized controlled studies, and classified models into two categories, prognostic prediction models and diagnostic prediction models, based on the type of study. The methods of model construction and/or validation, especially internal and/or external validation, were also distinguished and described. When the predictive objects of the models included both diabetes and prediabetes, we excluded data indicators of prediabetes if they could not be extracted separately. In addition, included articles had to report the model construction process or methods, otherwise they were not included in the study. The specific search strategy is shown in Exhibit I.

2.2. Data analysis

Data was extracted and cross-checked independently by 2 investigators to complete consistent data extraction. If multiple models were described in a single article, data were extracted separately for each model. Extracted information included first author, year of publication, country, name of assessment tool, study type, sample size, number of positive events, prediabetes definition criteria, model development method, model validation method, predictors incorporated in the model, and model prediction performance (calibration, discrimination). When there was disagreement, a third investigator was asked.

We define the concept of prediabetes using three commonly used defining criteria, namely the World Health Organization (WHO),

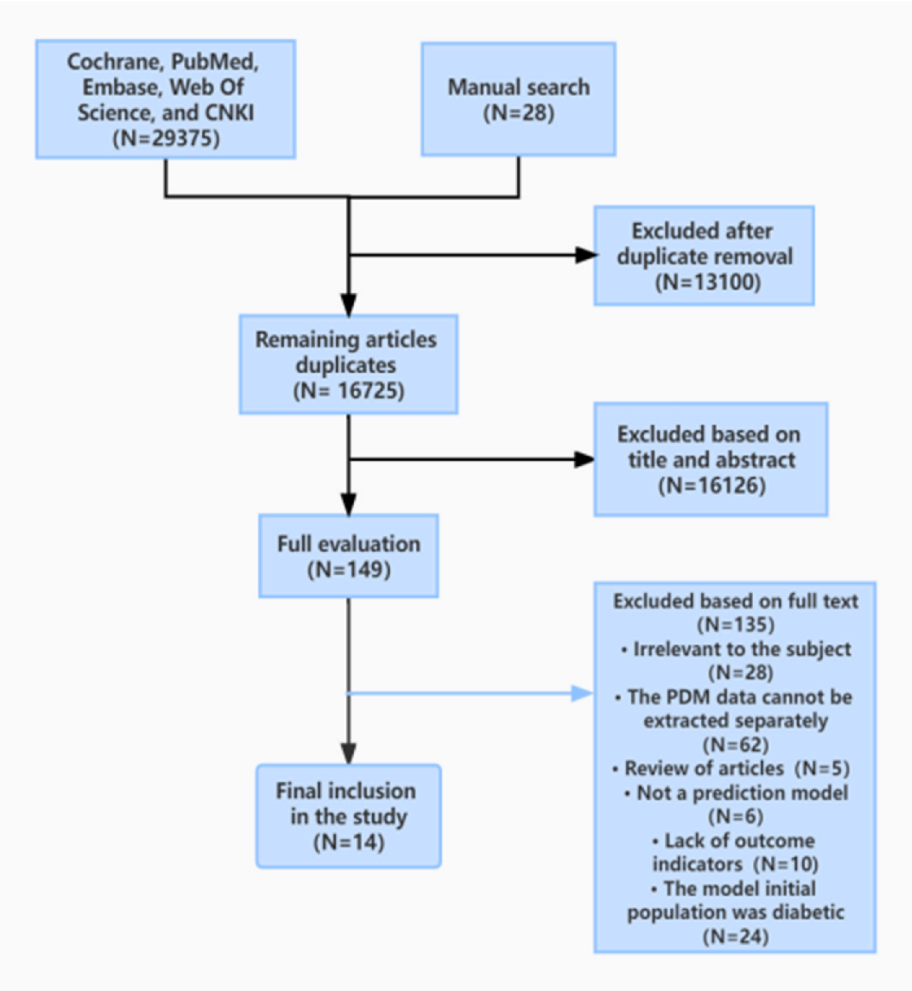


Fig. 1. Search results and study selection.

the American Diabetes Association (ADA), and the Canadian Diabetes Association (CDA) standard, to facilitate the extraction of prediabete data. The WHO definition of prediabetes: IFG is a fasting blood glucose (FPG) level of 110–125 mg/dL, IGT 2-h postprandial blood glucose level of 140–200 mg/dL [15]; ADA definition: IGT level is the same as WHO, IFG is FPG level of 100–125 mg/dL, while additionally considering glycosylated hemoglobin level in the range of 5.7%-6.4% [15,16]; CDA definition: fasting blood glucose (FPG) level of 5.6–6.9 mmol/L [17].

Meta Dics 1.4 and Revman 5.3 were used to combine the results of the obtained data, and $I^2 > 50\%$ in this study, suggesting a large heterogeneity, so the data were processed using a random effects model. Stata 14.0 was used to perform publication bias test on the data of this study. All included studies used discrimination and calibration as measures of model outcome. The discrimination was mainly measured by the area under the receiver operating curve (AUC) and also included sensitivity (SEN), specificity (SPE), positive likelihood ratio (LR+), negative likelihood ratio (LR−), and diagnostic ratio (DOR); the calibration of the model was described by the Hosmer-Lemeshow χ^2 test to describe the degree of conformity between the occurrence of the event and the actual occurrence afterwards.

2.3. Study quality

Using the prediction model risk of bias assessment tool (PROBAST) [18,19] specially designed for predictive model research, a total of 20 items were used to determine the risk of bias and applicability of the included literature. Risk of bias was assessed for four domains: participants, predictors, outcomes, and data analysis; applicability evaluation was performed for the first three domains. Independent assessments were conducted by LYJ, ZZC, and HYT, and any disagreements were resolved through conference discussions.

