Ageing Research Reviews 83 (2023) 101803

Contents lists available at ScienceDirect

Ageing Research Reviews

journal homepage: www.elsevier.com/locate/arr

Prediction models for depression risk among older adults: systematic review and critical appraisal

Jie Tan [a,b], Chenxinan Ma [a], Chonglin Zhu [c], Yin Wang [d], Xiaoshuang Zou [e], Han Li [f], Jiarun Li [g], Yanxuan He [h], Chenkai Wu [a,*]

[a] Global Health Research Center, Duke Kunshan University, Kunshan, Jiangsu, China
[b] School of Public Health, Wuhan University, Wuhan, Hubei, China
[c] College of Pharmacy, Southwest Medical University, Luzhou, Sichuang, China
[d] College of Management Science, Chengdu University of Technology, Chengdu, Sichuan, China
[e] College of Basic Medicine Science, Shenyang Medical College, Shenyang, Liaoning, China
[f] School of Public Health, Zunyi Medical University, Zunyi, Guizhou, China
[g] School of Basic Medicine, Guizhou Medical University, Guiyang, Guizhou, China
[h] School of Kinesiology, Shanghai University of Sport, Shanghai, China

ARTICLE INFO

Keywords:
Risk prediction model
Major depression disorder
Older adults
Systematic review

ABSTRACT

Objective: To provide an overview of prediction models for the risk of major depressive disorder (MDD) among older adults.
Methods: We conducted a systematic review combined with a meta-analysis and critical appraisal of published studies on existing geriatric depression risk models.
Results: The systematic search screened 23,378 titles and abstracts; 14 studies including 20 prediction models were included. A total of 16 predictors were selected in the final model at least twice. Age, physical health, and cognitive function were the most common predictors. Only one model was externally validated, two models were presented with a complete equation, and five models examined the calibration. We found substantial heterogeneity in predictor and outcome definitions across models; important methodological information was often missing. All models were rated at high or unclear risk of bias, primarily due to methodological limitations. The pooled C-statistics of 12 prediction models was 0.83 (95%CI=0.77–0.89).
Conclusion: The usefulness of all models remains unclear due to several methodological limitations. Future studies should focus on methodological quality and external validation of depression risk prediction models.

1. Introduction

Globally, the population aged 65 years or over was estimated to exceed 700 million in 2019, and the number is projected to double by 2050 (Economic and Affairs, 2020). One of the challenges associated with ageing is the burden of depression (Ludvigsson et al., 2015). 13.3% of older adults worldwide reported major depressive disorder (MDD) (Abdoli et al., 2022). Despite its high prevalence, depression among older adults is often underdiagnosed due to physical comorbidities and cognitive dysfunction, and hence, published statistics may underestimate the burden of geriatric depression (Allan et al., 2014; Kok and Reynolds, 2017; Malhi and Mann, 2018; VanItallie, 2005). Moreover, depression among older adults was inadequately treated (Briggs et al., 2018; Kok and Reynolds, 2017). Barriers to seeking these treatments among older adults include lack of affordable treatment for depression (Patel et al., 2004), stigma around depression (Vega et al., 2010), beliefs that depression is a normal part of ageing, and concerns regarding the effectiveness of treatments for depression (Wuthrich and Frei, 2015). Depression in old age could lead to a wide array of adverse outcomes such as falls (Xu et al., 2018), disability (Bruce, 2000), mortality (Gilman et al., 2017), and worsen the conditions of many diseases (Santomauro et al., 2021). Considering the tremendous burden of depressive disorders and the challenges of the rapidly ageing population, more attention should be given to depression prevention. Risk prediction for depression is critical for identifying at-risk groups and achieving appropriate early interventions (Halfin, 2007; Reynolds et al., 2012).

* Correspondence to: Global Health Research Center, Duke Kunshan University, Academic Building 3038 No. 8 Duke Avenue, Kunshan 215316, Jiangsu, China.
E-mail address: chenkai.wu@dukekunshan.edu.cn (C. Wu).

https://doi.org/10.1016/j.arr.2022.101803
Received 2 May 2022; Received in revised form 15 November 2022; Accepted 16 November 2022
Available online 21 November 2022

There has been an increasing number of studies attempting to develop prediction models for depression risk among older adults (Cattelani et al., 2019; Choi et al., 2018; de Man-van Ginkel et al., 2013; Hatton et al., 2019; Xu et al., 2019). However, existing models varied hugely in the source of data, methodology, definition of depression, and model performance (Andrews et al., 2017; Byeon, 2021; Kim et al., 2019). No consensus has been reached on which one is the best model for predicting the risk of depression among older adults. To our knowledge, no study summarized the published geriatric depression models or synthesized their results to generate evidence-based findings. Filling this gap will be crucial to moving forward in predicting incident depression among older adults and, therefore, allowing early implementation of prophylaxis and treatment of depression.

In this study, we conducted this systematic review and critical appraisal aiming to: (1) provide an overview of all currently available multivariable prediction models developed to predict the risk of developing MDD among older adults, (2) synthesize all available evidence, and (3) compare model performance. The findings from this study would help improve the reliability and accuracy of depression risk prediction models and provide recommendations for the choice of models.

2. Methods

This systematic review was conducted following the CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS) (Moons et al., 2014) and the guidance by Debray et al. (Debray et al., 2017) for reviews of prediction model studies. The protocol for this review was registered at the PROSPERO International Prospective Register of Systematic Reviews Website (registration number CRD42021286576).

2.1. Literature search

We systematically searched PubMed, EMBASE and PsycINFO from the inception of each database to January 20, 2022, for prediction models developed to predict the risk of developing MDD among older adults aged 60 years or older. Additionally, we manually reviewed the reference lists of included studies. The search strategy consisted of three concepts (1) prediction models, (2) older adults, and (3) major depression and related derivatives. We created the search algorithm according to the previous guidance of the search method for systematic reviews of prediction model studies (Bellou et al., 2019; Debray et al., 2017; Geersing et al., 2012). A complete list of search terms can be found in **Appendix A**.

2.2. Eligibility criteria

In medical research, prediction refers to either diagnosis (the probability of having undetected disease(s)) or prognosis (the likelihood of disease(s) developing in the future) (Bouwmeester et al., 2012; Knottnerus, 1995; Steyerberg et al., 2013; Wolff et al., 2019). This review included all primary studies that developed and/or validated multivariable (not less than two predictors) prediction models, tools, indices, or scores for estimating the individual risk of MDD among older adults. We presented a detailed description of the study population, intervention, comparator, outcome, timing, and setting (PICOTS) for this review in Table 1. Articles that reported original research (reports, reviews, and letters were excluded) with human studies and written in English were included. No restrictions were made on the prediction horizon, study setting, or included predictors.

We excluded studies examining the presence of depression among older adults using cross-sectional data and those focusing on subtypes of depression such as dysthymia and bipolar disorder. We also excluded impact studies, methodological studies, studies exploring the risk factors for depression, depression case-finding or screening studies, genetic or biomarker studies, and articles without full text.

Table 1
Criteria for study inclusion in the systematic review.

Item	Definition
Population	Older adults defined as the cut-off point of participants was 60 years or above, or the average age of participants was 60 years or above.
Intervention	Prediction model studies.
Comparator	Not applicable.
Outcome	Presence of major depression.
Timing	No limitation. Including short-term and long-term prediction horizon.
Setting	No limitation. Including general population, inpatients, and outpatients.

2.3. Screening process

Software Covidence was used to conduct the screening. Initially, pairs of two reviews (JT, CXNM, CLZ, YW, XSZ, YXH, HL, and JRL) independently applied the inclusion and exclusion criteria to screen prediction model studies on the title and abstract. Discrepancies were resolved by paired discussion or, if necessary, by a third reviewer (CW) involvement. After reaching a consensus, two reviewers (JT, CXNM) independently retrieved and screened full-text articles. We also reviewed articles' supplementary materials when needed. Moreover, we manually screened the reference list of included studies to locate potentially relevant research.

2.4. Data extraction

Two researchers (JT and CXNM) independently extracted data. A standardized data extraction form (**Appendix B**) was developed and piloted using the CHARMS checklist (Moons et al., 2014) for collecting relevant data items from eligible studies. Some studies reported more than one model, and some models were analyzed in various studies (external validation studies). The unit of this review was a model within a study unless with other specifications. For the articles that developed more than one prediction model within a study, we only extracted data from the model with the best prediction performance. If it was unclear which model was the best, data from all models would then be extracted. We extracted data including the first author, year, journal of publication, study type, data source, study setting, study population, geographical location, the definition of depression, size of study population and number of depression episodes, candidate and included predictors, and predictive performance measures, etc.

To analyze each model's predictive ability, we extracted discrimination, calibration, classification, and overall performance of included models (Collins et al., 2015). A prediction model's discrimination refers to the model's ability to differentiate people who may develop an outcome of interest from those who do not (Damen et al., 2016; Harrell et al., 1996) and is frequently measured by the concordance (C) statistics (C-statistics) (Debray et al., 2017). The C-statistics lays between 0.50 and 0.59, 0.60–0.69, 0.70–0.79, 0.80–0.89 and 0.90 or over, indicating poor discrimination, moderate discrimination, good discrimination, very good discrimination, and excellent discrimination, respectively (Hanley and McNeil, 1982; Yourman et al., 2012). Concordance refers to the area under the receiver operating characteristics curve (AUC) in logistic regression models (Debray et al., 2017). A prediction model's calibration refers to the model's agreement between predicted and observed numbers of events (Damen et al., 2016; Royston and Altman, 2013). Calibration is often measured by calibration intercept and calibration slope (Bedogni et al., 2009). It might be challenging to summarize estimates of calibration performance, given that calibration plots are usually absent, and studies often reported different forms of summary statistics in the calibration (Bouwmeester et al., 2012; Collins et al., 2013). Potential classification indicators include sensitivity, specificity, accuracy, and positive and negative predictive value. Potential overall performance was demonstrated by the value of R^2 and

Brier score (Steyerberg, 2019; Steyerberg et al., 2010).

Reviewers discussed any discrepancies in data extraction, and CW resolved the remaining conflicts. We used the Preferred Reporting Items for Systematic review and Meta-Analysis (PRISMA) (Liberati et al., 2009; Moher et al., 2009) to guide reporting of our study.

2.5. Risk of bias assessment and critical appraisal

The Prediction model Risk Of Bias ASsessment Tool (PROBAST) (Moons et al., 2019; Wolff et al., 2019) was used to assess the risk of bias (ROB) and the applicability of prediction models (**Appendix C)**. We applied PROBAST to assess the ROB and applicability of each prediction model in this review. PROBAST comprises four domains: participants, predictors, outcome, and analysis. There are 20 questions in ROB used to examine prediction model studies' performance in study design, conduct, and data analysis. Each question can be answered as "Yes", "Probably yes", "Probably no", "No", or "No information", based on the study's characteristics. The ROB would be judged as low, high, or unclear. A domain would be considered as high risk if one or more questions in that domain are answered as "No" or "Probably no"; to qualify as a low-risk domain, every component question needs to be answered as "Yes" or "Probably yes". The prediction model would be rated as having low ROB only if all domains showed low ROB. If one or more domains had a high ROB, the overall judgment would be high ROB. Only if one or more domains had an unclear ROB and all other domains were rated as low ROB, the overall judgment as unclear ROB can be achieved. The first three domains (participants, predictors, and outcome) are also adopted to rate the applicability of the prediction models (low, high, or unclear). Concerns regarding applicability were rated similarly to ROB but without questions. Assessors used their judgment to answer these

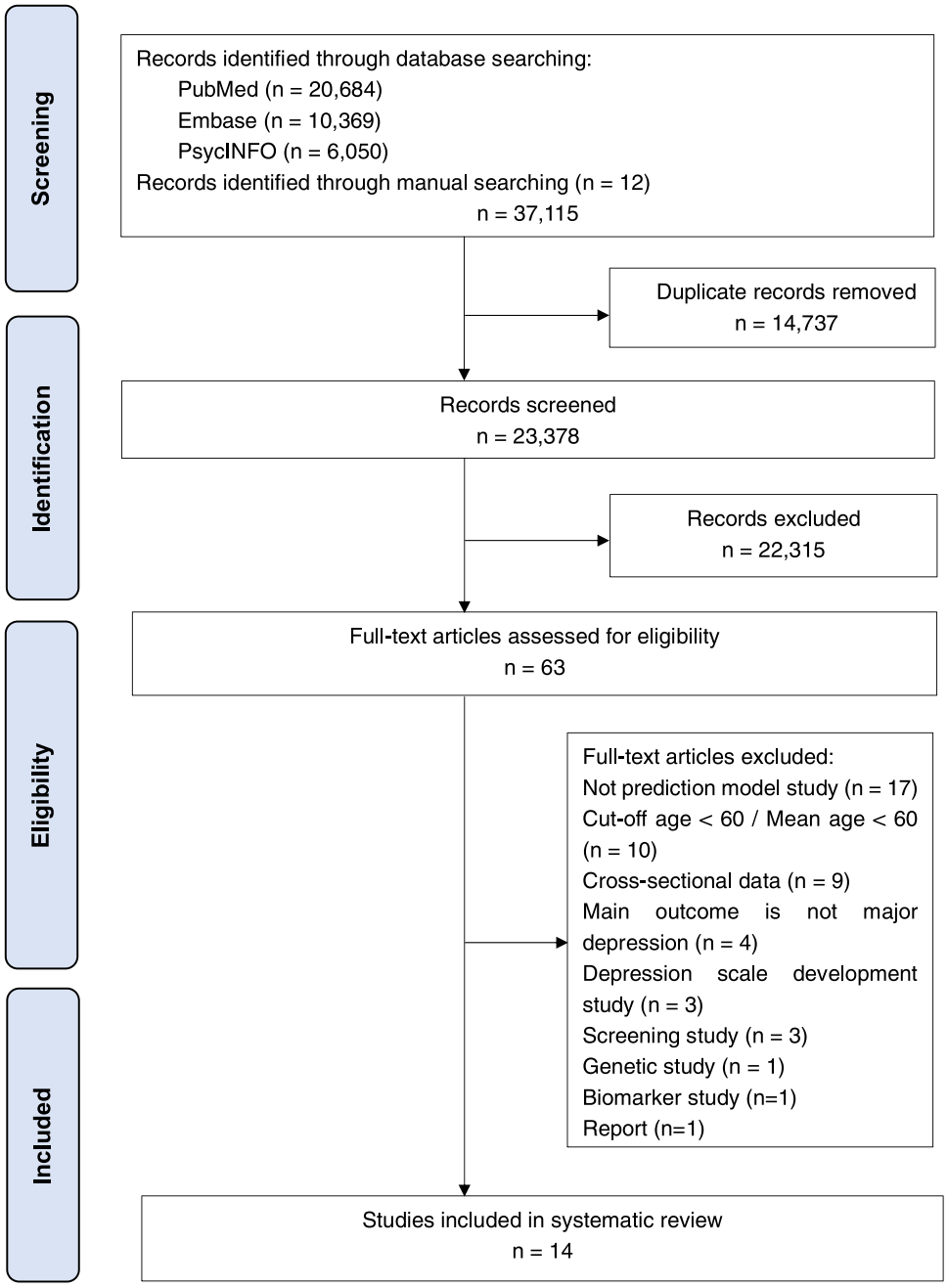

Fig. 1. Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) flowchart of literature searching and selection.

questions. Two researchers (JT, CXNM) independently assessed ROB and applicability. Discordance was eliminated by discussion.

