Predicting falls in community-dwelling older adults: a systematic review of prognostic models

Gustav Valentin Gade 1,2, Martin Grønbech Jørgensen,1 Jesper Ryg 3,4, Johannes Riis 1,2, Katja Thomsen 3,4, Tahir Masud,5 Stig Andersen 1,2

ABSTRACT

Objective To systematically review and critically appraise prognostic models for falls in community-dwelling older adults.

Eligibility criteria Prospective cohort studies with any follow-up period. Studies had to develop or validate multifactorial prognostic models for falls in community-dwelling older adults (60+ years). Models had to be applicable for screening in a general population setting.

Information source MEDLINE, EMBASE, CINAHL, The Cochrane Library, PsycINFO and Web of Science for studies published in English, Danish, Norwegian or Swedish until January 2020. Sources also included trial registries, clinical guidelines, reference lists of included papers, along with contacting clinical experts to locate published studies.

Data extraction and risk of bias Two authors performed all review stages independently. Data extraction followed the Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies checklist. Risk of bias assessments on participants, predictors, outcomes and analysis methods followed Prediction study Risk Of Bias Assessment Tool.

Results After screening 11 789 studies, 30 were eligible for inclusion (n=86 369 participants). Median age of participants ranged from 67.5 to 83.0 years. Falls incidences varied from 5.9% to 59%. Included studies reported 69 developed and three validated prediction models. Most frequent falls predictors were prior falls, age, sex, measures of gait, balance and strength, along with vision and disability. The area under the curve was available for 40 (55.6%) models, ranging from 0.49 to 0.87. Validated models 1 the area under the curve ranged from 0.62 to 0.69. All models had a high risk of bias, mostly due to limitations in statistical methods, outcome assessments and restrictive eligibility criteria.

Conclusions An abundance of prognostic models on falls risk have been developed, but with a wide range in discriminatory performance. All models exhibited a high risk of bias rendering them unreliable for prediction in clinical practice. Future prognostic prediction models should comply with recent recommendations such as Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis.

PROSPERO registration number CRD42019124021.

BACKGROUND

The propensity to fall is a serious and common health issue among older adults with one-third of community-dwelling adults ≥65 years and half of those ≥80 years falling annually.1 Consequences of falls are considerable with loss of independence, increased morbidity and mortality.2-4 Furthermore, the healthcare costs of falls increase substantially with age.5 Therefore, as the prevalence of older fallers is predicted to increase with changes in demography, preventing falls is of utmost importance.6

Falls interventions have proven effective when aimed at older adults with a high risk of falling.7 However, identifying these high-risk individuals is not straightforward since falling is multifactorial. Prognostic models combine risk factors to estimate the individual’s risk of a future outcome.8 Thus, a prognostic model may be a valuable tool to discriminate between older adults at high versus low risk of falling. To prevent the consequences of falls, healthcare professionals could perform screening in the general population using prognostic models.9 However, no systematic
review has addressed prognostic models on falls for community-dwelling older adults.

This systematic review aims to provide an updated overview of available models to be used by healthcare professionals and for researchers to improve on. The primary objective was to describe the discriminatory performance of prognostic models for falls in prospective cohort studies on community-dwelling older adults. Secondary objectives were to describe the study and model characteristics of these models.

**METHODS**

A protocol was preregistered before commencing the review process and is available in data supplements (online supplemental appendix 1). The review and its protocol followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement (PRISMA) and PRISMA-Protocols, respectively. A completed PRISMA Checklist is available in data supplements (online supplemental appendix 2). During the review process, we found the protocol unclear in terms of eligibility criteria for the study designs, participants, models, outcomes and settings, for which reason we have further described these. Rationales for the changes are given in the protocol. Box 1 provides an introduction to commonly used prediction modelling terms.

**Eligibility criteria**

**Participants and setting**

All participants had to be community dwelling, 60 years of age or older, and be recruited from a general population setting. For that reason, we excluded models intended for hospitals, general practitioners and nursing homes. Studies restricted to participants with prespecified diseases, conditions or symptoms such as Parkinson’s disease or stroke were excluded to raise external validity. However, we included studies that excluded certain types of community-dwelling older adults, such as those with known neurological, spinal or cognitive disorders.

**Index (model)**

Studies had to present a final multifactorial prognostic model defined by the inclusion of two prognostic factors or more. This definition was chosen as causes of falls are multifactorial and coexisting. Thus, prognostic factor studies investigating the association between predictors and prospective falls were excluded. Both development studies and validation studies with and without model updating were included.

**Outcome**

We included studies defining falls as ‘an unexpected event in which the participants come to rest on the ground, floor or lower level’. However, studies without an outcome definition were also included since this would not rule out the definition mentioned above. We excluded studies using fall definitions excluding certain

---

**Box 1 Commonly used prediction modelling terms with examples related to falls**

**Prognostic factor**

A prognostic factor, also called a predictor, is any measure that, among people with a given health condition, is associated with a subsequent clinical outcome such as falls.

**Prognostic prediction model**

A prognostic prediction model is a statistical combination of multiple predictors from which risks of a longitudinal outcome, for example, falls, can be calculated for individuals.

**Development and validation studies**

A prediction model development study aims to develop a prediction model by combining essential predictors from a data set into a model and testing its predictive performance within the same development data set. A model validation study aims to assess the predictive performance of a developed prediction model using new data not used in the development of the model.

**Model performance, overfitting and internal validation**

A model’s predictive performance is termed model performance. This term encompasses several measures with the two most important being discrimination and calibration. Estimates of model performance derived directly from a data set used for developing the model is termed the apparent performance. Since the model is fitted explicitly to the development data set, predictions on new data, that is, new older adults with different characteristics, may yield poorer model performance estimates, that is, poor generalisability. Hence, clinicians would typically find the apparent performance optimistic in terms of predicting a fall in their population which has not been used for developing the model. In consequence, fall preventive interventions could end up being provided to those not needing it and not offered to those actually in need hereof. The optimism in apparent performance is due to the model fitting too well to its data, a term known as overfitting. In such situations, predictions would be biased when the model is used on older adults with different characteristics, that is, frequency distributions of predictors. Estimating the amount of optimism in the development study’s model can be done using internal validation techniques such as bootstrap validation. However, since the population of older adults is heterogeneous, generalising a model’s performance to the entire population would be more clinically relevant. Here, internal validation procedures fail, and the model should instead be tested in a validation study.

**Model discrimination**

Model discrimination is a performance measure referring to the models’ ability to correctly predict if an individual will experience a fall or not. Therefore, as an example, it can be used by healthcare professionals to assess how confident a model assigns individuals to a high-risk group and guides the clinician when allocating fall preventive interventions. A perfectly discriminating model assigns a higher risk of falling to all older adults experiencing a fall. Likewise, a lower risk is appointed to those not suffering a fall. Usually, discrimination is reported as a concordance index (c-index) or an area under the curve, but other measures are also available. Here, a value of 1 equals perfect discrimination, and 0.5 indicates that the model discriminates no better than chance. If a model shows poor discriminative performance, it could predict low-risk older adults to fall and high-risk older adults to not fall.

**Model calibration**

Model calibration is a performance measure used to examine whether a model over- or underestimates the predicted risks in a sample.
Box 1 Continued

Specifically, it is the agreement between predictions made by the model and the frequency of the outcome to be predicted. Healthcare professionals can use this information to assess how confident the model predicts the specific risk of having a fall for the individual. In brief, it is crucial when counselling older adults on their fall risk that the risk estimate is as accurate as possible. If the model predicts a person to have a 10% risk of falling within 1 year, the observed frequency of people falling with such a predicted risk should be 10 out of 100 for the model to have good calibration. However, should the frequency be only 5 of 100 people, the model overestimates the risk. Calibration is typically assessed graphically using calibration plots. In development studies, models are usually calibrated well to the data from which they are developed and therefore yield limited information. Thus, it is more relevant how well the model is calibrated when introduced to a new sample of older adults used to validate the model. This information would enable healthcare professionals to evaluate whether the model over- or underestimates the risk of falling when used in their population of community-dwelling older adults. If the model is not correctly calibrated, it could predict low-risk older adults to have a higher risk and vice versa, or systematically overestimate or underestimate all predictions.

Language and publication year
Due to the composition of the study group, we only included published studies reported in English, Danish, Norwegian or Swedish languages. No restrictions on publication year were made.

Information sources
We searched electronic databases, trial registries and clinical guidelines. Furthermore, we consulted with additional clinical experts. Lastly, we screened conference abstracts along with reference lists of all the included studies and systematic reviews found during the search. Databases included Pubmed.gov (PubMed interface), EMBASE (Embase.com), CINAHL (EBSCOhost interface), The Cochrane Library (Wiley interface), PsycINFO (APA PsycNET interface), and Web of Science (Web of Science Core Collection). All databases were searched from inception dates to the 3 January 2020. Trial registries included PROSPERO, ClinicalTrials.gov, WHO International Clinical Trials Registry Platform and Open Grey. Guidelines included Guidelines International Network, the National Institute for Health and Care Excellence, Centre for Reviews and Dissemination and to Health Technology Assessments and Scottish Intercollegiate Guidelines Network. Conference abstracts and studies in trial registries were used to obtain full-text papers through contact with authors. Letters to the editor were excluded.

Study designs
We only included prospective cohort studies since this study design allows optimal control when measuring predictors and outcomes. Thus, it is the recommended study design for prognostic modelling studies. We excluded randomised controlled trials since these can have different limitations incorporated within their design. Typically, strict eligibility criteria are used that generate a highly selected sample of participants. This narrows predictors’ distribution and hence reduces the discriminatory performance in the prognostic models.

Also, strict criteria may compromise generalisability to the target population. Lastly, interventions in the study may also influence the discriminatory performance of the models. Retrospective cohort studies were also excluded due to issues of missing data and restrictions on which predictors to apply since data are already collected.

Timing
No restrictions on follow-up or predictive horizon were made since we found it clinically relevant to include models both able to predict falls within short and long periods ahead in time.
prioritised list of reasons (online supplemental appendix 4). Reviewers were not blinded to author names, institutions or journal titles.

**Data collection process**

We developed a standardised data collection form using Research Electronic Data Capture (REDCap), a research electronic data capture software, following the Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modelling Studies checklist. Data extraction was performed in duplicate and independently by two reviewers (GVG and JRi). Independence between reviewers was ensured using a double data entry module in REDCap, thereby denying access to each other’s responses. Disagreements among the reviewers were discussed, and the third reviewer (MGJ) was not consulted during data collection since consensus was reached in all studies. We contacted all study authors for retrieval of information on data items not reported. None of the included studies were published more than once.

**Data items**

We extracted data on the following items: country, publication year, authors, inclusion criteria, exclusion criteria, age, outcome definition, number of falls and fallers, candidate predictors, missing data, choice of statistical analysis, C-statistic and area under the curve (AUC), internal and external validation procedures, final model presentation and sources of funding. If available, 51 data items were extracted from each paper as detailed in online supplemental appendix 5.

**Risk of bias and reporting transparency**

To follow current recommendations, the Prediction study Risk Of Bias Assessment Tool was used for the risk of bias assessment in individual studies. The tool comprises 20 signalling questions in four domains: participants, predictors, outcomes, and analysis. The tool also includes an evaluation of each model’s applicability for the intended population, predictors and outcome of the review. Reporting transparency was assessed using the Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis, TRIPOD) adherence assessment form. The bias, applicability and reporting assessments were performed in duplicate and independently by two reviewers (GVG and JRi). Independence between reviewers was ensured using a double data entry module in REDCap, thereby denying access to each other’s responses. Disagreements among the reviewers were discussed, and consensus was reached for all studies. Thus, a third reviewer was not consulted for a final decision. The reviewers were not blinded to study authors, institutions or journal titles. The results of the risk of bias assessments of all included studies were incorporated into the qualitative synthesis. We sought to investigate outcome reporting bias by comparing the study papers to their pertaining protocols to examine whether outcomes were prespecified and not differing from the published paper.

**Summary measures and planned method of analysis**

The principal summary measure of this systematic review was the discriminatory performance measured either in a C-index or AUC. In the prespecified protocol, we decided not to perform meta-analyses due to the presumed heterogeneity of the prognostic models. This assumption was confirmed after the review was complete. Furthermore, we summarised the study and model characteristics using ranges and percentage proportions when appropriate. When data were available, we summarised continuous measures using medians and IQRs.

**Patient public involvement**

We did not involve patients or the public in the research.

**RESULTS**

**Study selection**

The search yielded 19,612 publications with 11,789 remaining after removal of duplicates. Screening titles and abstracts led to the exclusion of 11,611 publications leaving 178 for full-text reading. Of these, 148 were excluded due to: wrong outcome (n=45), wrong study design (n=45), not being a prediction model (n=25), no full-text paper published (n=14), wrong population (n=8), not multifactorial (n=8) or wrong setting (n=3). Thirty studies met the eligibility criteria and were included. Figure 1 displays the PRISMA flow diagram.
for the study selection. Details on excluded papers can be found as online supplemental appendices 6–8.

**Characteristics of included studies**

A summary of included studies along with models’ performance can be found in online supplemental table 1 (online supplemental appendix 9). All studies were published in English from 1994 to 2019. Seventy-two prognostic models were reported, of which 69 models were developed, and three were validated.

**Participants**

Studies were conducted in Australia, Belgium, Canada, France, Germany, Israel, Italy, Japan, the Netherlands, Malaysia, Spain, the UK and the USA. Sample sizes ranged from 65 to 23,417 participants with median ages from 67.5 to 83 years. Studies used primarily a probability sampling method (n=16), followed by convenience sampling (n=8) and consecutive sampling (n=2). Four studies did not report their sampling methods.

**Index/model**

The median (IQR) number of predictors in the final models were five (3–9) and ranged from two to 96 predictors. Figure 2 shows the number of studies including a specific predictor. The most frequently applied predictors were prior falls, age, sex, measures of gait, balance and strength, along with vision and disability. Predictors were measured in homes (n=19), research centres (n=19), or both (n=12). Locations for measuring predictors were not reported for 22 models.

**Outcomes and timing**

The percentage of fallers ranged from 5.9% to 59%, and the number of recurrent fallers (≥2 falls) ranged from 6.3%–54.1%. Models primarily predicted any falls (n=34), that is, single and recurrent falls, and recurrent falls only (n=34). Two models predicted first-time falls and two predicted time to a fall. Participants were followed for a median (IQR) time of 12 (9.75–12) months. Individual study data items extracted are available in online supplemental appendix 10.

**Model performance**

Discriminatory measures were reported for 40 (55.6%) models. AUCs were 0.49–0.87 and 0.62–0.69 for developed (n=37) and validated (n=3) models, respectively. Corresponding CIs were reported for 27 (37.5%) models. Calibration measures were available for seven (9.7%) models. For validated models (n=1), calibration was imperfect due to over- and underestimated predicted risks of falling for high-risk and low-risk participants, respectively. Regarding developed models (n=6), calibration was found acceptable, but studies did not assess model calibration in new participants.

**Risk of bias, reporting transparency and applicability within studies**

**Risk of bias**

Table 1 summarises ratings on risk of bias, applicability and reporting transparency for the individual studies. All studies had a high risk of bias mainly due to methods of analysis and outcome assessment along with restrictive eligibility criteria. Regarding analysis methods, missing data were excluded in 13 out of 30 studies, and no internal validation methods were applied. As to the outcome, only four studies recorded falls daily with monthly notifications. Also, the majority of studies did not report the outcome definition used or whether outcomes assessors were blinded. Eligibility criteria were found restrictive for the majority of studies due to the exclusion of individuals with falls-risk-increasing conditions. These selective criteria limit the usability of models for the target population of community-dwellers. Overall, risk of bias and applicability assessments were complicated by studies only reporting, on average, 50% of all items recommended in reporting guidelines. Furthermore, this
<table>
<thead>
<tr>
<th>Study</th>
<th>Reporting</th>
<th>Risk of bias</th>
<th>Applicability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TRIPOD, %</td>
<td>Participants</td>
<td>Predictors</td>
</tr>
<tr>
<td>Yamashita et al&lt;sup&gt;38&lt;/sup&gt;</td>
<td>34</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Weiss et al&lt;sup&gt;32&lt;/sup&gt;</td>
<td>36</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Muhiadat et al&lt;sup&gt;35&lt;/sup&gt;</td>
<td>36</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Brauer et al&lt;sup&gt;25&lt;/sup&gt;</td>
<td>37</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Panzer et al&lt;sup&gt;20&lt;/sup&gt;</td>
<td>37</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Maki et al&lt;sup&gt;34&lt;/sup&gt;</td>
<td>41</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Bougarides et al&lt;sup&gt;22&lt;/sup&gt;</td>
<td>41</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Viccaro et al&lt;sup&gt;31&lt;/sup&gt;</td>
<td>41</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Lindemann et al&lt;sup&gt;61&lt;/sup&gt;</td>
<td>44</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Singh et al&lt;sup&gt;44&lt;/sup&gt;</td>
<td>44</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Nandy et al&lt;sup&gt;23&lt;/sup&gt;</td>
<td>45</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Okochi et al&lt;sup&gt;40&lt;/sup&gt;</td>
<td>45</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Rodriguez-Molinero et al&lt;sup&gt;41&lt;/sup&gt;</td>
<td>46</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Gadkaree et al&lt;sup&gt;37&lt;/sup&gt;</td>
<td>48</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>de Vries et al&lt;sup&gt;24&lt;/sup&gt;</td>
<td>50</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Hinzdo et al&lt;sup&gt;43&lt;/sup&gt;</td>
<td>52</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Lohman et al&lt;sup&gt;42&lt;/sup&gt;</td>
<td>52</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Gillain et al&lt;sup&gt;45&lt;/sup&gt;</td>
<td>54</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Delbaere et al&lt;sup&gt;27&lt;/sup&gt;</td>
<td>55</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Cawthon et al&lt;sup&gt;48&lt;/sup&gt;</td>
<td>56</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Kim et al&lt;sup&gt;13&lt;/sup&gt;</td>
<td>56</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Tromp et al&lt;sup&gt;36&lt;/sup&gt;</td>
<td>57</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Stel et al&lt;sup&gt;46&lt;/sup&gt;</td>
<td>57</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Lamb et al&lt;sup&gt;25&lt;/sup&gt;</td>
<td>61</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Yamashita et al&lt;sup&gt;29&lt;/sup&gt;</td>
<td>61</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Stalenhoef et al&lt;sup&gt;47&lt;/sup&gt;</td>
<td>62</td>
<td>?</td>
<td>+</td>
</tr>
<tr>
<td>Palumbo et al&lt;sup&gt;10&lt;/sup&gt;</td>
<td>68</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Bongue et al&lt;sup&gt;36&lt;/sup&gt;</td>
<td>70</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Palumbo et al&lt;sup&gt;39&lt;/sup&gt;</td>
<td>73</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Pluijm et al&lt;sup&gt;19&lt;/sup&gt;</td>
<td>76</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Note: + and green colour indicates low risk of bias / low concern regarding applicability; − and red colour indicates high risk of bias / high concern regarding applicability; ? and blue colour indicates unclear risk of bias / unclear concern regarding applicability; Bright colours pertain to the PROBAST domains, and dark colours pertain to the overall PROBAST assessments; ROB indicates risk of bias; TRIPOD indicates the percentage proportion of reported items recommended in reporting guidelines. ROB, risk of bias; TRIPOD, Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis.
was complicated by a low response rate with four out of 30 study authors responding when contacted for clarification on study characteristics and data extraction items. Finally, outcome reporting bias assessments were not possible due to studies not referring to a preregistered protocol for their prognostic modelling study.

Applicability
Seven (23%) studies, with a total of 21 models, had low applicability concerns for the review question. Regarding participants, 17 (56.7%) studies were rated as having high or unclear applicability concerns for the review question. This concern was primarily due to restrictive eligibility criteria impeding generalisation to the general population of community dwellers. