

Identification and risk prediction models for social frailty in older adults: a systematic review

Wu Xiaoqian

Dali University

Wang Li

kunwangli@126.com

Affiliated Hospital of Yunnan University

Yang Hongwei

Dali University

Li Zhongjian

Dali University

Huang Yani

Dali University

Deng Yishu

Affiliated Hospital of Yunnan University

Huang Yanzi

Dali University

Shen Xia

Dali University

Research Article

Keywords: Social frailty, Older adults, Prediction model, Systematic review, Risk of bias, PROBAST, Calibration, External validation, Early identification

Posted Date: October 5th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10765626/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Background

Social frailty is closely associated with adverse health outcomes in older adults. Risk prediction models may facilitate its early identification; however, the methodological quality and clinical applicability of existing models remain unclear. This systematic review aimed to evaluate the methodological quality, predictive performance, and clinical applicability of identification and risk prediction models for social frailty in older adults.

Methods

Ten Chinese- and English-language databases, including the Cochrane Library, PubMed, Embase, CINAHL, Web of Science, Scopus, CNKI, Wanfang, CQVIP, and CBM, were searched from January 2017, when social frailty emerged as a distinct concept, to 30 June 2026. Eligible studies developed or validated identification or risk prediction models for social frailty in adults aged 60 years or older. Two reviewers independently screened studies and extracted data, with disagreements adjudicated by a third reviewer. Methodological quality was assessed with PROBAST, and the findings were synthesised narratively.

Results

Nine studies reporting 18 models and 4830 participants were included. Eight studies (88.9%) were cross-sectional and yielded identification models; only one was a prospective cohort study. Models were built with logistic regression or machine learning algorithms and included 5–14 final predictors, most frequently depression, educational level, age, family function, and social support. Internally validated AUCs were 0.771–0.993 for the identification models and 0.825 for the prognostic model; these were not pooled because their interpretations differ. Eight studies had insufficient outcome events (events per variable < 20); only two performed external validation, and only one reported the calibration intercept and slope. All nine studies were rated at high risk of bias, concentrated in the analysis domain; applicability concerns were low throughout.

Conclusions

The overall methodological quality of identification and risk prediction models for social frailty in older adults is suboptimal, with bias concentrated in the analysis domain. Future studies should adopt prospective cohort designs, adhere strictly to the TRIPOD reporting guideline, and strengthen external validation and clinical translation so that these models can move from research into practice.

Registration

PROSPERO registration number: CRD420261348102.

Background

With the acceleration of global population ageing, the World Health Organization (WHO) defines older adults as persons aged 60 years and over [1], and the risk of frailty rises with advancing age. Social frailty refers to a state of vulnerability at the social level, manifested by the loss of social relationships, insufficient social support, and reduced social participation [2]. Its reported prevalence reaches 16.8%–20% worldwide [3, 4] and approximately 15.2% in China [5]. Social frailty predisposes older adults to a range of serious adverse outcomes, including death, new-onset disability, loss of independence in activities of daily living, cognitive decline, and depression [6, 7], thereby markedly impairing health status and quality of life and placing a heavy burden on families and health-care systems [8, 9]. Encouragingly, available evidence indicates that social frailty can be ameliorated through early screening and intervention [10, 11]. Because social frailty spans the dimensions of social relationships, social support, and social participation, accurately identifying its key risk factors and stratifying high-risk individuals are prerequisites for targeting intervention resources and improving health outcomes in the older population.

Risk prediction models for social frailty help to quantify an individual's risk, enable the early identification of high-risk groups, and support the precise allocation of intervention resources. In recent years, a growing number of such models have been developed for older populations, targeting community-dwelling older adults, residents of nursing homes, older adults with chronic diseases, patients with type 2 diabetes mellitus, and patients receiving maintenance haemodialysis. Owing to the diversity of these target populations, the predictors incorporated into the models vary widely, spanning sociodemographic, psychosocial, disease-related, and dialysis-related indicators. However, the key risk factors addressed, the manner in which model performance is reported, and the methodological quality of the primary studies are inconsistent. Faced with multiple candidate models, clinicians and nurses find it difficult to judge which model suits their own practice setting [12, 42]. Heterogeneity in study design, sample representativeness, and modelling approaches has resulted in variable predictive performance and methodological quality, which seriously hampers the translation of these models from research into clinical practice. A systematic review that comprehensively summarises the predictor composition, predictive performance, and methodological quality of existing models can provide a basis for selecting, applying, and improving them. We therefore conducted a systematic review of identification and risk prediction models for social frailty in older adults, evaluating their methodological quality, predictive performance, and clinical applicability, with the aim of providing a reference for clinical practice.

Methods

Design and reporting

This systematic review was conducted in strict accordance with the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis checklist for systematic reviews and meta-analyses (TRIPOD-SRMA) [13] and is reported following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement [14]. The review protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO; registration number CRD420261348102).

Search strategy

We first searched the Cochrane Library for evidence syntheses and then searched the primary-study databases PubMed, Embase, CINAHL, Web of Science Core Collection, and Scopus, together with the Chinese databases China National Knowledge Infrastructure (CNKI), Wanfang, CQVIP, and the Chinese Biomedical Literature Database (CBM), yielding ten databases in total. Given that social frailty was proposed as a distinct concept only around 2017, the search was restricted to the period from January 2017 to June 2026 to maximise precision, and languages were limited to Chinese and English. Before the formal search, PubMed was used for a preliminary search, after which the search terms were adjusted and finalised. Following the PICOTS principle, the search strategy combined subject headings with free-text terms, and four conceptual groups were joined with "AND": (1) population (P)—older adults, with the Chinese terms "老年", "老", and "高龄" and the English terms aged, "older adult", elderly, and seniors; (2) topic of interest (I)—social frailty and related measures, with the Chinese terms "社会脆弱", "社会脆弱性", "社会孤立", and "社会支持" and the English terms "social frailty", "social vulnerability", "social isolation", and "social support"; (3) model type (C)—risk prediction models, with the Chinese terms "预测模型", "风险评估", "模型", and "机器学习" and the English terms "prediction model", "risk model", "prognostic model", "machine learning", and nomogram; and (4) model evaluation (O)—model validation and performance, with the English terms "model validation", discrimination, calibration, AUC, and "Hosmer-Lemeshow test". Taking PubMed as an example, the full strategy was: (("Social Isolation"[Mesh] OR "Social Support"[Mesh] OR "social frailty"[Title/Abstract] OR "social vulnerability"[Title/Abstract] OR "social isolation"[Title/Abstract] OR "social support"[Title/Abstract]) AND ("Aged"[Mesh] OR aged[Title/Abstract] OR "older adult"[Title/Abstract] OR elderly[Title/Abstract]) AND ("prediction model"[Title/Abstract] OR "risk model"[Title/Abstract] OR "machine learning"[Title/Abstract] OR nomogram[Title/Abstract]) AND ("model validation"[Title/Abstract] OR discrimination[Title/Abstract] OR calibration[Title/Abstract] OR AUC[Title/Abstract] OR "Hosmer-Lemeshow"[Title/Abstract])). The complete search strategies for all databases are provided in Additional file 2. In addition, the reference lists of included studies were hand-searched to identify further eligible studies.

Eligibility criteria

Studies were eligible if they met the following criteria: (1) study design—cohort or case-control studies; given that cross-sectional studies cannot establish the temporal sequence between predictors and outcome and therefore cannot predict future risk, they were included only as studies of identification

models and were synthesised separately from prognostic prediction models, with prospective studies prioritised during the search; (2) population—adults aged 60 years or older [1]; (3) content—studies that used statistical methods to develop or validate an identification or risk prediction model for social frailty in older adults, where the model contained at least two predictors and at least one measure of model performance, such as the area under the receiver operating characteristic curve (AUC), the Hosmer-Lemeshow test, or sensitivity, was reported; (4) outcome—social frailty determined by a dedicated assessment tool such as the HALFT scale, the Social Frailty Index (SFI), or the six-item Lubben Social Network Scale (LSNS-6), excluding studies in which the outcome was determined by instruments based on other concepts such as the Social Vulnerability Index (SVI); and (5) language—Chinese or English, with the full text available. Studies were excluded if (1) the full text was unavailable or key data were incompletely reported; (2) they were duplicate publications; (3) they reported only the association between a single predictor and social frailty without building a multivariable model; (4) the outcome was another frailty subtype, such as cognitive or physical frailty, without social frailty reported separately; (5) they focused on frailty progression or state transition as the outcome; or (6) they aimed to evaluate the diagnostic accuracy of a frailty assessment tool or model.

Study selection

Two reviewers independently performed the following steps. First, the retrieved records were imported into EndNote X9, duplicates were removed, and titles and abstracts were screened to identify potentially eligible studies. Second, the full texts of the remaining records were assessed against the eligibility criteria. Finally, the reference lists of the included studies were examined to identify additional eligible studies. Disagreements were resolved through discussion and, when necessary, adjudicated by a third reviewer.

Data extraction

One reviewer extracted data using a piloted form designed according to the CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS) [15], and a second reviewer checked the extracted information for accuracy. The extracted data comprised two parts: (1) general information—author (year), country, study design, data source, participants, outcome measure, number of social frailty cases/sample size (%), candidate predictors, and number of models; and (2) model information—handling of missing data, handling of continuous variables, predictor selection method, model-building method, model validation method, discrimination and calibration, the predictors finally included in the model, and the form of model presentation. When a study reported several models (for example, models built with different algorithms), information on the development, validation, and performance of each model was extracted separately on a per-model basis; when several models shared the same data source and analytical pipeline, the quality assessment was performed at the study level.

Quality assessment

Two reviewers independently assessed the risk of bias and applicability of the included models using the Prediction model Risk Of Bias ASsessment Tool (PROBAST) [16], which comprises 20 signalling questions across four domains: participants, predictors, outcome, and analysis. Each question can be answered "yes", "probably yes", "no", "probably no", or "no information"; "yes" and "probably yes" indicate low risk of bias, whereas "no" and "probably no" indicate high risk of bias. The overall risk of bias was classified as low (all domains at low risk) or high (at least one domain at high risk), and disagreements were resolved through discussion. Cross-sectional studies were appraised as identification models and the cohort study as a prognostic model.

Data synthesis

Given the differences among the included studies in participant characteristics, social frailty assessment tools, predictor types, and modelling approaches, the results were synthesised narratively. A meta-analysis (quantitative pooling) was not performed for the following reasons: (1) the included models differed considerably in study population, outcome assessment tool, predictor set, and modelling algorithm, and the resulting clinical and methodological heterogeneity meant that pooling predictor effect estimates (such as odds ratios) could be misleading; (2) eight studies were cross-sectional, in which the AUC reflects the ability to discriminate prevalent status, a meaning different from that of the AUC of a prognostic model derived from a cohort study, so pooling AUCs across studies is methodologically unsound; and (3) only two studies performed external validation, leaving no poolable external-validation performance data. We therefore conducted a narrative synthesis and present the results for identification models and prognostic prediction models separately. We summarised the data sources of the included models and information on model development and validation, including study setting, participant characteristics, outcome measures, handling of missing data and continuous variables, methods of model development and validation, measures of discrimination and calibration, and the form of model presentation. For discrimination, the AUC or C-index was extracted [17, 18]. For calibration, the Hosmer-Lemeshow test or the Brier score was extracted [17, 19], with calibration intercepts, calibration slopes, and calibration plots extracted preferentially, and development, internal-validation, and external-validation performance distinguished [20]. In addition, we summarised the predictors finally included in each model; high-frequency predictors were counted per independent study.

Results

Search results

[Authors to complete before submission: report the number of records identified through database searching and other sources, the numbers screened and excluded with reasons, and the nine studies

finally included, cross-referring to the PRISMA 2020 flow diagram (Fig. 1). This information is missing from the Chinese manuscript and must be added to comply with PRISMA 2020.]

Characteristics of the included studies

Table 1 summarises the characteristics of the included studies. The nine studies were published between 2019 and 2026; all were conducted in China, and most were single-centre, cross-sectional studies, which limits the generalisability of the resulting models to other regions and populations. Seven studies (77.8%) were published in 2024 or later. Eight studies [21–25, 27–29] (88.9%) were cross-sectional and yielded identification models of prevalent social frailty; only the study by Xian [26] used a prospective cohort design, and the resulting model is therefore capable of predicting future risk. Five studies (55.6%) were conducted in hospitals: three focused on older patients receiving maintenance haemodialysis [23, 27, 29], one on older patients with type 2 diabetes mellitus [22], and one on hospitalised older patients with gastric cancer [28]. Three studies (33.3%) were conducted in the community [21, 24, 25], including the study by Han et al. [25] among community-dwelling older adults with chronic diseases, and one study (11.1%) was conducted in nursing homes [26]. Sample sizes ranged from 340 to 785, with a total of 4830 participants. In total, 1799 cases of social frailty were reported, with prevalences ranging from 23.69% to 68.07%; the lowest prevalence was reported by Han et al. [25] (23.69%) and the highest by Kuo et al. [24] (68.07%), while the prospective cohort study by Xian [26] reported a one-year incidence of 36.6%. Six studies [22, 23, 25–27, 29] (66.7%) assessed social frailty with the HALFT scale, two studies [21, 28] used the Social Frailty Index (SFI), and one study [24] used the LSNS-6.