Table 1
Summaries the study and participant characteristics of the included trials.

Study	Prediction Model	Location	Design	Model Types	Development Methodology	Gender	Age	Sample size		Golden Standard	Definition Criteria	Validation Methodology
								Modeling (Positive Events)	Validation (Positive Events)			
Qing Wang et al. (2018) [20]	Jiangsu Diabetes Risk Score Self -Assessment Scale	China	Cross-sectional	Diagnostic Model	Logistic regression	Modeling: Female 436/Male 457 Validation: Female 198/Male 196	18–65	5430 (893)	2259 (394)	OGTT	WHO	Internal validation (Random split validation)
Fujiati II et al. (2017) [21]	INA-PRISC	Indonesia	Cross-sectional	Diagnostic Model	Logistic regression	Female 5033 Male 3878	>18	21,720 (8911)	6933 (548)	RPG	ADA	External validation
Rajput et al. (2018) [22]	PRESS	India	Cross-sectional	Diagnostic Model	Logistic regression	Female 82/Male 78	>18	892 (160)	–	OGTT	ADA	–
Koopman RJ et al. (2008) [23]	TAG-IT	USA	Cross-sectional	Diagnostic Model	Logistic regression	–	20–64	4045 (1014)	Data not shown	FBG	ADA	External validation
Hui Wang et al. (2015) [24]	Guangdong 2015	China	Cross-sectional	Diagnostic Model	Logistic regression	–	≥20	–	6197 (979)	FBG	ADA	Internal validation (Random split validation)
	TAG-IT					–		–	6197 (979)			External validation
	Shanghai 2009					–		–	6197 (979)			External validation
Ouyang Peng et al. (2016) [17]	Harbin 2015	China	Cross-sectional	Diagnostic Model	Logistic regression	Modeling: Female 140/Male 218 Validation: Female 171/Male 242	54.1 ± 16.1	1360 (358)	1739 (413)	FBG	CDA	Internal validation (Random split validation)
Yuqi Wang et al. (2021) [25]	IFG morbidity risk prediction model	China	Historical cohort	Prognostic model	Generalized Estimating Equations (GEE)	Female 156/Male 286	≥20	4926 (442)	4926 (442)	FBG	WHO	Internal validation (No method specified)
G.Štiglic et al. (2018) [26]	IFG simplified model	Slovenia	Cross-sectional	Diagnostic Model	Logistic regression	–	>18	2073 (435)	2073 (435)	FPG	WHO	Internal validation (Bootstrap Validation)
Min Liu et al. (2011) [27]	Beijing 2011	China	Prospective cohort	Prognostic model	Logistic regression	–	40–90	1851 (394)	699 (223)	OGTT	ADA	External validation
Kai Hu et al. (2012) [28]	Shanghai Huangpu Model	China	Case control	Prognostic model	Logistic regression	–	≥35	1769 (151)	–	OGTT	WHO	–

(*continued on next page*)

Table 1 (*continued*)

Study	Prediction Model	Location	Design	Model Types	Development Methodology	Gender	Age	Sample size		Golden Standard	Definition Criteria	Validation Methodology
								Modeling (Positive Events)	Validation (Positive Events)			
AbbasM et al. (2020) [29]	Qatar Rating	Qatar	Case control	Prognostic model	Logistic regression	Modeling: Female 959/Male 943 Validation: Female 244/Male 227	18–86	5814 (1902)	1454 (471)	HbA1C	ADA	Internal validation (v-fold cross-validation)
Danielle HE et al. (2017) [30]	PERSEUS	New Zealand	Prospective cohort	Prognostic model	Logistic regression	–	53 (41–65)	82 (29)	80 (37)	HbA1C/ FBG	ADA	Internal validation (Random split validation)
Jun Li et al. (2021) [31]	noninvasive diabetics risk prediction model	China	Case control	Prognostic model	Machine learning and deep learning	Female 359/Male 211	61 (57–68)	1710 (570)	Data not shown	OGTT	ADA	Internal validation (v-fold cross-validation)
Liyang Wang et al. (2020) [32]	IGRNet	China	Cross-sectional	Diagnostic Model	Deep learning	–	>60/ ≤60	2251 (501)	663 (160)	FPG/ 2hPG	WHO	Internal validation (v-fold cross-validation)

Role of the funding source

There was no funding source for this study. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

3. Results

A total of 29,375 relevant literatures were obtained from the preliminary search. After removing duplicates, 16,275 articles remained; by reading the titles and abstracts, 16,126 articles were excluded as irrelevant to the study topic, and 149 relevant articles were initially included. According to the nadir criteria and after reading through the full text, 14 publications [17,20–32] were finally included, containing 8 cross-sectional studies [17,20–24,26,32] and 6 non-cross-sectional studies [25,27–31], with a total of 15 models. The study population was the general population, except for the study by Danielle et al. [30], which was a predictive model of prediabetes risk developed for patients after an episode of pancreatitis. Table 1 summaries the study and participant characteristics of the included trials.

Since 2 studies [31,32] used classical machine learning and deep learning techniques and could not be included in the PROBAST quality assessment, the remaining 12 studies [17,20–30] were assessed for risk of bias. 10 studies were ultimately considered to be at high risk of bias. The details are shown in Fig. 2. The suitability evaluation yielded that the overall suitability of the included models was good in the three domains of study subjects, predictors, and outcomes.

The included models used 3–13 predictors, most models contained similar predictors, and candidate predictors were selected based on univariate or multivariate logistic regression analysis and clinical correlations. The most commonly used clinical predictors were age, family history of diabetes, sex, history of hypertension, and BMI. In addition, a few predictors (eg, race, education level, history of coronary artery disease, renal function status, smoking, dyslipidemia, etc.) were included in the model only once or twice. The predictors included in these models can usually be obtained by non-invasive means such as manual measurement, calculation or asking medical history, but there are also two studies that need to obtain indicators (blood lipids, renal function status) through invasive means such as puncture and blood collection. There are certain disadvantages in terms of efficiency. Table 2 lists all model predictors.

We summarized the predictive performance of the models and categorized models constructed based on cross-sectional study information as diagnostic predictive models, while models constructed based on cohort study or case-control study information as prognostic predictive models. The model construction and validation studies were also distinguished according to the model

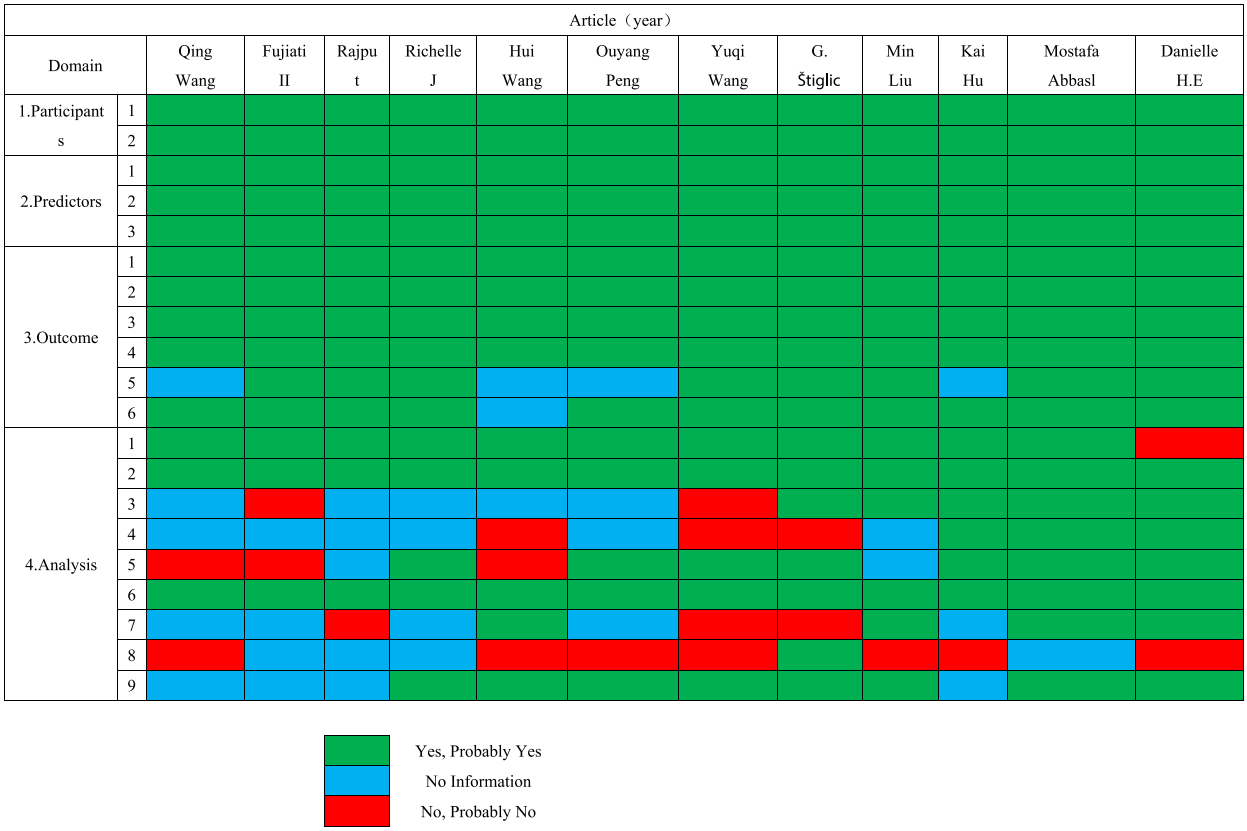


Fig. 2. Risk of model bias.

Table 2
Distribution of model predictors.

Study	age	gender	race	smoking history	EDU	RHR	BMI	WHR	obesity	PA	vegetable and fruit intake	hypertension history	SBP/ DBP	hyperglycemia history	FH of DM	History of CAD	DYSIS	renal function	Others
Qing Wang 2018	+	+				+	+	+					+		+				
Fujiati II 2017	+	+		+	+		+			+		+			+				
Rajput 2018	+							+					+		+				
Koopman RJ 2008	+	+				+	+					+			+				
Hui Wang 2015	+	+				+	+					+			+				
Ouyang Peng 2016	+						+					+	+		+		+		
Yuqi Wang 2021	+	+					+	+	+				+				+	+	
G.Štiglic 2018	+	+						+				+		+	+				
Min Liu 2011	+	+					+					+		+		+			
Kai Hu 2012	+							+					+		+				
Abbas M 2020	+	+						+				+			+				
Danielle HE 2017	+	+	+	+						+	+	+			+				
Jun Li 2021	+	+					+												12-lead ECG lasting 5s
Liyang Wang 2020																			(1) tongue color; (2) Tongue shape; (3) Tongue texture

EDU, Education; RHR, Resting Heart Rate; BMI, Body Mass Index; WHR, Waist-to-hip Ratio; PA, Physical Activity; SBP, Systolic Blood Pressure; DBP, Diastolic Blood Pressure; FH, Family History; DM, Diabetes Mellitus; CAD, Coronary Artery Disease; DYSIS; Dyslipidemia.

Table 3
Summary of the discrimination of the included models.

Diagnostic model		SEN		SPE		AUC	
		Modeling Data	Validating data	Modeling Data	Validating data	Modeling Data	Validating data
Qing Wang et al. (2018) [20]	Jiangsu Diabetes Risk Score Self -Assessment Scale	0.7	0.69	0.565	0.615	0.69	0.655
Fujiati II et al. (2017) [21]	INA-PRISC	0.5	0.55	0.67	0.66	0.623	0.646
Rajput et al. (2018) [22]	PRESS	0.844	–	0.585	–	0.785	–
Koopman RJ et al. (2008) [23]	TAG-IT	0.87	0.495	–	–	0.74	0.744
Hui Wang et al. (2015) [24]	TAG-IT	–	–	0.92	0.26	–	0.74
	Guangdong 2015	–	–	M:0.57/ F:0.69	M:0.72/F:0.6	–	M:0.7/F:0.7
	Shanghai 2009	–	–	0.69	0.55	–	0.7
Ouyang Peng et al. (2016) [17]	Harbin 2015	–	–	–	–	0.748	0.713
Liyang Wang et al. (2020) [32]	IGRNet	0.806	0.808	0.865	0.775	0.809	0.773
G. Štiglic et al. (2018) [26]	IFG simplified model	–	–	–	–	0.84	0.829
Prognostic model							
Yuqi Wang et al. (2021) [25]	IFG morbidity risk prediction model	0.705	0.715	0.661	0.668	0.74	0.751
Min Liu et al. (2011) [27]	Beijing 2011	–	–	–	–	–	0.828
Kai Hu et al. (2012) [28]	Shanghai Huangpu Model	0.682	–	0.643	–	0.788	–
Abbas M et al. (2020) [29]	Qatar Rating	0.553	0.579	0.856	0.862	0.79	0.8
Danielle HE et al. (2017) [30]	PERSEUS	0.552	–	0.981	–	0.81	0.88
Jun Li et al. (2021) [31]	Noninvasive diabetics risk prediction model	–	–	–	–	–	0.914

M:male; F:female.

development process, and the overall predictive ability of the models was assessed using the degree of distinction and calibration. The main indicators of differentiation were SEN, SPE, LR+, LR−, DOR, and AUC. The calibration of the model was assessed using the Hosmer-Lemeshow goodness-of-fit test.

A total of 2 studies [23,24] containing the TAG-IT model were included in the included studies, so the effect sizes of the 2 studies were combined to assess the performance of the TAG-IT model as follows: SEN = 0.89 (0.88–0.91), $I^2 = 92.7\%$; SPE = 0.35 (0.34–0.36), $I^2 = 99.8\%$; LR+ = 1.47 (1.05–2.06), $I^2 = 99.5\%$; LR− = 0.28 (0.24–0.31); DOR = 5.24 (3.34–8.22), $I^2 = 0.0\%$. Several models were involved in the study of Hui Wang et al. [24] (TAG-IT, Guangdong 2015, Shanghai 2009). Other 12 risk prediction models involved in the literature were one. Models mostly report discriminant performance with limited reported calibration degree performance. Of all the included risk prediction models, the study by Danielle et al. [30] reported model calibrations that were assessed based on Hosmer-Lemeshow tests and calibration plots with a Hosmer-Lemeshow χ^2 of 5.50 ($p = 0.599$); Hui Wang et al. [24] reported in their study that the Hosmer-Lemeshow goodness-of-fit test with p-values between 0.56–0.96, and Liyang Wang et al. [32] reported $p > 0.5$ for the Hosmer-Lemeshow goodness-of-fit test, indicating a good calibration. 1 study [31] described the model calibration in the form of calibration plots; 4 studies [21,25–27] used only calibration descriptions and did not report specific values; 6 studies [17,20,22,23,28,29] did not describe the model calibration.

In order to show the performance of each model more clearly, we summarized the discrimination (Table 3) and calibration (Fig. 3) information of the models included in the study.

Publication bias test (Fig. 4) using Stata 14.0 yielded bias = 20.6, $p = 0.111$, implying that the funnel plot was symmetrical and there was no significant publication bias.

4. Discussion

We conducted a systematic review and quality assessment of prediabetes prediction models to explore the accuracy and clinical application value of different models. 14 studies with a total of 15 models were finally identified, and the predictive efficacy of these models was less than optimal, resulting in the absence of a definitive model for clinical practice and considerable doubt among clinical decision makers as to which models to advocate.

The included studies mostly constructed models based on cross-sectional information (8 items) [17,20–24,26,32], and heterogeneity is more likely to occur when assessing predictive performance due to increased variability in study design compared to

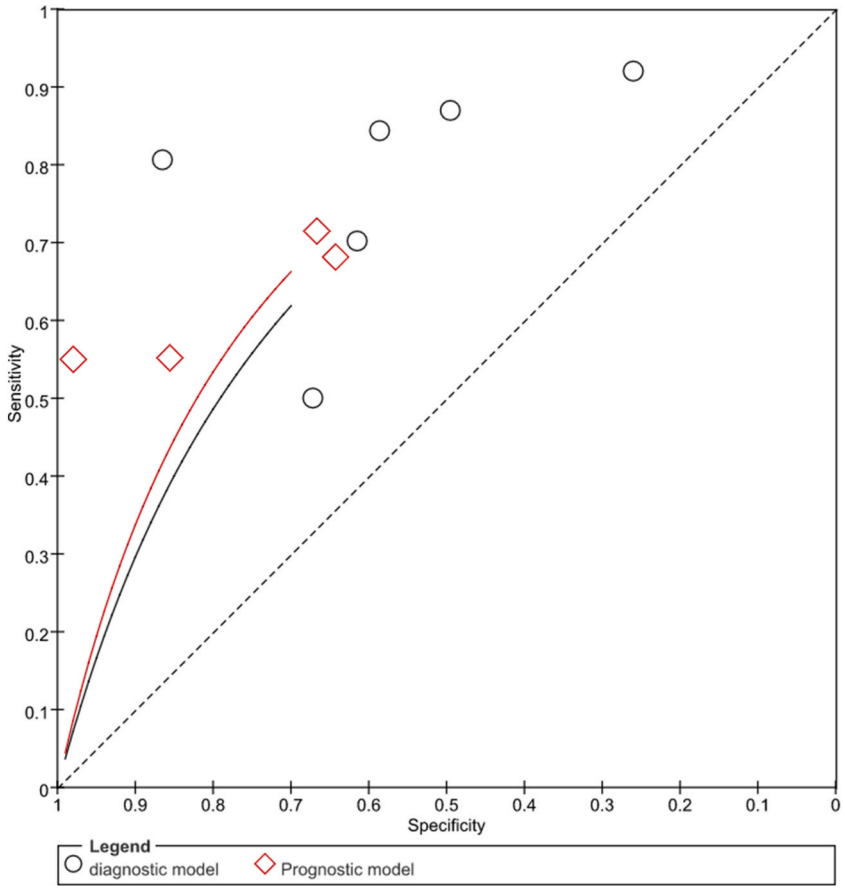


Fig. 3. AUC distribution of model validation data.

randomized trials. Prospective validation may reflect the stability, reproducibility, and external validity of the models in independent samples. Whereas only 1 study of all models used prospective data, it is not yet known whether it will improve final patient outcomes, and more prospective data information should be used in the future to minimize the risk of bias. In addition, external validation better reflects the generality and generalizability of the models, and prediction models without external validation are not recommended for use in clinical practice. In our study, most model validation was based on internal data validation using random number splitting method, which has the risk of model overfitting and affects the judgment of model stability and applicability [33]. Therefore, we believe that when constructing a new prediction model, we need to consider the relevant risk factor associations between the results and predictors, and also the random variation within the training data set. Multiple external validations should also be conducted in different populations with different age ranges, races, and genders in order to improve the generalizability, stability, and validity of the models [34].

There is an urgent need to focus on the risk of bias in predictive modeling studies, which was assessed in this study using the PROBAST tool and found a high overall risk of bias, mainly in the area of analysis. It may be related to under-reporting in the original article, improper data handling and poor methodological design [35]. In many studies, important clinical information needed to develop, validate, or use models developed by others is often omitted, leaving incomplete reporting of study data and leading to incomplete overall model evaluation [34]. It is recommended to follow the process of constructing models, while strengthening the methodological review as well as the reporting criteria; avoid using only univariate analysis when selecting predictors; factors such as the results of univariate analysis, sample size, and statistical methods need to be considered simultaneously.

Clinical prediction models were originally constructed to predict disease status or prognosis with the help of a small number of predictors that are easy to collect and inexpensive to detect. In summary, the predictors of prediabetes were found to focus on demographic information, with almost all studies using age as a predictor, and the prevalence of prediabetes tended to increase with age, consistent with the results of existing studies [33–36]. The other most commonly used predictors were family history of diabetes, history of hypertension, gender, and BMI in that order [37,38]. These predictors can be obtained through patient self-report or routine collection without laboratory tests, and have the advantage of being easily available and well applicable [39]. However, there are many predictors of prediabetes with complex mechanisms of occurrence and risk involving multiple aspects such as clinical characteristics, disease history, laboratory tests, metabolomics, and lifestyle factors [40,41], and therefore there are shortcomings in the screening of predictors. For example, important lifestyle factors associated with the risk of prediabetes such as alcohol consumption

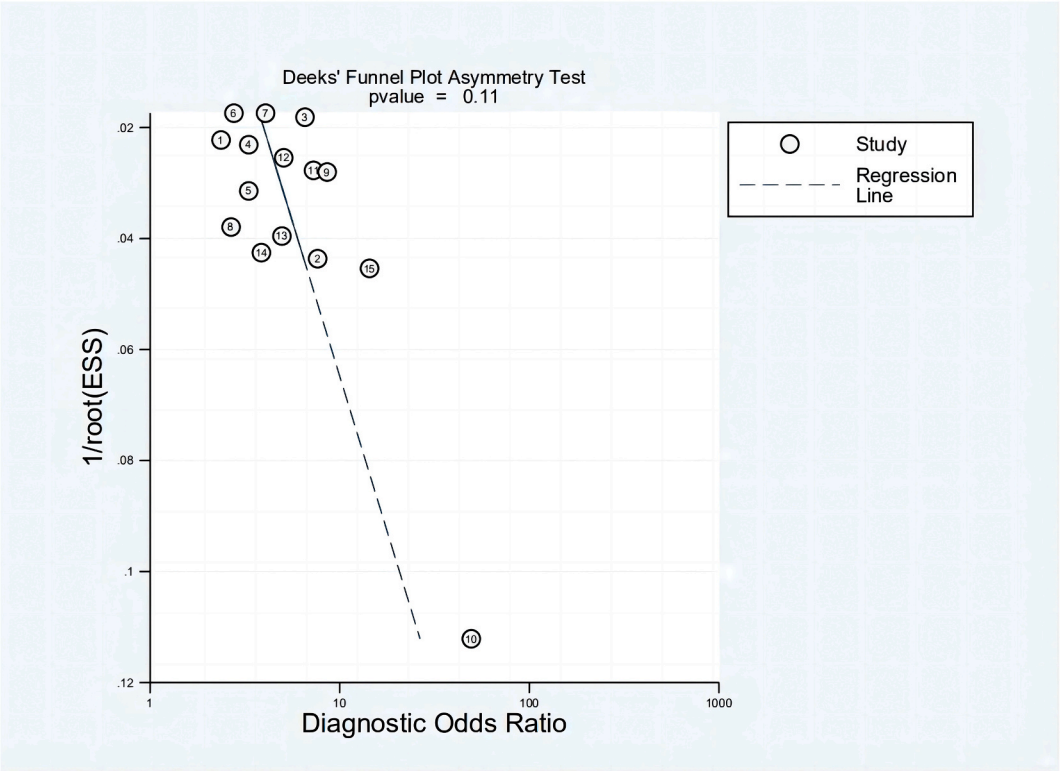


Fig. 4. Publication bias.