2.6. Statistical analysis

We summarized the characteristics of all included models using medians and interquartile ranges for continuous variables, and counts and percentages for categorical variables. We conducted a random-effects meta-analysis to obtain a pooled estimation of the models' discrimination (Debray et al., 2017). Heterogeneity among studies was measured by the I^2 statistics (I^2 >50% indicates statistically significant heterogeneity), which reflects the proportion of total variation across studies due to heterogeneity (Higgins et al., 2003). The risk of publication bias was evaluated using a funnel plot, a routinely used graphical approach for detecting publication bias in systematic reviews. In a funnel plot, the models' performance and the standard error of their performance are on the x-axis and the y-axis, respectively. The data points are symmetrically distributed in the shape of a triangle without publication bias (Liu, 2011; Muka et al., 2020). All statistical analyses were performed in Stata 17.0.

3. Results

3.1. Selection process

Fig. 1 displays the PRISMA flowchart of literature searching and selection. We retrieved 37,115 potentially relevant records from PubMed (n = 20,684), EMBASE (n = 10,369), PsycINFO (n = 6050) and through manual searching (n = 12). After deleting 14,737 duplicated records, there were 23,378 original records retained for title and abstract screening; 22,315 records were excluded after title and abstract screening. Finally, 63 full texts were screened, of which 14 articles met the eligibility criteria and were included in this review. Of the 14 articles, one article included six models and the rest included two, with a total of 20 models.

3.2. Summary of findings

3.2.1. Study designs and study populations

The detailed characteristics of included studies are presented in Table 2. The majority of prediction models (n = 17, 85%) were developed from the longitudinal cohort study; two models (10%) were developed from a randomized controlled trial and one model (5%) was developed based on an appraisal of relevant literature. Seven models originated from China (35%), and two models were developed in the United States and the United Kingdom, respectively. Most models were developed in a community setting (n = 15, 75%); five models (25%) were developed in a hospital setting. For most models (n = 12, 65%), the target population is the general population, and the rest models were developed in stroke patients, Parkinson's disease patients, or patients visiting the emergency department (EDs), etc. The size of the study population used to develop the prediction models ranged from 40 to 11,704, and the percentage of depression events ranged between 4.5% and 44.1%. Although all the study populations were older adults, they differed between studies for sex, race/ethnicity, health status, and other characteristics. Except for one model developed specifically for men, the proportion of men ranged from 6% to 68.5% in other studies.

3.2.2. Definition of depression outcome

Although we only included articles using the presence of MDD as outcome and excluded those focusing on subtypes of depression such as dysthymia and bipolar disorder, we still observed a high degree of heterogeneity in the operational definition of depression. Eight operational definitions were used in the 20 models. Although many prediction models (n = 6, 30%) use the Geriatric Depression Scale-15 (GDS-15) as the criteria to screen depression, the cut-off points ranged from 5 to 7

Table 2
Characteristics of reviewed prediction models for depression risk among older adults.

	No (%) of models* (N = 20) or median (interquartile range)
Study type	
Developed and internally validated study	17 (85)
Developed and externally validated study	1 (5)
Developed study without validation	2 (10)
Data source	
Longitudinal cohort study	17 (85)
Randomized controlled trial	2 (10)
Other	1 (5)
Location	
China	7 (35)
United States	2 (10)
United Kingdom	2 (10)
Canada	1 (5)
Netherlands	1 (5)
Australia	1 (5)
Korea	1 (5)
Singapore	1 (5)
Japan	1 (5)
More than one country	3 (15)
Setting	
Community setting	15 (75)
Inpatients	5 (25)
Population	
General population	12 (60)
Stoke patients	2 (10)
Parkinson's disease patients	2 (10)
Sub-threshold depression patients	2 (10)
Patients visiting the emergency department	1 (5)
Family caregivers	1 (5)
Definition of depression outcome	
Geriatric Depression Scale-15 (GDS-15)	6 (30)
Diagnostic and Statistical Manual of Mental Disorders-IV (DSM-5)	6 (30)
Center for Epidemiologic Studies Depression Scale (CES-D)	3 (15)
Patient Health Questionnaire-9 (PHQ-9)	1 (5)
Hamilton Rating Scale for Depression (HRSD)	1 (5)
Hospital Anxiety and Depression Scale (HADS-D)	1 (5)
Composite International Diagnostic Interview (CIDI)	1 (5)
More than one method	1 (5)
Handling of missing data	
Imputation	11 (55)
Exclusion	6 (30)
Not reported	3 (15)
Handling of continuous data	
Continuous	10 (50)
Categorical / dichotomous	4 (20)
Not reported	6 (30)
Shrinkage methods	
LASSO	7 (35)
Not reported	13 (65)
Modelling method	
Machine learning	11 (55)
Logistic regression model	7 (35)
Linear regression model	2 (10)
Validation	
Random split	9 (45)
Bootstrapping	4 (20)
Cross validation	3 (15)
Combination of methods	1 (5)
External validation	1 (5)
None	2 (10)
Model discrimination	
C-statistics / AUC	19 (95)
None	1 (5)
Model calibration	
Hosmer-Lemeshow test	1 (5)
Calibration plot	1 (5)

(continued on next page)

Table 2 (*continued*)

	No (%) of models* (N = 20) or median (interquartile range)
U-statistics	1 (5)
More than one method	2 (10)
None	15 (75)
Model presentation	
Sum score	4 (20)
Full equation	2 (10)
More than one method	2 (10)
Nomogram	1 (5)
Decision tree	1 (5)
Risk chart	1 (5)
None	9 (45)
No of participants	1538 (312–1665)
No of depression events	228 (50–289)
Proportion of depression events	18.8% (12.7%–19.2%)
No of males	728 (188–728)
Proportion of males	47.3% (40.0%–66.0%)
No of missing data	94 (7–120)
Proportion of missing data	17.5% (7.5%–27.1%)
Mean age of participants	75.0 (67.4–83.7)
Sensitivity	0.74 (0.62–0.81)
Specificity	0.68 (0.56–0.83)

* : Analysis unit is a model within a study

(total score: 15) (Greenberg, 2012). Diagnostic and Statistical Manual of Mental Disorders-IV (DSM-5) was used as a depression screening tool for six models (30%). Center for Epidemiologic Studies Depression Scale (CES-D) was used for three models (15%), with a cut-point of 16 used for 20-item CES-D (total score: 60) for two models and a cut-point of 3 used for 8-item CES-D for another model (total score: 24) (Radloff, 1977).

3.2.3. Predictors

More than 300 candidate predictors were considered for model development. These candidate predictors can be divided into four major categories: demographic characteristics, health-related risk factors, chronic illness-related risk factors, and socioeconomic status-related risk factors. Sixteen predictors were selected at least twice: age, physical health, cognitive function, gender, self-reported health, smoking, marital status, diabetes, cancer, history of depression, education, social support, sleep quality, alcohol consumption, medication use and dietary information (Table 3). Age (n = 11, 55%), physical health (n = 10, 50%) and cognitive function (n = 10, 50%) were the most common predictors.

3.2.4. Handling of missing data and continuous variables

Most models (n = 17, 85%) reported missing values in the development datasets. The percentage of missing data ranged from 1.8% to 26.2%. There were 11 models (55%) using multiple imputation methods to impute missing data, and six models (30%) using the complete case analysis method to exclude individuals with any missing data. Continuous predictors were retained in 14 models (70%); four models converted continuous variables to categorical variables and ten models dealt with them as continuous variables. The remaining six models (30%) did not explicitly state how continuous variables were modeled.

Table 3
Predictors identified not less than twice in the prediction models for depression risk among older adults.

Risk predictors	Frequency identified, n (%)* (N = 20)
Age	11 (55)
Physical health	10 (50)
Cognitive function	10 (50)
Gender	9 (45)
Self-reported health	9 (45)
Smoking	8 (40)
Marital status	7 (35)
Diabetes	7 (35)
Cancer	7 (35)
History of depression	4 (20)
Education	3 (15)
Social support	3 (15)
Sleep quality	3 (15)
Alcohol consumption	2 (10)
Medication use	2 (10)
Dietary information	2 (10)

* : Analysis unit is a model within a study

3.2.5. Modelling method and prediction horizon

Over half of the prediction models were developed using machine learning techniques (n = 11, 55%). Seven models (35%) were developed using multivariable logistic regression, and two models (10%) used linear regression. The prediction horizon ranged from one week to 12 years; 11 studies (55%) predicted depression for no more than three years.

3.2.6. Predictive performance and validation

Only one model (5%) was externally validated, 17 models (85%) were internally validated, and two models (10%) were not validated. Of the internal validation models, nine models (45%) used the random split method, four models (20%) used bootstrapping method, three models (15%) used the cross-validation method, and one model (5%) used a combination of methods. As for models' performance, nearly all models (n = 19, 95%) reported discrimination, remaining one model (5%) did not provide any information about its discriminative performance. Thirteen models (65%) reported a C-statistics above 0.70; six models (30%) reported a C-statistics between 0.50 and 0.69. Calibration was only reported in five models (25%); all reported that the models were well-calibrated. The sensitivity of models ranged from 43% to 92% and specificity ranged from 37% to 94%.

3.2.7. Model presentation

Only two models reported the complete regression equation, which includes regression coefficients and intercepts or baseline hazards. Four models (20%) were presented as sum scores, and two models (10%) were presented using more than one method. Three models (15%) were presented as nomograms, decision trees, or risk charts. For the remaining nine models (45%), the calculation of individual risks can't achieve because of the absence of sufficient information.

3.3. Risk of bias and applicability assessment

We used PROBAST to assess the ROB (Fig. 2.1) and applicability (Fig. 2.2) of all 20 included models. A total of 18 models were rated as high (n = 17, 85%) or unclear (n = 1, 5%) ROB. Two models (10%) with a low ROB were developed without any external validation, and the development was based on a relatively small dataset. We therefore downgraded them to a high ROB. Three models (15%) had a high ROB in the participants' domain, suggesting the study participants couldn't represent the models' targeted populations. Ambiguous reporting on the participants' eligibility criteria caused an unclear ROB in one model (5%), and the remaining 16 (80%) models had a low ROB in the participants' domain. No model was rated as having a high ROB in the predictor domain, whereas eight models (40%) were rated as having an unclear ROB in the predictor domain, indicating the predictors might not be available at the time when the models are supposed to be used, not well-defined, or might be influenced by the outcome measurements. The remaining 12 models (60%) were rated as having a low ROB in the predictor domain. Three models (15%) were rated as having a high ROB in the outcome domain as their predictors were not excluded from the outcome definition, or concerns about bias induced by the outcome measurements. The other 17 models (85%) were rated as having a low ROB, suggesting that the depression outcome was determined and assessed appropriately. A majority of models (n = 17, 85%) were at high ROB in the analysis domain. The remaining three models (15%) had a

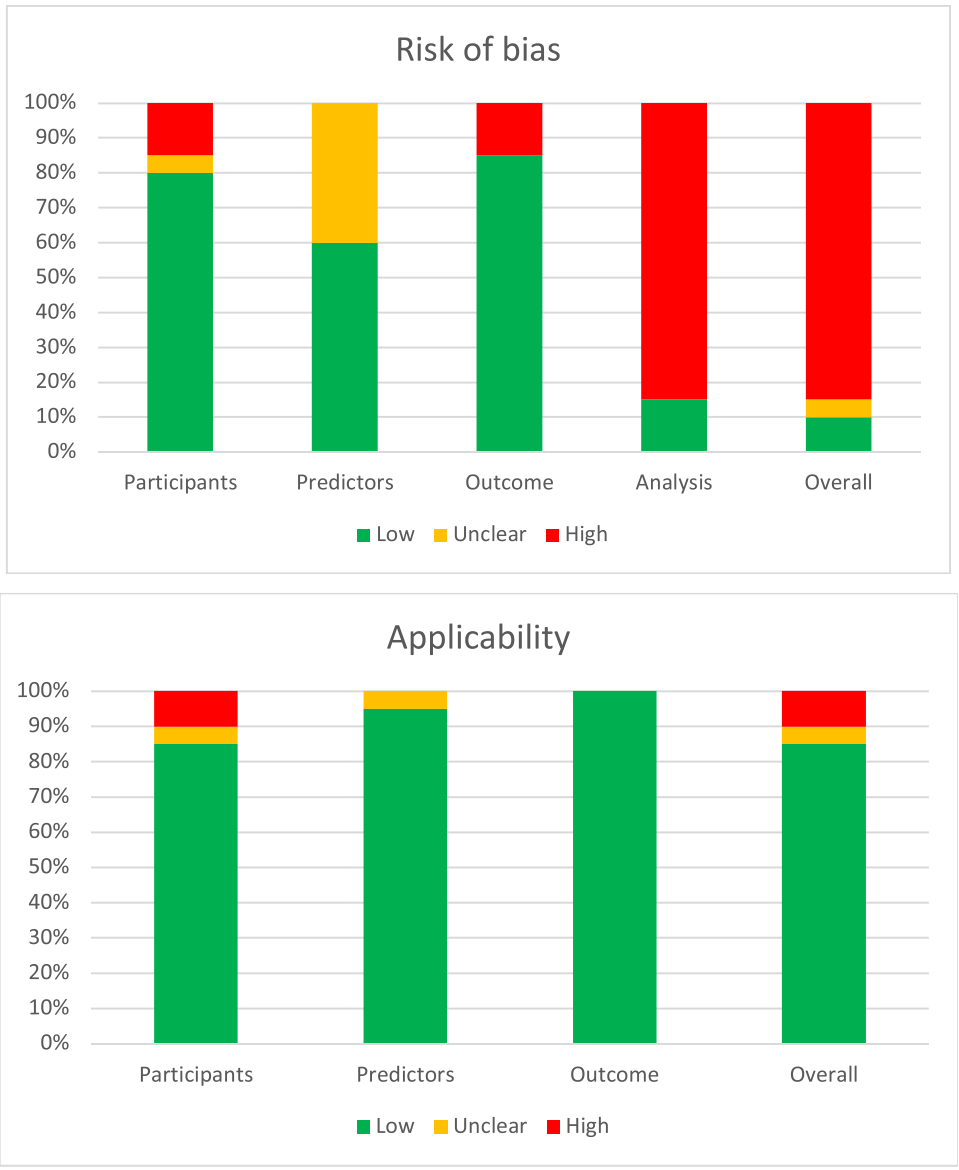

Fig. 2.1. Prediction model Risk Of Bias ASsessment Tool (PROBAST) risk of bias assessment for all included models (N = 20).

low ROB in the analysis domain.

As for applicability, two models (10%) were rated as having high concern, one (5%) was of unclear concern, and 17 models (85%) were rated as low concern. Three models were rated as high concern (n = 2, 10%) or unclear concern (n = 1,5%) in the participants' domain, which implies that participants or settings in these prediction model studies may differ from the targeted population or setting defined in the review question. Most models (n = 17, 85%) had a low applicability concern for the participants' domain. In the predictors' domain, one model (5%) was of unclear concern, indicating predictors' definition, assessment or timing might not match the review question. Nearly all models (n = 19, 95%) were of low concern in the predictors' domain. All models were rated as low concern for the outcomes domain, suggesting that the MDD outcome's definition, timing, and determination matched the review question.

Overall, most models (n = 17, 85%) were at a high ROB, but most models (n = 17, 85%) were at a low concern of applicability, which indicated issues of methodological quality limit the performance and application of the included models (Moons et al., 2019; Wolff et al., 2019; Wynants et al., 2020).

3.4. Meta-analysis of validation models included in the review

We conducted a meta-analysis that included 12 validation models with reported AUC and its 95% confidence interval (95%CI) to get the pooled discrimination. The rest eight models were excluded from meta-analysis due to missing data. We used the random-effects model to pool all AUCs (Fig. 3.) The pooled AUC was 0.83 (95%CI: 0.77, 0.89). The I^2 value was 99.4% ($p < 0.001$), suggesting a high heterogeneity. A funnel plot was used to evaluate publication bias (Fig. 4). The scatter plots were distributed approximately symmetrically, suggesting there was no obvious publication bias of the included prediction models.

4. Discussion

4.1. Principal findings and interpretations

In this systematic review of prediction models to predict the risk of developing MDD among older adults, we have identified and critically appraised 20 models from 14 studies. The included models were developed in either community settings or clinical settings, focusing on the general population, diseased population, and family caregivers. Over

Fig. 2.2. Prediction model Risk Of Bias ASsessment Tool (PROBAST) applicability assessment based on three models for all included models (N = 20).

half of the models were developed by machine learning technique, and the rest were developed by logistic regression method and linear regression method. Most studies developed new models with internal validation, but only one model conducted external validation, and calibration was barely evaluated. Although all models showed moderate to excellent predictive performance, they all had a high or unclear ROB as assessed by the PROBAST. Four models (20%) had a low ROB in all domains except analysis, which suggests that these prediction models performed well in other aspects such as study design and data collection, and a better statistical analysis could have avoided this issue (Moons et al., 2019; Wynants et al., 2020). We found one model that was generally high quality, constructed by large datasets, having a good predictive performance, and been rated as low ROB in three domains but judged as unclear ROB, due to insufficient reporting of one question in the predictors' domain (Choi et al., 2018). A high ROB indicates that the performance of prediction models among independent participants would be lower than that claimed in the original study (Wynants et al., 2020). Therefore, the pooled C-statistics above 0.80 suggests good discrimination (Hanley and McNeil, 1982), which is probably overestimated. Unfortunately, despite prediction models holding the prospect of improving the prevention and intervention of depression among older adults, we still can't recommend the widespread use of any prediction models in practice at this time given the insufficient evidence. But we admitted that some of the included prediction models were of better quality and promising (Gu et al., 2020; Su et al., 2021).

4.2. *Challenges and opportunities*

This systematic review found several methodological limitations in the development and/or validation of included prediction models, which is also shown by the ROB evaluation. Firstly, although most models were internally validated, only one model was externally validated. Prediction models typically perform better in the derivation dataset than in an independent population (Altman et al., 2009; Moons, Kengne, Grobbee et al., 2012). Before applying prediction models in practice, we must externally validate the models in independent and different populations, and also compare the predictive performance of every model to identify the models with the best discrimination and calibration (Bellou et al., 2019). Based on our review, because there are not enough external validation studies and no direct head-to-head comparison studies, we still cannot compare the predictive performance of these models or choose the best one among existing prediction

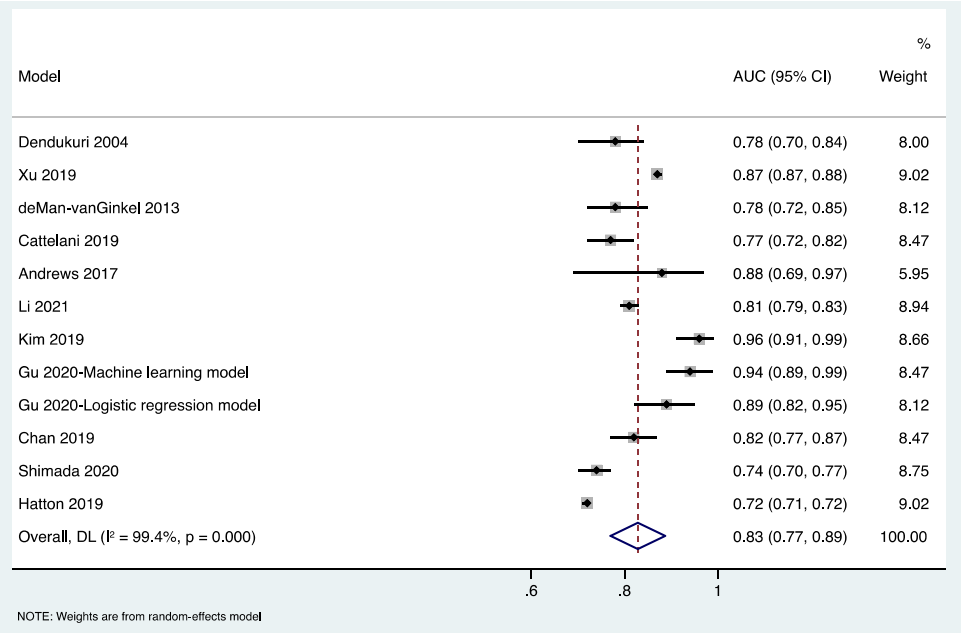

Fig. 3. Forest plot of the random effects meta-analysis of pooled AUC estimates for 12 validation models.

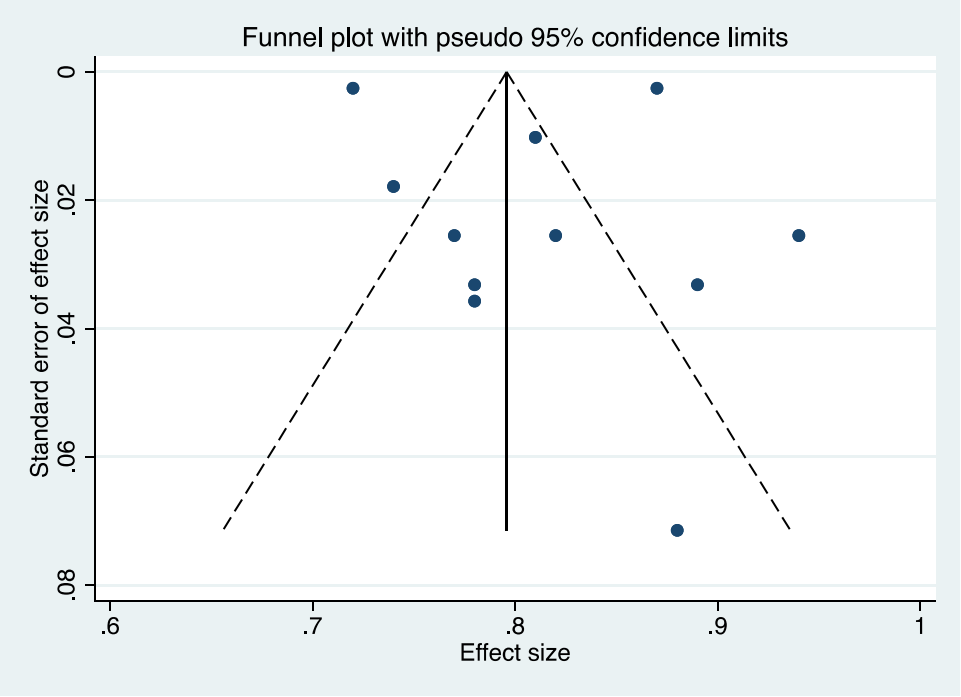

Fig. 4. Funnel plot to assess publication bias.

models, not to mention deciding on which is the best model to be advocated for use in research or clinical settings (Damen et al., 2016).

Secondly, in about half of the included models, missing values were either excluded or without mention of its handling methods. If missing data are not imputed, they usually bias effect estimation and reduce the discrimination ability of prediction models. This is because missing data can distort the prediction models' performance if they are connected to other variables (Janssen et al., 2010; Moons, Kengne, Grobbee et al., 2012; Rubin, 1976). Meanwhile, in half of the prediction models, continuous predictors were dichotomized or categorized. However, it has been proven that categorizing continuous predictors into two or more categories results in lower prediction performance than analyzing predictors on a continuous scale, due to the loss of information and declination of analysis power (Collins et al., 2016; Moons et al., 2009; Royston et al., 2006).

Also, we have noticed that various predictor selection procedures and statistical significance levels for choosing a predictor for the final model were used. Although the most appropriate method to select candidate predictors remains unknown, univariate analysis was commonly used in these included models known to yield overfitting models (Bellou et al., 2019). To overcome overfitting, studies should apply shrinkage, a technique to reduce overfitting by re-adjusting the

regression coefficients (Moons, Kengne, Woodward et al., 2012; Steyerberg and Vergouwe, 2014). Only a few prediction models included in our review performed shrinkage.

Furthermore, the performance of geriatric depression risk models can be assessed by discrimination, calibration, and classification measures (Collins et al., 2015). However, these measures vary depending on participants' demographic characteristics, study sample size, depression prevalence, depression severity, candidate predictors, diagnostic criteria, definitions of depression and length of follow-up (Hou et al., 2019). Thus, it is difficult to compare the predictive performance of the prediction models.

Besides, the cost-effectiveness of using the geriatric depression risk models in practice should also be considered. Generally, the models with high-cost predictors have higher predictive accuracy than those without predictors (Pashayan et al., 2018). But the feasibility and cost constraints would limit the usage of the models, particularly in the primary care setting (Blum et al., 2020). The most applicable model for primary care settings should adopt predictors already available or easily obtainable (Hou et al., 2019). In this review, the machine-learning models included a high volume of predictors; however, some predictors are not easily obtainable, which might limit the practicality of these models.

Finally, incomplete reporting of studies is emphasized as a severe waste of research sources as it hinders future efforts for validation, update, recalibration and guiding clinical practice (Collins et al., 2015). For example, most included models were partially presented, with only 10% reporting a complete regression equation. It is advised to present the final prediction model as an original regression equation with intercept or baseline hazard (Bonnett et al., 2019). Using simplified risk calculators or charts can make the risk prediction models more accessible, but the process of simplification must be weighed against the information loss (Moons, Kengne, Woodward et al., 2012; Steyerberg and Vergouwe, 2014; Sullivan et al., 2004). Moreover, only a quarter of included models reported calibration, which is the accuracy of absolute risk estimates (Royston and Altman, 2013). To improve the reporting of prediction modeling studies, the Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) statement should be strictly followed (Collins et al., 2015).

4.3. Strengths and limitations

The main strengths of the current review include the thorough literature search, deliberate screening, and standardized data extraction on critical characteristics of geriatric MDD risk prediction models based on CHARMS. Another significant advantage is the ROB assessment and critical appraisal of prediction models using the PROBAST. We also did a meta-analysis of AUCs to get an overall estimate of the included models' performance.

However, this review also has some limitations. First, we can't do a meta-analysis of calibration measurements for prediction models due to poor reporting of calibration among prediction models. Also, there is a high heterogeneity between included studies. As validation studies tend to vary in study design, execution and thus case mix, variation between them was unlikely to happen by chance alone (Debray et al., 2015; Vergouwe et al., 2010). Given this reason, heterogeneity should typically be accepted in the meta-analysis of prediction models (Debray et al., 2017). Though significant heterogeneity exists, we could still get meaningful results from the meta-analysis (Riley et al., 2016). Then, we excluded articles not written in English, which might lead to certain underestimation of the number of models, and affected the representativeness of geographical distribution. Furthermore, research on the topic is relatively scarce.

4.4. Implications for policy and future research

The ultimate goal of systematic reviews is to aid in guiding evidence-based health decision-making (Damen et al., 2016). In this study, we try to determine the best model(s) to be advocated or used to predict the risk of developing MDD among older adults across various settings and countries. However, we are unable to recommend any model in practice right now for the following reasons. First, all reviewed prediction models had an unclear or high ROB. Then, the included newly developed models lack evidence from independent external validations. Moreover, it is pretty challenging to compare the predictive performance of the prediction models given the large heterogeneity among the included models. Furthermore, the nonstandard analysis methods, the poor reporting of most models, and the dearth of cost-effectiveness analysis all result in failing to choose the best model.

Based on the above-mentioned methodological defects, we put forward the following recommendations to improve the quality of prediction modeling studies. Firstly, models needed to be externally validated in independent and different populations. Secondly, imputation techniques should be applied when data are missing and continuous predictors should not be dichotomized. Thirdly, prediction models should adjust for overfitting not only by internal validation but also should use shrinkage techniques. Fourthly, researchers should carry out more studies to reduce the heterogeneity among models and allow for a head-to-head comparison of the models. Then, the cost-effectiveness analysis of newly developed models should be conducted. Finally, the TRIPOD statement should be adopted to complete the reporting of prediction modeling studies.

5. Conclusions

We found 20 prediction models to predict the risk of developing MDD among older adults, and they all reported moderate to excellent discrimination. Nevertheless, all models had a high or unclear ROB due to methodological defects. Therefore, their reported predictive performance is probably overestimated and can't represent the target population (Wolff et al., 2019). We cannot advocate any of the existing prediction models to be put into practice at present. Subsequent prediction modeling studies for geriatric depression risk should primarily deal with the problems raised, and also should follow the methodological guidance of prediction modeling studies, since unreliable predictions may lead to more harm than benefit when guiding medical practice (Wynants et al., 2020). Meanwhile, future studies in this area should also concentrate on assessing model applicability and generalizability and conducting a cost-effectiveness analysis to estimate the most effective model for use in practice.

CRediT authorship contribution statement

CW and JT conceived the study. **JT** designed the systematic review and ran the literature search. **JT, CXNM, CLZ, YW, XSZ, HL, JRL, and YXH** performed the screening. **JT and CXNM** performed the data extraction and risk of bias assessment. **JT** conducted all statistical analyses in consultation with **CW**. **JT and CXNM** wrote the manuscript, and **CW** critically revised the manuscript.

Competing interests and funding

The authors report no declarations of interest; this paper is supported by Jiangsu Provincial Association for Maternal and Child Health Studies (grant number: JSFY202107).

Data Availability

No data was used for the research described in the article.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the

online version at doi:10.1016/j.arr.2022.101803.

References

Abdoli, N., Salari, N., Darvishi, N., Jafarpour, S., Solaymani, M., Mohammadi, M., Shohaimi, S., 2022. The global prevalence of major depressive disorder (MDD) among the elderly: a systematic review and meta-analysis. Neurosci. Biobehav Rev. 132, 1067–1073. https://doi.org/10.1016/j.neubiorev.2021.10.041.

Allan, C.E., Valkanova, V., Ebmeier, K.P., 2014. Depression in older people is underdiagnosed. *Practitioner*, *258*(1771), 19-22 12–13.

Altman, D.G., Vergouwe, Y., Royston, P., Moons, K.G., 2009. Prognosis and prognostic research: validating a prognostic model. Bmj 338, b605. https://doi.org/10.1136/bmj.b605.

Andrews, J.A., Harrison, R.F., Brown, L.J.E., MacLean, L.M., Hwang, F., Smith, T., Williams, E.A., Timon, C., Adlam, T., Khadra, H., Astell, A.J., 2017. Using the NANA toolkit at home to predict older adults' future depression. J. Affect Disord. 213, 187–190. https://doi.org/10.1016/j.jad.2017.02.019.

Bellou, V., Belbasis, L., Konstantinidis, A.K., Tzoulaki, I., Evangelou, E., 2019. Prognostic models for outcome prediction in patients with chronic obstructive pulmonary disease: systematic review and critical appraisal. Bmj 367, l5358. https://doi.org/10.1136/bmj.l5358.

Blum, M.R., Øien, H., Carmichael, H.L., Heidenreich, P., Owens, D.K., Goldhaber-Fiebert, J.D., 2020. Cost-effectiveness of transitional care services after hospitalization with heart failure. Ann. Intern Med 172 (4), 248–257. https://doi.org/10.7326/m19-1980.

Bonnett, L.J., Snell, K.I.E., Collins, G.S., Riley, R.D., 2019. Guide to presenting clinical prediction models for use in clinical settings. Bmj 365, l737. https://doi.org/10.1136/bmj.l737.

Bouwmeester, W., Zuithoff, N.P., Mallett, S., Geerlings, M.I., Vergouwe, Y., Steyerberg, E. W., Altman, D.G., Moons, K.G., 2012. Reporting and methods in clinical prediction research: a systematic review. PLoS Med 9 (5), 1–12. https://doi.org/10.1371/journal.pmed.1001221.

Briggs, R., Tobin, K., Kenny, R.A., Kennelly, S.P., 2018. What is the prevalence of untreated depression and death ideation in older people? Data from the Irish Longitudinal Study on Aging. Int Psychogeriatr. 30 (9), 1393–1401. https://doi.org/10.1017/s104161021700299x.

Bruce, M.L., 2000. Depression and Disability. In: Williamson, G.M., Shaffer, D.R., Parmelee, P.A. (Eds.), Physical Illness and Depression in Older Adults: A Handbook of Theory, Research, and Practice. Springer, US, pp. 11–29. https://doi.org/10.1007/0-306-47178-7_2.

Byeon, H., 2021. Development of a Nomogram for Predicting Depression in the Elderly Using Patient Health Questionnaire-9 among a Nationwide Sample of Korean Elderly. J. Pers. Med 11 (7). https://doi.org/10.3390/jpm11070645.

Cattelani, L., Murri, M.B., Chesani, F., Chiari, L., Bandinelli, S., Palumbo, P., 2019. Risk Prediction Model for Late Life Depression: Development and Validation on Three Large European Datasets. IEEE J. Biomed. Health Inf. 23 (5), 2196–2204. https://doi.org/10.1109/jbhi.2018.2884079.

Choi, J., Choi, J., Choi, W.J., 2018. Predicting Depression Among Community Residing Older Adults: A Use of Machine Learning Approch. Stud. Health Technol. Inf. *250*–265.

Collins, G.S., Omar, O., Shanyinde, M., Yu, L.M., 2013. A systematic review finds prediction models for chronic kidney disease were poorly reported and often developed using inappropriate methods. J. Clin. Epidemiol. 66 (3), 268–277. https://doi.org/10.1016/j.jclinepi.2012.06.020.

Collins, G.S., Reitsma, J.B., Altman, D.G., Moons, K.G., 2015. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. Bmj 350, g7594. https://doi.org/10.1136/bmj.g7594.

Collins, G.S., Ogundimu, E.O., Cook, J.A., Manach, Y.L., Altman, D.G., 2016. Quantifying the impact of different approaches for handling continuous predictors on the performance of a prognostic model. Stat. Med 35 (23), 4124–4135. https://doi.org/10.1002/sim.6986.

Damen, J.A., Hooft, L., Schuit, E., Debray, T.P., Collins, G.S., Tzoulaki, I., Lassale, C.M., Siontis, G.C., Chiocchia, V., Roberts, C., Schlüssel, M.M., Gerry, S., Black, J.A., Heus, P., van der Schouw, Y.T., Peelen, L.M., Moons, K.G., 2016. Prediction models for cardiovascular disease risk in the general population: systematic review. Bmj 353, i2416. https://doi.org/10.1136/bmj.i2416.

Debray, T.P., Vergouwe, Y., Koffijberg, H., Nieboer, D., Steyerberg, E.W., Moons, K.G., 2015. A new framework to enhance the interpretation of external validation studies of clinical prediction models. J. Clin. Epidemiol. 68 (3), 279–289. https://doi.org/10.1016/j.jclinepi.2014.06.018.

Debray, T.P., Damen, J.A., Snell, K.I., Ensor, J., Hooft, L., Reitsma, J.B., Riley, R.D., Moons, K.G., 2017. A guide to systematic review and meta-analysis of prediction model performance. Bmj 356, i6460. https://doi.org/10.1136/bmj.i6460.

Economic, U.N.D.O., & Affairs, S. (2020). World Population Ageing 2019. United Nations. https://doi.org/10.18356/6a8968ef-en.

Geersing, G.J., Bouwmeester, W., Zuithoff, P., Spijker, R., Leeflang, M., Moons, K.G., 2012. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. PloS One 7 (2), e32844. https://doi.org/10.1371/journal.pone.0032844.

Gilman, S.E., Sucha, E., Kingsbury, M., Horton, N.J., Murphy, J.M., Colman, I., 2017. Depression and mortality in a longitudinal study: 1952-2011. Cmaj 189 (42), 1304–1310. https://doi.org/10.1503/cmaj.170125.

Greenberg, S.A., 2012. The geriatric depression scale (GDS). Best. Pract. Nurs. Care Older Adults 4 (1), 1–2.

Gu, S.C., Zhou, J., Yuan, C.X., Ye, Q., 2020. Personalized prediction of depression in patients with newly diagnosed Parkinson's disease: a prospective cohort study. J. Affect Disord. 268, 118–126. https://doi.org/10.1016/j.jad.2020.02.046.

Halfin, A., 2007. Depression: the benefits of early and appropriate treatment. Am. J. Manag Care 13 (4 Suppl), S92–S97.

Hanley, J.A., McNeil, B.J., 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143 (1), 29–36. https://doi.org/10.1148/radiology.143.1.7063747.

Harrell Jr, F.E., Lee, K.L., Mark, D.B., 1996. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat. Med 15 (4), 361–387. https://doi.org/10.1002/(SICI)1097-0258(19960229)15:4<361::AID-SIM168>3.0.CO;2-4.

Hatton, C.M., Paton, L.W., McMillan, D., Cussens, J., Gilbody, S., Tiffin, P.A., 2019. Predicting persistent depressive symptoms in older adults: a machine learning approach to personalised mental healthcare. J. Affect Disord. 246, 857–860. https://doi.org/10.1016/j.jad.2018.12.095.

Higgins, J.P., Thompson, S.G., Deeks, J.J., Altman, D.G., 2003. Measuring inconsistency in meta-analyses. Bmj 327 (7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557.

Hou, X.H., Feng, L., Zhang, C., Cao, X.P., Tan, L., Yu, J.T., 2019. Models for predicting risk of dementia: a systematic review. J. Neurol. Neurosurg. Psychiatry 90 (4), 373–379. https://doi.org/10.1136/jnnp-2018-318212.

Janssen, K.J., Donders, A.R., Harrell Jr., F.E., Vergouwe, Y., Chen, Q., Grobbee, D.E., Moons, K.G., 2010. Missing covariate data in medical research: to impute is better than to ignore. J. Clin. Epidemiol. 63 (7), 721–727. https://doi.org/10.1016/j.jclinepi.2009.12.008.

Kim, H., Lee, S., Lee, S., Hong, S., Kang, H., Kim, N., 2019. Depression prediction by using ecological momentary assessment, actiwatch data, and machine learning: observational study on older adults living alone. JMIR Mhealth Uhealth 7 (10), e14149. https://doi.org/10.2196/14149.

Knottnerus, J.A., 1995. Diagnostic prediction rules: principles, requirements and pitfalls. Prim. Care 22 (2), 341–363.

Kok, R.M., Reynolds 3rd, C.F., 2017. Management of depression in older adults: a review. Jama 317 (20), 2114–2122. https://doi.org/10.1001/jama.2017.5706.

Liberati, A., Altman, D.G., Tetzlaff, J., Mulrow, C., Gøtzsche, P.C., Ioannidis, J.P., Clarke, M., Devereaux, P.J., Kleijnen, J., Moher, D., 2009. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. Bmj 339, b2700. https://doi.org/10.1136/bmj.b2700.

Liu, J.L., 2011. The role of the funnel plot in detecting publication and related biases in meta-analysis. Evid. Based Dent. 12 (4), 121–122. https://doi.org/10.1038/sj.ebd.6400831.

Ludvigsson, M., Milberg, A., Marcusson, J., Wressle, E., 2015. Normal aging or depression? A qualitative study on the differences between subsyndromal depression and depression in very old people. Gerontologist 55 (5), 760–769. https://doi.org/10.1093/geront/gnt162.

Malhi, G.S., Mann, J.J., 2018. Depression. Lancet 392 (10161), 2299–2312. https://doi.org/10.1016/s0140-6736(18)31948-2.

de Man-van Ginkel, J.M., Hafsteinsdóttir, T.B., Lindeman, E., Ettema, R.G., Grobbee, D. E., Schuurmans, M.J., 2013. In-hospital risk prediction for post-stroke depression: development and validation of the post-stroke depression prediction scale. Stroke 44 (9), 2441–2445. https://doi.org/10.1161/strokeaha.111.000304.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Bmj 339, b2535. https://doi.org/10.1136/bmj.b2535.

Moons, K.G., Royston, P., Vergouwe, Y., Grobbee, D.E., Altman, D.G., 2009. Prognosis and prognostic research: what, why, and how. Bmj 338, b375. https://doi.org/10.1136/bmj.b375.

Moons, K.G., Kengne, A.P., Woodward, M., Royston, P., Vergouwe, Y., Altman, D.G., Grobbee, D.E., 2012. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. Heart 98 (9), 683–690. https://doi.org/10.1136/heartjnl-2011-301246.

Moons, K.G., Kengne, A.P., Grobbee, D.E., Royston, P., Vergouwe, Y., Altman, D.G., Woodward, M., 2012. Risk prediction models: II. External validation, model updating, and impact assessment. Heart 98 (9), 691–698. https://doi.org/10.1136/heartjnl-2011-301247.

Moons, K.G., de Groot, J.A., Bouwmeester, W., Vergouwe, Y., Mallett, S., Altman, D.G., Reitsma, J.B., Collins, G.S., 2014. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med 11 (10), e1001744. https://doi.org/10.1371/journal.pmed.1001744.

Moons, K.G.M., Wolff, R.F., Riley, R.D., Whiting, P.F., Westwood, M., Collins, G.S., Reitsma, J.B., Kleijnen, J., Mallett, S., 2019. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. Ann. Intern Med 170 (1), W1–w33. https://doi.org/10.7326/m18-1377.

Muka, T., Glisic, M., Milic, J., Verhoog, S., Bohlius, J., Bramer, W., Chowdhury, R., Franco, O.H., 2020. A 24-step guide on how to design, conduct, and successfully publish a systematic review and meta-analysis in medical research. Eur. J. Epidemiol. 35 (1), 49–60. https://doi.org/10.1007/s10654-019-00576-5.

Pashayan, N., Morris, S., Gilbert, F.J., Pharoah, P.D.P., 2018. Cost-effectiveness and benefit-to-harm ratio of risk-stratified screening for breast cancer: a life-table model. JAMA Oncol. 4 (11), 1504–1510. https://doi.org/10.1001/jamaoncol.2018.1901.

Patel, V., Araya, R., Bolton, P., 2004. Treating depression in the developing world. Trop. Med Int Health 9 (5), 539–541. https://doi.org/10.1111/j.1365-3156.2004.01243.x.

Radloff, L.S., 1977. The CES-D scale: a self-report depression scale for research in the general population. Appl. Psychol. Meas. 1 (3), 385–401. https://doi.org/10.1177/014662167700100306.

Reynolds 3rd, C.F., Cuijpers, P., Patel, V., Cohen, A., Dias, A., Chowdhary, N., Okereke, O.I., Dew, M.A., Anderson, S.J., Mazumdar, S., Lotrich, F., Albert, S.M., 2012. Early intervention to reduce the global health and economic burden of major depression in older adults. Annu Rev. Public Health 33, 123–135. https://doi.org/10.1146/annurev-publhealth-031811-124544.

Riley, R.D., Ensor, J., Snell, K.I.E., Debray, T.P.A., Altman, D.G., Moons, K.G.M., Collins, G.S., 2016. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. Bmj 353, i3140. https://doi.org/10.1136/bmj.i3140.

Royston, P., Altman, D.G., 2013. External validation of a Cox prognostic model: principles and methods. BMC Med Res Method. 13, 33. https://doi.org/10.1186/1471-2288-13-33.

Royston, P., Altman, D.G., Sauerbrei, W., 2006. Dichotomizing continuous predictors in multiple regression: a bad idea. Stat. Med 25 (1), 127–141. https://doi.org/10.1002/sim.2331.

RUBIN, D.B., 1976. Inference and missing data. Biometrika 63 (3), 581–592. https://doi.org/10.1093/biomet/63.3.581.

Santomauro, D.F., Herrera, A.M.M., Shadid, J., Zheng, P., Ashbaugh, C., Pigott, D.M., Abbafati, C., Adolph, C., Amlag, J.O., Aravkin, A.Y., 2021. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. Lancet 398 (10312), 1700–1712. https://doi.org/10.1016/S0140-6736(21)02143-7.

Steyerberg, E.W. (2019). Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. Springer International Publishing. https://link.springer.com/book/10.1007/978–0-387–77244-8.

Steyerberg, E.W., Vergouwe, Y., 2014. Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur. Heart J. 35 (29), 1925–1931. https://doi.org/10.1093/eurheartj/ehu207.

Steyerberg, E.W., Vickers, A.J., Cook, N.R., Gerds, T., Gonen, M., Obuchowski, N., Pencina, M.J., Kattan, M.W., 2010. Assessing the performance of prediction models: a framework for traditional and novel measures. Epidemiology 21 (1), 128–138. https://doi.org/10.1097/EDE.0b013e3181c30fb2.

Steyerberg, E.W., Moons, K.G., van der Windt, D.A., Hayden, J.A., Perel, P., Schroter, S., Riley, R.D., Hemingway, H., Altman, D.G., 2013. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. PLoS Med 10 (2), e1001381. https://doi.org/10.1371/journal.pmed.1001381.

Su, D., Zhang, X., He, K., Chen, Y., 2021. Use of machine learning approach to predict depression in the elderly in China: A longitudinal study. J. Affect Disord. 282, 289–298. https://doi.org/10.1016/j.jad.2020.12.160.

Sullivan, L.M., Massaro, J.M., D'Agostino Sr., R.B., 2004. Presentation of multivariate data for clinical use: The Framingham Study risk score functions. Stat. Med 23 (10), 1631–1660. https://doi.org/10.1002/sim.1742.

VanItallie, T.B., 2005. Subsyndromal depression in the elderly: underdiagnosed and undertreated. Metabolism 54 (5 Suppl 1), 39–44. https://doi.org/10.1016/j.metabol.2005.01.012.

Vega, W.A., Rodriguez, M.A., Ang, A., 2010. Addressing stigma of depression in Latino primary care patients. Gen. Hosp. Psychiatry 32 (2), 182–191. https://doi.org/10.1016/j.genhosppsych.2009.10.008.

Vergouwe, Y., Moons, K.G., Steyerberg, E.W., 2010. External validity of risk models: Use of benchmark values to disentangle a case-mix effect from incorrect coefficients. Am. J. Epidemiol. 172 (8), 971–980. https://doi.org/10.1093/aje/kwq223.

Wolff, R.F., Moons, K.G.M., Riley, R.D., Whiting, P.F., Westwood, M., Collins, G.S., Reitsma, J.B., Kleijnen, J., Mallett, S., 2019. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. Ann. Intern Med 170 (1), 51–58. https://doi.org/10.7326/m18-1376.

Wuthrich, V.M., Frei, J., 2015. Barriers to treatment for older adults seeking psychological therapy. Int Psychogeriatr. 27 (7), 1227–1236. https://doi.org/10.1017/s1041610215000241.

Wynants, L., Van Calster, B., Collins, G.S., Riley, R.D., Heinze, G., Schuit, E., Bonten, M. M., Dahly, D.L., Damen, J.A., Debray, T.P., 2020. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. Bmj *369*. https://doi.org/10.1136/bmj.m1328.

Xu, T., Clemson, L., O'Loughlin, K., Lannin, N.A., Dean, C., Koh, G., 2018. Risk factors for falls in community stroke survivors: a systematic review and meta-analysis. e565 Arch. Phys. Med Rehabil. 99 (3), 563–573. https://doi.org/10.1016/j.apmr.2017.06.032.

Xu, Z., Zhang, Q., Li, W., Li, M., Yip, P.S.F., 2019. Individualized prediction of depressive disorder in the elderly: a multitask deep learning approach. Int J. Med Inf. 132, 103973 https://doi.org/10.1016/j.ijmedinf.2019.103973.

Yourman, L.C., Lee, S.J., Schonberg, M.A., Widera, E.W., Smith, A.K., 2012. Prognostic indices for older adults: a systematic review. Jama 307 (2), 182–192. https://doi.org/10.1001/jama.2011.1966.