Model development characteristics

Table 2 summarises the development characteristics of the social frailty prediction models. With respect to candidate predictors, only the study by Wang et al. [23] relied primarily on psychosocial factors, whereas the remaining studies selected multidimensional predictors spanning sociodemographic, psychosocial, physical health, lifestyle, disease-related, and dialysis-related indicators, with 14 to 24 candidate predictors per study. Regarding the handling of missing data, two studies [26, 27] reported a missing-data strategy: Xian [26] applied multiple imputation to follow-up data (including the outcome), and Fan et al. (2026a) [27] excluded incomplete questionnaires; the remaining seven studies (77.8%) did not explicitly report how missing values were handled. Regarding sample size estimation, only the study by Kuo et al. [24] achieved an events-per-variable (EPV) ratio of approximately 29, meeting the recommended threshold of $EPV \geq 20$; the remaining eight studies all had EPV values below 20, and five of them had EPV values below 10, highlighting the widespread problem of insufficient outcome events.

Univariable analysis was the most commonly used predictor selection method among the included studies ($n = 5$, 55.6%); of these five studies [22, 23, 25, 27, 28], two [22, 25] subsequently applied least absolute shrinkage and selection operator (LASSO) regression. Two further studies used LASSO regression directly, either alone [26] or in combination with ten-fold cross-validation [29], and one study [24] determined its predictors a priori through literature review and expert consensus. Three studies [22,

24, 29] compared model performance using machine learning; the algorithms used across the included studies comprised logistic regression, extreme gradient boosting (XGBoost), random forest, decision trees, the C5.0 algorithm, bagged classification and regression trees (Bagged CART), model-averaged neural networks, and stochastic gradient boosting. The number of final predictors ranged from 5 to 14, and the 18 models incorporated approximately 40 distinct predictors in total. As shown in Fig. 2, the most frequently included predictor was depression ($n = 7$, 77.8%), followed by educational level ($n = 5$, 55.6%); age was included in the models of four studies [21, 24, 25, 28] ($n = 4$, 44.4%). Other frequent predictors were family function or family care ($n = 4$, 44.4%), social support ($n = 3$, 33.3%), and marital status ($n = 3$, 33.3%). In addition, predictors such as dialysis vintage and dialysis symptom burden appeared only in studies of haemodialysis populations, whereas predictors such as glycosylated haemoglobin appeared only in studies of diabetic populations. The final predictors of the 18 models are presented in Table 5.

Model validation characteristics

Table 3 summarises the validation characteristics of the models. All nine studies reported internal validation: five studies [23, 25–28] (55.6%) used bootstrap resampling with 1000 repetitions; Kuo et al. [24] used a 70/30 hold-out split combined with ten-fold cross-validation of the training set; and Han et al. [25] and Fan et al. (2026b) [29] used ten-fold cross-validation at the variable-selection stage. Four studies [21, 22, 25, 29] performed internal validation by randomly splitting the sample into training and validation sets (7:3). Only two studies conducted external validation: Feng et al. [28] performed temporal validation using data from a different time period, and Fan et al. (2026a) [27] performed geographical external validation using multicentre data.

Model performance

As shown in Table 4, all included models reported discrimination, expressed as the AUC, and all AUCs reported at the model-development stage exceeded 0.70. Because the AUC of a cross-sectional identification model reflects the ability to discriminate prevalent status, whereas the AUC of a cohort-based prognostic model reflects the ability to predict future risk, the two are reported separately and were not pooled across studies. At internal validation, the AUCs of the eight cross-sectional identification models ranged from 0.771 to 0.993, and the AUC of the prospective cohort prognostic model was 0.825. In external validation, the models of Feng et al. [28] and Fan et al. (2026a) [27] showed good discrimination, with AUCs of 0.824 and 0.876, respectively. Discrimination varied across models built in different populations. The community-based model of Zhang [21] yielded the highest AUCs (0.987–0.993); these values are conspicuously high and are probably related to the limited sample size (650 participants), the large number of candidate predictors (24), and internal validation by 7:3 random splitting alone, suggesting possible overfitting and warranting cautious interpretation. The model of Fan et al. (2026a) [27], developed in haemodialysis patients, achieved an internal-validation AUC of 0.888 and an external-validation AUC of 0.876, and among the machine learning models of Fan et al. (2026b) [29], the random forest model performed best (AUC = 0.884). The model of Feng et al. [28] achieved an AUC

of 0.930 in the development cohort, but performance fell to 0.824 in the validation cohort. By comparison, discrimination was lower in models for community-dwelling and hospitalised populations: the model of Han et al. [25] achieved AUCs of 0.828–0.856, and the model of Xian [26] an AUC of 0.825. Eight studies [21–23, 25–29] reported calibration measures, none of which indicated miscalibration; Kuo et al. [24] did not assess calibration. Four studies [22, 26, 27, 29] additionally reported the Brier score. Furthermore, five studies [21, 23, 25–27] performed decision curve analysis (DCA), and some models reported auxiliary metrics such as sensitivity, specificity, positive and negative predictive values, and the F1 score.

Model presentation

The nomogram was the most common form of visual presentation: six studies [21, 23, 25–28] presented their prediction models as nomograms. Two studies [26, 29] developed web-based calculators providing online computation platforms. Zhou et al. [22] presented a variable-importance plot, and Kuo et al. [24] provided a web-based interactive interface accessible from mobile devices and personal computers, offering good platform compatibility and user experience.

Risk of bias and applicability

The risk-of-bias and applicability assessments of the included models are summarised in Table 6. Two reviewers independently applied PROBAST, and disagreements were resolved by consensus. Because the eight cross-sectional studies yielded identification models and the study by Xian [26] yielded a prognostic prediction model, the assessments were performed according to the intended use of each model: cross-sectional data were considered an appropriate data source for identification purposes, and prospective cohort data an appropriate source for prognostic prediction. The assessment was conducted on a per-model basis. The multi-algorithm models of Zhou et al. [22], Kuo et al. [24], and Fan et al. (2026b) [29] shared the same data source, predictor set, outcome definition, and analytical pipeline within each study, so the domain-level judgements apply to all models within the respective study.

In the participants domain, all nine studies were rated at low risk of bias: the data sources matched the intended use of the models, and the inclusion and exclusion criteria for participants were clear and reasonable; question-level judgements are provided in Additional file 3 (Table S7). In the predictors domain, all nine studies were rated at low risk of bias: predictors were assessed in the same way for all participants using standardised instruments, assessors were unaware of outcome information, and all predictors would be routinely available at the time the model is intended to be used. Notably, although the predictors of Kuo et al. [24] were determined a priori through literature review and expert consensus, a priori specification of predictors does not in itself constitute high risk of bias in this domain.

In the outcome domain, all nine studies were rated at low risk of bias. Outcomes were determined with validated social frailty instruments (HALFT, SFI, or LSNS-6) using prespecified cut-offs applied consistently to all participants; predictors were not part of the outcome definition; and outcomes were collected mainly by self-report questionnaire, so that determination was not influenced by knowledge of

predictor values. In the study by Xian [26], predictors were measured at baseline and the outcome was ascertained at one-year follow-up, establishing a clear temporal sequence. In the analysis domain, however, all nine studies were rated at high risk of bias. The main problems were: (1) an insufficient number of outcome events—only Kuo et al. [24] reached an EPV of approximately 29, whereas all other studies had EPV values below 20, and five studies [21–23, 27, 29] had EPV values below 10; (2) inappropriate predictor selection—five studies [22, 23, 25, 27, 28] selected predictors on the basis of univariable analysis; (3) inadequate reporting of missing-data handling—only Fan et al. (2026a) [27] explicitly reported that the dataset was complete, whereas Xian [26] used multiple imputation to fill in missing outcomes among participants lost to follow-up, an approach that does not meet methodological requirements, and the remaining seven studies did not report any missing-data strategy; and (4) insufficient model validation—four studies [21, 22, 25, 29] used random data splitting for internal validation, which may overestimate model performance, Kuo et al. [24] did not evaluate calibration, and only Fan et al. (2026a) [27] reported the full calibration hierarchy (calibration intercept and slope) and conducted multicentre geographical external validation.

With respect to applicability, concerns were low in the participants, predictors, and outcome domains for all nine studies, whose populations, predictors, and outcome definitions matched the review question. It should be emphasised that low applicability concerns merely reflect the match between the included studies and the question of this review and do not imply that the models are ready for clinical implementation; in particular, cross-sectional identification models should not be used directly to predict future risk.

Discussion

Principal findings and risk of bias of the models

Social frailty is significantly associated with all-cause mortality in older adults [30], and it is reversible in its early stages; timely identification of high-risk individuals followed by targeted intervention is therefore a key step in reducing the risk of adverse health outcomes [7], and building risk prediction models is an effective means of achieving early identification [10]. According to the PROBAST assessment, all included models were judged to be at high risk of bias, with the bias concentrated in the analysis domain while the participants, predictors, and outcome domains were at low risk in all nine studies; the main manifestations were insufficient numbers of outcome events, predictor selection based on univariable analysis, inadequate reporting of missing-data handling, and insufficient model validation. Applicability concerns were low in all three domains. It should be emphasised that low applicability concerns reflect only the match between the included studies and the review question and are not equivalent to clinical implementability. Given that all models are at high risk of bias and that most lack independent external validation and a complete model equation, they should not be recommended for direct clinical use at this stage. Future research should focus on improving the methodological quality of existing social frailty models and on developing new prognostic prediction models using rigorous prospective designs [31].

Although PROBAST was published in 2019 [16], the included studies still fall short of its methodological requirements. This review found that the high risk of bias arose mainly from the analytical stage of model development and validation. Seven studies did not report how missing values were handled, and one study [26] multiply imputed the missing outcomes of participants lost to follow-up, which is itself methodologically inappropriate; simply deleting incomplete cases or ignoring the problem may introduce sampling bias and information loss. Regarding predictor selection, five studies [22, 23, 25, 27, 28] selected predictors based on univariable analysis, which may introduce bias by omitting important variables, whereas four studies [22, 25, 26, 29] used LASSO regression, which selects predictors and shrinks regression coefficients simultaneously and thus helps to prevent overfitting [35]. Regarding model building, most included studies used logistic regression, whose simple form and interpretable coefficients are readily converted into practical clinical tools [36]. We recommend that future work start from large-scale, multicentre, prospective cohort studies and establish a complete development pipeline spanning high-quality data collection to dynamic model validation [37].

Model validation is an indispensable step between model development and implementation and comprises internal and external validation [38]. Of the 18 models included in this review, only two [27, 28] underwent external validation; the remaining 16 underwent internal validation only. Internal validation is easier to perform because it evaluates model performance within the same population [38]. All nine studies reported internal-validation results, and eight reported both discrimination and calibration, whereas Kuo et al. [24] did not assess calibration. For a prediction model, calibration refers to the agreement between predicted risks and observed outcomes [19]; the lack of external validation and of complete calibration assessment means that the current models cannot yet be used directly in clinical practice. The harm of inadequate calibration assessment is immediate: if a model systematically overestimates or underestimates individual risk, clinical decisions will be based on distorted probabilities, which may lead to missed screening of high-risk individuals or excessive intervention in low-risk individuals. Moreover, most studies reported only the Hosmer-Lemeshow test without the calibration intercept, calibration slope, or calibration plot, making it difficult to identify models that discriminate well but calibrate poorly and to transport calibration across settings with different prevalences. In addition, five studies [21, 23, 25–27] performed decision curve analysis (DCA), suggesting potential net benefit within specific threshold ranges; compared with the ROC curve, DCA evaluates clinical net benefit across threshold probabilities and is therefore more informative [19, 30]. Overall, existing models urgently require stronger external validation and complete calibration assessment to improve their clinical utility and generalisability [40].

Predictors and assessment tools

Predictors can markedly influence the performance of a prediction model [41]. Social frailty results from the interplay of social, psychological, behavioural, and biological factors, and the included studies all selected multidimensional candidate predictors. Depression was among the most frequently included predictors in existing social frailty models; its frequency is influenced by the range of candidate variables, the measurement instruments, and the selection methods, and causal inferences cannot be

drawn from it. Routine screening for social frailty in older populations is recommended [10]. Five studies [22, 24–26, 28] included educational level as a predictor, suggesting that educational attainment and the reserve of cultural knowledge affect an individual's social participation. In four studies [21, 24, 25, 28], age was a risk factor for social frailty, indicating that older individuals are more susceptible. In four studies [23, 25, 27, 29], family function or family care was also a high-frequency predictor, reflecting the role of the family support system in the development of social frailty. Disease-specific predictors such as dialysis symptom burden and dialysis vintage appeared in the four models of Fan et al. [27, 29]; although dialysis-related indicators may improve model performance to some extent, their applicability to other settings is limited because they apply only to specific disease populations. Furthermore, high-frequency predictors such as social support, marital status, physical activity, and living arrangement require dedicated assessment tools; questionnaire-based variables may be more feasible than clinical measurements in community settings, but their inclusion may increase the burden on users, and inter-rater variability may introduce bias [38]. Assessors should therefore receive specific training to ensure consistency, and predictors that are routinely collected and easy to obtain should be preferred, so as to simplify the use of prediction models in clinical settings and facilitate external validation in other regions.

The prevalence of social frailty varied considerably among the included studies, ranging from 23.69% [25] to 68.07% [24], which may be attributable to differences in study setting, study population, and the social frailty assessment tools used. Several instruments are currently available for assessing social frailty; the studies included in this review used the HALFT scale [32], the SFI [33], and the LSNS-6 [34], with six studies [22, 23, 25–27, 29] applying the HALFT scale. However, social frailty measured with different instruments is inherently heterogeneous: the HALFT scale emphasises the fulfilment of social needs [32], whereas the LSNS-6 mainly measures the size of social networks and social support [34]. When applying social frailty prediction models, clinicians and nurses should therefore take into account the influence of the assessment tool on both the characteristics of the assessed individuals and the social frailty outcome itself.

Clinical implications and translational application

This review found that existing social frailty prediction models cover a broad range of populations, including community-dwelling older adults, hospitalised patients, and patients with specific diseases. The models of three studies [21, 24, 25] are applicable to community-dwelling older adults: Zhang [21] and Han et al. [25] each developed nomogram-based models for community-dwelling older adults with good internal-validation performance, and Kuo et al. [24] built six machine learning models for community-dwelling older adults, of which the C5.0 model showed excellent discrimination (AUC = 0.972) and was presented through a web-based interactive interface. In addition, the models of five studies target hospitalised older patients with specific diseases, namely type 2 diabetes mellitus [22], maintenance haemodialysis [23, 27, 29], and gastric cancer [28]. Among the five models for patients receiving maintenance haemodialysis, the nomogram model of Fan et al. (2026a) [27], built with logistic regression, demonstrated discrimination (AUC = 0.876) and calibration after external validation and is

useful for identifying high-risk individuals. The model of Wang et al. [23] achieved an internal-validation AUC of 0.921 but has not been externally validated, so its generalisability requires further verification. The model of Feng et al. [28] for older patients with gastric cancer showed good discrimination (AUC = 0.930), although performance in the validation cohort was somewhat lower than in the development cohort. The model of Zhang [21] reached AUCs as high as 0.987–0.993; however, that study was based on 650 community-dwelling older adults, 24 candidate predictors, and internal validation by 7:3 random splitting only, with a limited number of events, so the high discrimination is more likely to reflect overfitting caused by the small sample, random splitting, and excessive candidate variables, and should not be interpreted as evidence of superior model performance in the absence of external validation. By contrast, the model of Fan et al. (2026a) [27] retained an AUC of 0.876 after external validation, suggesting more robust generalisability. This indicates that the quality of a model should not be judged by the AUC alone but should be appraised in conjunction with the validation approach.

An appropriate form of presentation is a key consideration in promoting the implementation of prediction models in clinical settings [42]. Six studies [21, 23, 25–28] presented their models as nomograms (one model each), and three studies [24, 26, 29] provided a web-based calculator or interactive interface, whereas Zhou et al. [22] presented a variable-importance plot from a random forest. However, most models did not provide the complete model equation, which limits independent external validation and model updating by other researchers in different populations. The choice of presentation format should take into account the clinical context and the intended users: social frailty prediction models should be presented in a user-friendly form that can be applied across clinical scenarios, while the complete model equation should be provided to support external validation.

Limitations

This review has several limitations. First, only Chinese- and English-language publications were included, and the search period began in 2017; relevant studies published in other languages or earlier under different concept labels may have been missed, introducing a risk of language bias. Second, all included studies were conducted in China, which limits the generalisability of the findings to other regions. Third, the included models were predominantly built from cross-sectional data (eight of nine studies), whose AUCs reflect only the discrimination of prevalent status and cannot be extrapolated to the prediction of future risk. Fourth, because all included models were at high risk of bias and there was considerable heterogeneity in participant characteristics, assessment tools, predictor types, and modelling approaches, only a narrative synthesis was possible; predictor effects and model performance were not pooled, and the strength of the reported associations should be interpreted with caution. Finally, research on social frailty prediction models is still in its infancy, the number of included studies was small, and some conclusions may lack robustness owing to limited statistical power. In addition, the two studies by Fan et al. [27, 29] were conducted by the same research team under the same ethics approval, with highly overlapping recruitment periods and sites in the haemodialysis population; their samples may therefore partially overlap, and the pooled sample size reported in this review should be interpreted with this caveat in mind.

Implications for future research and practice

Future studies should strictly adhere to the TRIPOD reporting guideline and the PROBAST methodological requirements and implement the following measures in the context of social frailty research: (1) at the design stage, perform sample size estimation based on an EPV of at least 20 or the pmsampsize method, and prefer prospective cohort designs to ensure the temporal sequence between predictors and outcome; (2) handle missing data appropriately (for example, by multiple imputation) and avoid selecting predictors on the basis of univariable analysis; (3) standardise the outcome assessment tool for social frailty (the HALFT scale is recommended) and report the full calibration hierarchy, including the calibration intercept, calibration slope, and calibration plot; (4) prioritise predictors routinely accessible in the community (such as age, educational level, family function, and depression) and limit the number of variables to preserve clinical operability; (5) conduct multicentre external validation and dynamic model updating by leveraging community health-management platforms and annual health-examination data for older adults; and (6) promote the conversion of nomograms into web-based calculators and digital assessment tools embedded in electronic health records, thereby enabling dynamic monitoring and early warning of social frailty risk.

Conclusions

This systematic review included 18 identification and risk prediction models for social frailty in older adults reported in nine studies. Most models showed good discrimination at internal validation and were presented in intuitive formats such as nomograms, but calibration was incompletely assessed in some models. All models were at high risk of bias, external validation was severely lacking, and methodological quality needs improvement. Depression, educational level, age, and family function were the most common predictors, reflecting the multidimensional nature of the factors influencing social frailty. Overall, prediction models for social frailty remain at an early stage of development and are still far from meeting the requirements for clinical application. Future studies should strictly follow the PROBAST guideline and the TRIPOD statement, improve the methodology of model development, actively conduct multicentre external validation, adopt standardised assessment tools, and provide complete model equations, so as to promote the translation of social frailty prediction models from research into clinical practice and provide strong support for the early identification of and precise intervention for social frailty in older adults.

Abbreviations

ADL

activities of daily living

AIC

Akaike information criterion

APGAR

Family Adaptation, Partnership, Growth, Affection, and Resolve Index

AUC
area under the receiver operating characteristic curve
B-APQ
Brief Ageing Perceptions Questionnaire
BIC
Bayesian information criterion
BMI
body mass index
CBM
Chinese Biomedical Literature Database
CHARMS
CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling
Studies
CI
confidence interval
CNKI
China National Knowledge Infrastructure
DCA
decision curve analysis
DDS
Diabetes Distress Scale
DSI
Dialysis Symptom Index
ECOG PS
Eastern Cooperative Oncology Group performance status
EPV
events per variable
FCI
Functional Comorbidity Index
FFQ
food frequency questionnaire
GDS-5/GDS-15
five-/fifteen-item Geriatric Depression Scale
HADS
Hospital Anxiety and Depression Scale
HALFT
Help, Participation, Loneliness, Financial, and Talk scale
HD-SMI
Hemodialysis Self-Management Instrument
H-L

Hosmer-Lemeshow
HLS-EU-Q16
sixteen-item short form of the European Health Literacy Survey
IDI
integrated discrimination improvement
IPAQ-SF
International Physical Activity Questionnaire–Short Form
LASSO
least absolute shrinkage and selection operator
LSNS-6
six-item Lubben Social Network Scale
MNA-SF
Mini Nutritional Assessment–Short Form
NPV
negative predictive value
NRI
net reclassification improvement
PMS
Personal Mastery Scale
PPV
positive predictive value
PRISMA
Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROBAST
Prediction model Risk Of Bias ASsessment Tool
PSQI
Pittsburgh Sleep Quality Index
RCS
restricted cubic spline
SAS
Self-Rating Anxiety Scale
SFI
Social Frailty Index
SGB
stochastic gradient boosting
SPPB
Short Physical Performance Battery
SSRS
Social Support Rating Scale
SVI

Social Vulnerability Index

TAS-20

twenty-item Toronto Alexithymia Scale

TRIPOD

Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis

VIF

variance inflation factor

WHO

World Health Organization

XGBoost

extreme gradient boosting.

Declarations

Ethics approval and consent to participate

Not applicable. This systematic review was based on previously published studies and did not involve human participants, human data, or human tissue directly. The review protocol was registered with PROSPERO (CRD420261348102).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Clinical trial number:

not applicable

Supplementary information

Additional file 1 (DOCX): Tables S1–S6 — characteristics of the included studies, model development characteristics, validation characteristics, model performance, final predictors, and question-level PROBAST summary. Additional file 2 [to be supplied by the authors]: full electronic search strategies for

all ten databases. Additional file 3 [to be supplied by the authors]: question-level PROBAST judgements for each included study (Table S7), as cited in the Risk of bias section.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author Contribution

Authors' contributions Wu Xiaoqian and Wang Li were responsible for writing, reviewing, and editing the manuscript. They also contributed to the initial draft, developed the methodology, secured funding, conducted formal analysis, managed data, and conceptualized the study. Yang Hongwei and Li Zhongjian were responsible for writing, reviewing, and editing the manuscript. They also contributed to the methodology, conducted formal analysis, and participated in the conceptualization of the study. Huang Yani and Deng Yishu: Responsible for writing, review, editing, form analysis, and conceptualization. Huang Yanzi and Shen Xia contributed to the writing, reviewing, and editing, as well as the methodology and conceptualization.

Acknowledgements

Not applicable.

Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary information files.

Tables

Tables 1-6 are available in the Supplementary Files section.

References

1. World Health Organization. WHO clinical consortium on healthy ageing 2021: report of consortium meeting held virtually, 5–6 November 2021. Geneva: World Health Organization; 2022.
2. Rojas-Avila J, Araya AX, Pinilla Carrasco N. Social frailty in older people: a concept analysis. *Rev Cuid*. 2025;16(3):e4939. 10.15649/cuidarte.4939.

3. Gong XY, Gao J, Hou CM, et al. Prevalence of social frailty in older adults: a meta-analysis. *Military Nurs.* 2023;40(4):98–102. 10.3969/j.issn.2097-1826.2023.04.024. (in Chinese).
4. Jia B, Wang Z, Zhang T, et al. Prevalence of social frailty and risk factors among community-dwelling older adults: a systematic review and meta-analysis. *Arch Gerontol Geriatr.* 2024;123:105419.
5. Qi X, Li Y, Hu J, et al. Prevalence of social frailty and its associated factors in the older Chinese population: a national cross-sectional study. *BMC Geriatr.* 2023;23(1):532. 10.1186/s12877-023-04241-1.
6. Li X, Gao L, Qiu Y, et al. Social frailty as a predictor of adverse outcomes among older adults: a systematic review and meta-analysis. *Aging Clin Exp Res.* 2023;35(7):1417–28. 10.1007/s40520-023-02421-y.
7. Ragusa FS, Veronese N, Smith L, et al. Social frailty increases the risk of all-cause mortality: a longitudinal analysis of the English Longitudinal Study of Ageing. *Exp Gerontol.* 2022;167:111901.
8. Yamada M, Arai H. Social frailty predicts incident disability and mortality among community-dwelling Japanese older adults. *J Am Med Dir Assoc.* 2018;19(12):1099–103. 10.1016/j.jamda.2018.09.013.
9. Zou X, Zheng M, He Y, Wang H. Prevalence and determinants of social frailty among community-dwelling empty-nest older adults in China. *Front Public Health.* 2026;14:1860833. 10.3389/fpubh.2026.1860833.
10. Hong C, Zhang N, Sheng J, et al. Effect of multicomponent exercise intervention on multidimensional frailty in older adults with mild cognitive impairment: a secondary analysis of the MIND-STEP randomised clinical trial. *Age Ageing.* 2026;55(4):afag111.
11. Sun Q, Tan K, Tang H, et al. Incidence and predictive value of social frailty among community-dwelling older adults in Southwest China: a prospective cohort study. *Front Public Health.* 2023;11:1103651. 10.3389/fpubh.2023.1103651.
12. Ye ZC, Wang JH, Lu Q, et al. Methods and application practice of clinical prediction model development. *Chin J Epidemiol.* 2025;46(9):1640–9. 10.3760/cma.j.cn112338-20250415-00246. (in Chinese).
13. Snell KIE, Levis B, Damen JAA, et al. Transparent reporting of multivariable prediction models for individual prognosis or diagnosis: checklist for systematic reviews and meta-analyses (TRIPOD-SRMA). *BMJ.* 2023;381:e073538. 10.1136/bmj-2022-073538.
14. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* 2021;372:n71. 10.1136/bmj.n71.
15. Moons KGM, de Groot JAH, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med.* 2014;11(10):e1001744.
16. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med.* 2019;170(1):51–8.

17. Abe T, Seino S, Nofuji Y, et al. Development of risk prediction models for incident frailty and their performance evaluation. *Prev Med.* 2021;153:106768.
18. Li S, Fan W, Zhu B, et al. Frailty risk prediction model among older adults: a Chinese nation-wide cross-sectional study. *Int J Environ Res Public Health.* 2022;19(14):8410.
19. Van Calster B, Nieboer D, Vergouwe Y, et al. A calibration hierarchy for risk models was defined: from utopia to empirical data. *J Clin Epidemiol.* 2016;74:167–76.
20. Liu Q, Yang L, Shi Z, et al. Development and validation of a preliminary clinical support system for measuring the probability of incident 2-year (pre)frailty among community-dwelling older adults: a prospective cohort study. *Int J Med Inf.* 2023;177:105138.
21. Zhang YX. Development of a social frailty risk prediction model for urban community-dwelling older adults [master's thesis]. Hefei: Anhui Medical University; 2025 (in Chinese). 10.26921/d.cnki.ganyu.2025.002030
22. Zhou ZC, Wang FL, Liu HY, et al. Development of a social frailty risk prediction model for older patients with type 2 diabetes mellitus. *Mod Prev Med.* 2026;53(2):219–24. 10.20043/j.cnki.MPM.202507372. (in Chinese).
23. Wang X, Wang XH, Qi JQ, et al. Influencing factors of social frailty and development of a risk prediction model in older patients receiving maintenance haemodialysis. *Mod Prev Med.* 2025;52(11):1957–62. 10.20043/j.cnki.MPM.202406026. (in Chinese).
24. Kuo K-M, Talley PC, Kuzuya M, et al. Development of a clinical support system for identifying social frailty. *Int J Med Inf.* 2019;132:103979. 10.1016/j.ijmedinf.2019.103979.
25. Han CY, Shi KJ, Zhao YN, et al. Development and validation of a social frailty risk prediction model for community-dwelling older adults with chronic diseases. *Nurs Res.* 2026;40(8):1257–64. 10.12102/j.issn.1009-6493.2026.08.003. (in Chinese).
26. Xian YY. Development and validation of a social frailty risk prediction model for older adults in nursing institutions [master's thesis]. Chengdu: Chengdu University of Traditional Chinese Medicine; 2023 (in Chinese). 10.26988/d.cnki.gcdzu.2023.000658
27. Fan CR, Liu JY. Development and validation of a social frailty prediction model for older patients receiving maintenance haemodialysis. *Nurs Res.* 2026;40(12):2013–21. 10.12102/j.issn.1009-6493.2026.12.002. (in Chinese).
28. Feng Q, Guo QY, Zhou YZ, et al. Development and validation of a social frailty risk prediction model for older patients with gastric cancer. *Chin J Mod Nurs.* 2026;32(12):1595–602. 10.3760/cma.j.cn115682-20250825-04550. (in Chinese).
29. Fan CR, Liu JY. Development and comparison of machine learning-based prediction models for social frailty in older patients receiving maintenance haemodialysis. *Chin J Mod Nurs.* 2026;32(17):2253–61. 10.3760/cma.j.cn115682-20250909-04835. (in Chinese).
30. García-Botina HD, Curcio CL, Calderón-Larrañaga A et al. Social frailty as a predictor of mortality among older adults in four Colombian departments: evidence from the SABE survey. *BMC Geriatr.* 2026.

31. Kong LN, Yang L, Lyu Q, et al. Risk prediction models for frailty in older adults: a systematic review and critical appraisal. *Int J Nurs Stud*. 2025;167:105068.
32. Ma L, Sun F, Tang Z. Social frailty is associated with physical functioning, cognition, and depression, and predicts mortality. *J Nutr Health Aging*. 2018;22(8):989–95.
33. Zhang LL, Gao J, Wu CX, et al. Translation and psychometric testing of the Chinese version of the Social Frailty Index. *J Nurs Sci*. 2023;38(8):102–5. (in Chinese).
34. Lubben J, Blozik E, Gillmann G, et al. Performance of an abbreviated version of the Lubben Social Network Scale among three European community-dwelling older adult populations. *Gerontologist*. 2006;46(4):503–13.
35. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B*. 1996;58(1):267–88.
36. Christodoulou E, Ma J, Collins GS, et al. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol*. 2019;110:12–22.
37. Yang SZ, Wei JX, Wang Y, et al. Risk prediction models for frailty in patients receiving maintenance haemodialysis: a systematic review. *Chin J Nurs Educ*. 2026;23(1):101–9. (in Chinese).
38. Collins GS, Dhiman P, Ma J, et al. Evaluation of clinical prediction models (part 1): from development to external validation. *BMJ*. 2024;384:e074819. 10.1136/bmj-2023-074819.
39. Kerr KF, Brown MD, Zhu K, et al. Assessing the clinical impact of risk prediction models with decision curves: guidance for correct interpretation and appropriate use. *J Clin Oncol*. 2016;34(21):2534–40.
40. Zhao L, Leng Y, Hu Y, et al. Understanding decision curve analysis in clinical prediction model research. *Postgrad Med J*. 2024;100(1185):512–5.
41. Ahady Dolatsara H, Chen YJ, Leonard RD, et al. Explaining predictive model performance: an experimental study of data preparation and model choice. *Big Data*. 2023;11(3):199–213.
42. Bonnett LJ, Snell KIE, Collins GS, et al. Guide to presenting clinical prediction models for use in clinical settings. *BMJ*. 2019;365:l737. 10.1136/bmj.l73.

Figures

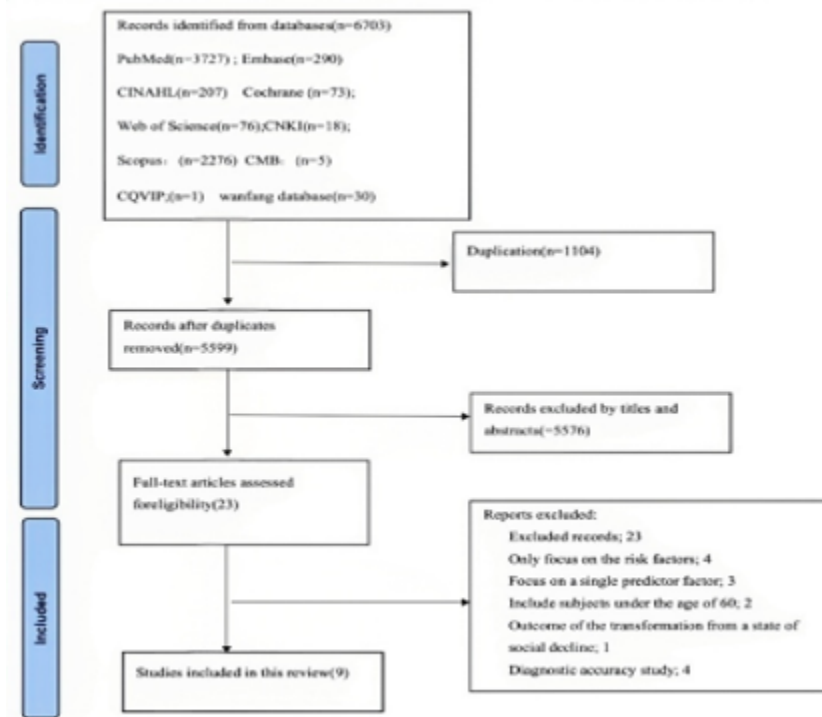


Figure 1

PRISMA 2020 flow diagram of the study selection process.

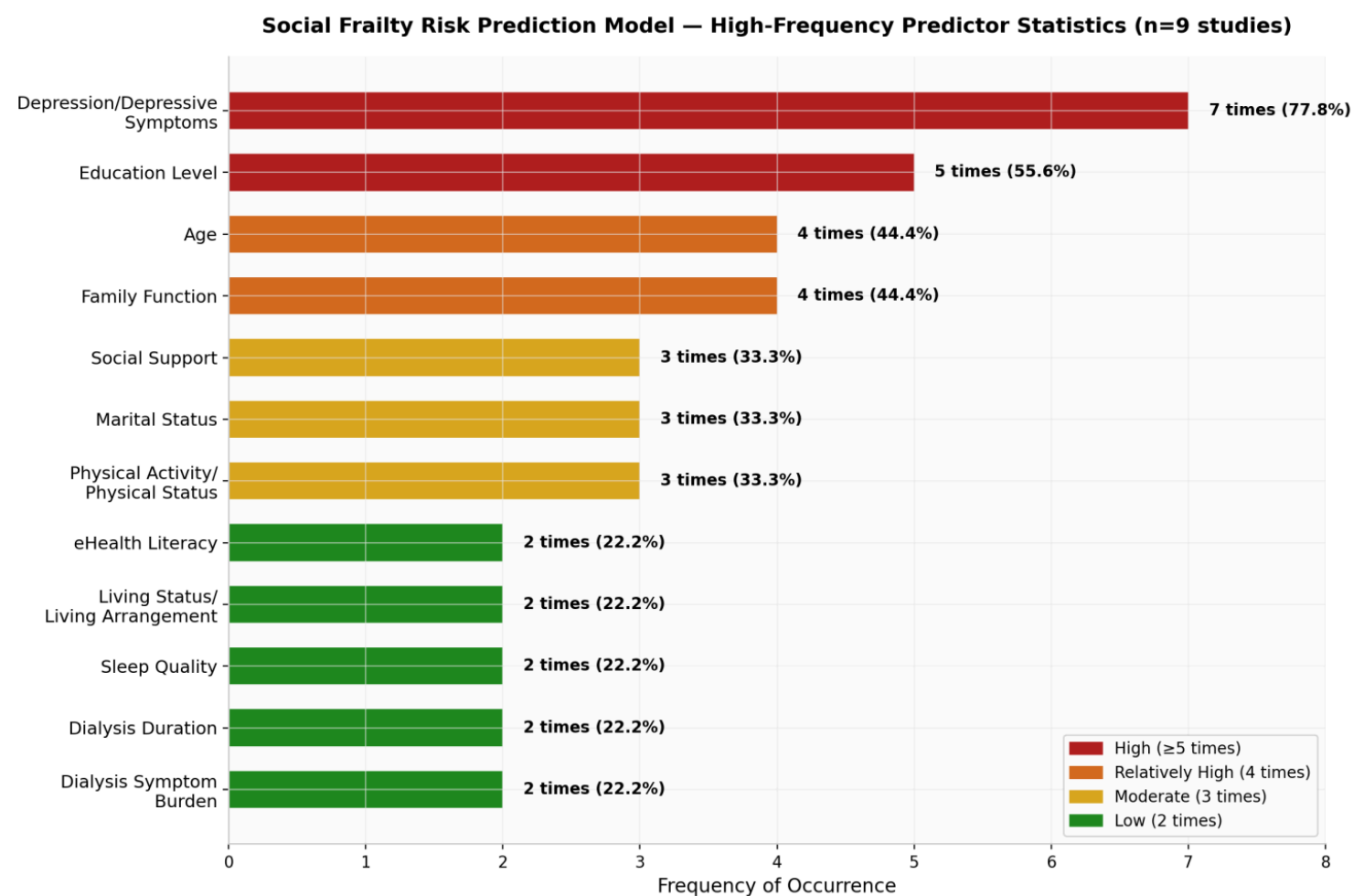


Figure 2

Frequency of predictors included in the final social frailty prediction models.

Supplementary Files

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

- [Additionalfile1Supplementarytables.docx](#)
- [Additionalfile1Searchstrategy.docx](#)
- [Additionalfile2PROBASTassessment.xlsx](#)
- [Tables.docx](#)