[42,43], sleep [44], psychosocial stress [45], sleep deprivation, and work stress were ignored; the identification of model predictors was based only on conventional clinical characteristics, without considering the close association between prediabetes and T2DM, cardiovascular disease, and other metabolic co-morbidities, and without assessing their risk for prediabetes contribution to the assessment of prediabetes risk [46]. This resulted in an incomplete selection of predictors, which affected the predictive performance of the final model, which may have contributed to the high risk of study bias. Therefore, the existing literature base, expert opinions, and clinical relevance need to be considered in variable selection, and the reliability, consistency, generalizability, applicability, and cost of the predictors should also be considered according to the relevant clinical setting [47].

The predictive performance of the models is usually assessed in terms of discrimination and calibration. After categorizing the model construction and validation data, it was found that the predictive performance varied significantly between different models or between the construction and validation data of the same model. All models were biased to use AUC to assess model accuracy, and when a model included both development and validation studies, the reporting display was not complete for the build data, and all models reported AUC for the validation set. In terms of discrimination, 11 models showed AUC >0.7 in the validation population, indicating that the models had good discrimination and were able to accurately identify the risk of developing prediabetes. For the evaluation of calibration, most of the developed prediction models did not report key information related to calibration, resulting in incomplete outcome prediction; a few studies used only textual descriptions of calibration, which does not provide a clear picture of its predictive efficacy, and the absence of key information may also affect the results of external validation of the models.

Other issues of concern are the greater willingness to repeat the process of identifying predictors and developing new models rather than validating and improving existing models for predicting prediabetes risk. These models are highly similar in terms of study design, patient population, and predictors, resulting in an overwhelming increase in the number and often poor quality of models and limited implementation in clinical practice. More importantly, there are no clear guidelines to assess the quality of the models, and studies related to the models are still in the exploratory stage. Therefore, until these prediabetes risk scoring models are widely accepted and used, external validation against relevant clinical settings to assess their calibration level and validity is the rigorous approach.

Due to the lack of mature expert consensus or guidelines, there is no unified method for conducting a systematic review of diagnostic prediction models. Therefore, our systematic review is mainly based on existing research to summarize the reliability and clinical applicability of the included models. Currently, research on systematic review of prediction models is still in the exploratory stage, indicating the need for more research on prediction models. Although some models have been externally validated, these findings indicate high heterogeneity when the models are applied to different regions or populations. Therefore, caution must be exercised when applying prediction models to clinical settings or populations different from those used to develop the original model.

Future research should encourage external validation of models in different centers and countries to ensure their generalizability. It is also recommended to adjust these models based on local settings and the predicted population to achieve better predictive

performance. Furthermore, future research is needed to determine effective integration strategies for the models in clinical practice, and to establish a unified method for systematic review of prediction models, in order to promote the clinical application of these models.

The main strengths of this review are the systematic search, screening and extraction of a wide range of data, including predictors, study results and comprehensive data presentation. However, there are some limitations, as we strictly limited the model building process of the models, and models that did not report the building process were not included, so some models recommended by guidelines, such as the Finnish Diabetes Risk Score (FINDRISC) [48,49] and the Chinese Diabetes Risk Score (CDRS) [50] were not included in this study. Second, most of the included studies were cross-sectional, leading to increased heterogeneity. Although this study failed to recommend the optimal model, it laid some foundation for future prediabetes-related studies by systematically summarizing and identifying the key predictors of prediabetes and clarifying that the next necessary stage of model research is external validation.

In conclusion, the existing prediction models dedicated to prediabetes are unsatisfactory and vary widely between models in terms of discrimination and calibration. All models show a high risk of bias, mainly due to improper handling of missing data and inadequate overfitting correction, for which more comprehensive and accurate statistical methods and clinical data based on which to base model construction are needed. Most models have not been externally validated or lack adequate reporting of model predictive performance, leaving the validity of the models still unknown. Future research should therefore focus not on developing new models but on increasing the accuracy of the models. When screening for the risk of prediabetes, in the absence of a specific prediction model, clinical practitioners can select the most appropriate risk prediction model based on the patient's actual condition, and it is recommended that external validation be performed first to help more accurately intervene earlier and delay the transition process from prediabetes to diabetes.

If in future research, a diabetes prediabetes risk prediction model with good overall performance is constructed and promoted for use in public health service agencies, it would be highly beneficial. By utilizing the large amount of collected data, these agencies can identify potential high-risk populations, better monitor disease trends, prevent and control diseases, and allocate health resources. They can also promote preventive measures through society-wide propaganda and organize relevant education and publicity activities to increase public awareness of diabetes, and strengthen its prevention and management. By doing so, they can help reduce the incidence of diabetes and prevent the development of complications, ultimately reducing the burden on the healthcare system and improving the quality of life.

5. Research in context

5.1. Evidence before this study

Prediabetes is a high-risk stage of diabetes, and early screening provides a window to prevent or delay diabetes. Previous reviews of predictive models for prediabetes have not reported a breakdown of the populations in which the models were constructed, and data from studies that developed and/or validated the models have not been separated separately. We systematically searched five databases, Cochrane, PubMed, Embase, Web of Science, and CNKI, using the keywords "Prediabetic State", "Glucose Intolerance", "impaired glucose tolerance", "Hyperglycemia", "Insulin Resistance", "impaired fasting glucose", "impaired glucose regulation". The search time frame was built to March 01, 2023, with no language or publication date restrictions, to identify relevant studies on prediabetes risk prediction models.

5.2. Added value of this study

We systematically retrieved risk prediction models specifically for the prediabetes population and performed data pooling and quality evaluation. Unlike previous studies, we took into account the heterogeneity among alternative populations and strictly excluded some models that were constructed based on diabetic populations but used to screen for people at high risk of prediabetes, because most of these models lacked a complete development process.

5.3. Implications of all the available evidence

To reduce the public health burden of diabetes, both healthcare practitioners, policy makers and the public should be aware of the importance of early screening of patients with prediabetes. Although there is a wide range of models for predicting the risk of prediabetes both domestically and internationally, the lack of an evidence base of well-designed studies has led to suboptimal predictive performance of existing models. Future research should focus not on developing new models, but on increasing the accuracy of models.

Contributors

LYJ and ZZC conceived the idea of the study. LYJ, ZZC and HYT did the literature search, selected studies, and extracted relevant information. LYJ, ZZC and HYT synthesised the data and wrote the first draft of the manuscript. LYJ, ZZC, QW, FWM, HYT, HYX, ZM and Laura Flavorta Billong revised successive drafts of the paper and approved the final version. SJT supervised the overall work and is the guarantor of the review.

Data sharing statement

Data sharing statement Data extracted for this study are presented in the supplementary material.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

References

[1] T. Tönnies, et al., Quantifying the underestimation of projected global diabetes prevalence by the International Diabetes Federation (IDF) Diabetes Atlas, BMJ Open Diabetes Res. Care 9 (1) (2021).
[2] K. Ogurtsova, et al., IDF diabetes Atlas: global estimates of undiagnosed diabetes in adults for 2021, Diabetes Res. Clin. Pract. 183 (2022), 109118.
[3] M.R. Rooney, et al., Risk of progression to diabetes among older adults with prediabetes, JAMA Intern. Med. 181 (4) (2021) 511–519.
[4] L. Lin, et al., Genetic variants relate to fasting plasma glucose, 2-hour postprandial glucose, glycosylated hemoglobin, and BMI in prediabetes, Front. Endocrinol. (Lausanne) 13 (2022), 778069.
[5] Z. Punthakee, R. Goldenberg, P. Katz, Definition, classification and diagnosis of diabetes, prediabetes and metabolic syndrome, Can. J. Diabetes 42 (Suppl 1) (2018) S10–s15.
[6] K.P. Zucatti, et al., Long-term effect of lifestyle interventions on the cardiovascular and all-cause mortality of subjects with prediabetes and type 2 diabetes: a systematic review and meta-analysis, Diabetes Care 45 (11) (2022) 2787–2795.
[7] N.W. Carris, R.R. Magness, A.J. Labovitz, Prevention of diabetes mellitus in patients with prediabetes, Am. J. Cardiol. 123 (3) (2019) 507–512.
[8] W.T. Cefalu, et al., Update and next steps for real-world translation of interventions for type 2 diabetes prevention: reflections from a diabetes care, Expert Forum. Diabetes Care 39 (7) (2016) 1186–1201.
[9] G. Li, et al., The long-term effect of lifestyle interventions to prevent diabetes in the China Da Qing Diabetes Prevention Study: a 20-year follow-up study, Lancet 371 (9626) (2008) 1783–1789.
[10] H. Sun, et al., IDF Diabetes Atlas: global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045, Diabetes Res. Clin. Pract. 183 (2022), 109119.
[11] Q. Gong, et al., Morbidity and mortality after lifestyle intervention for people with impaired glucose tolerance: 30-year results of the Da Qing Diabetes Prevention Outcome Study, Lancet Diabetes Endocrinol. 7 (6) (2019) 452–461.
[12] K.W. Davidson, et al., Screening for prediabetes and type 2 diabetes: US preventive services task Force recommendation statement, JAMA 326 (8) (2021) 736–743.
[13] K.G. Moons, et al., Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist, PLoS Med. 11 (10) (2014) e1001744.
[14] D. Moher, et al., Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement, PLoS Med. 6 (7) (2009) e1000097.
[15] T.D. Filippatos, et al., Differences in metabolic parameters and cardiovascular risk between American Diabetes Association and World Health Organization definition of impaired fasting glucose in European Caucasian subjects: a cross-sectional study, Arch. Med. Sci. 9 (5) (2013) 788–795.
[16] Standards of medical care in diabetes–2010, Diabetes Care 33 (Suppl 1) (2010) S11–S61.
[17] P. Ouyang, et al., A simple score model to assess prediabetes risk status based on the medical examination data, Can. J. Diabetes 40 (5) (2016) 419–423.
[18] K.G.M. Moons, et al., PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration, Ann. Intern. Med. 170 (1) (2019) W1–w33.
[19] R.F. Wolff, et al., PROBAST: a tool to assess the risk of bias and applicability of prediction model studies, Ann. Intern. Med. 170 (1) (2019) 51–58.
[20] W. Qing, The Prevalenceof Newly Diagnosed Diabetes Mellitus in Jiangsu Province and the Establishment and Evaluation of a Screening Model for the Pre-Diabetic Population, 2018.
[21] Fujiati II, et al., Development and validation of prediabetes risk score for predicting prediabetes among Indonesian adults in primary care: cross-sectional diagnostic study, Interv Med Appl Sci 9 (2) (2017) 76–85.
[22] R. Rajput, K. Garg, M. Rajput, Prediabetes Risk Evaluation Scoring System [PRESS]: a simplified scoring system for detecting undiagnosed prediabetes, Prim Care Diabetes 13 (1) (2019) 11–15.
[23] R.J. Koopman, et al., Tool to assess likelihood of fasting glucose impairment (TAG-IT), Ann. Fam. Med. 6 (6) (2008) 555–561.
[24] H. Wang, et al., A simple risk score for identifying individuals with impaired fasting glucose in the Southern Chinese population, Int. J. Environ. Res. Publ. Health 12 (2) (2015) 1237–1252.
[25] Y. Wang, et al., Prediction model for the onset risk of impaired fasting glucose: a 10-year longitudinal retrospective cohort health check-up study, BMC Endocr. Disord. 21 (1) (2021) 211.
[26] G. Štiglic, et al., Development of a screening tool using electronic health records for undiagnosed Type 2 diabetes mellitus and impaired fasting glucose detection in the Slovenian population, Diabet. Med. 35 (5) (2018) 640–649.
[27] M. Liu, C. Pan, M. Jin, A Chinese diabetes risk score for screening of undiagnosed diabetes and abnormal glucose tolerance, Diabetes Technol. Therapeut. 13 (5) (2011) 501–507.
[28] H. Kai, A Study on the Establishment of a Screening Model for People with Abnormal Glucose Regulation in the Huangpu District Community and the Effect of Intervention, 2012.
[29] M. Abbas, et al., Simple risk score to screen for prediabetes: a cross-sectional study from the Qatar Biobank cohort, J Diabetes Investig. 12 (6) (2021) 988–997.
[30] D.H.E. Soo, et al., Derivation and validation of the prediabetes self-assessment screening score after acute pancreatitis (PERSEUS), Dig. Liver Dis. 49 (10) (2017) 1146–1154.
[31] J. Li, et al., A tongue features fusion approach to predicting prediabetes and diabetes with machine learning, J. Biomed. Inf. 115 (2021), 103693.
[32] L. Wang, et al., IGRNet: a deep learning model for non-invasive, real-time diagnosis of prediabetes through electrocardiograms, Sensors (Basel) 20 (9) (2020).
[33] M. Pavlou, et al., How to develop a more accurate risk prediction model when there are few events, BMJ 351 (2015) h3868.
[34] L.E. Cowley, et al., Methodological standards for the development and evaluation of clinical prediction rules: a review of the literature, Diagn. Progn. Res. 3 (2019) 16.
[35] E.W. Steyerberg, Y. Vergouwe, Towards better clinical prediction models: seven steps for development and an ABCD for validation, Eur. Heart J. 35 (29) (2014) 1925–1931.
[36] K. Nakajima, K. Suwa, K. Toyama, Age-dependent changes in the association between sleep duration and impaired glucose metabolism, World J. Diabetes 8 (8) (2017) 397–406.
[37] D. LeRoith, et al., Treatment of diabetes in older adults: an endocrine society* clinical practice guideline, J. Clin. Endocrinol. Metab. 104 (5) (2019) 1520–1574.
[38] J.B. Echouffo-Tcheugui, E. Selvin, Prediabetes and what it means: the epidemiological evidence, Annu. Rev. Publ. Health 42 (2021) 59–77.
[39] A. Sinclair, et al., Diabetes and global ageing among 65-99-year-old adults: findings from the international diabetes Federation diabetes Atlas, 9(th) edition, Diabetes Res. Clin. Pract. 162 (2020), 108078.

[40] L. Chen, et al., AUSDRISK: an Australian Type 2 Diabetes Risk Assessment Tool based on demographic, lifestyle and simple anthropometric measures, Med. J. Aust. 192 (4) (2010) 197–202.
[41] W. Dong, et al., Non-laboratory-based risk assessment model for case detection of diabetes mellitus and pre-diabetes in primary care, J. Diabetes Investig. 13 (8) (2022) 1374–1386.
[42] C. Barbiellini Amidei, et al., Association between age at diabetes onset and subsequent risk of dementia, JAMA 325 (16) (2021) 1640–1649.
[43] J. Wu, et al., A prediction model for prediabetes risk in middle-Aged and elderly populations: a prospective cohort study in China, Internet J. Endocrinol. 2021 (2021), 2520806.
[44] Y. Huang, et al., Association between prediabetes and risk of cardiovascular disease and all cause mortality: systematic review and meta-analysis, BMJ 355 (2016) i5953.
[45] Z.M. Younossi, et al., The global epidemiology of NAFLD and NASH in patients with type 2 diabetes: a systematic review and meta-analysis, J. Hepatol. 71 (4) (2019) 793–801.
[46] C. Knott, S. Bell, A. Britton, Alcohol consumption and the risk of type 2 diabetes: a systematic review and dose-response meta-analysis of more than 1.9 million individuals from 38 observational studies, Diabetes Care 38 (9) (2015) 1804–1812.
[47] Consensus on the management of type 2 diabetes/prediabetes in adults, Clin. Med. Insights Endocrinol. Diabetes 13 (2020), 1179551420942232.
[48] P. López-Jaramillo, et al., Prediabetes in Colombia: expert consensus, Colomb. Méd. (Cali) 48 (4) (2017) 191–203.
[49] M. Abusaib, et al., Iraqi experts consensus on the management of type 2 diabetes/prediabetes in adults, Clin. Med. Insights Endocrinol. Diabetes 13 (2020), 1179551420942232.
[50] T. Mao, J Chen, H Guo, C Qu, C He, X Xu, G Yang, S Zhen, X Li, The Efficacy of New Chinese Diabetes Risk Score in Screening Undiagnosed Type 2 Diabetes and Prediabetes: A Community-Based Cross-Sectional Study in Eastern China, J Diabetes Res. 2020 (2020 Apr 29), 7463082.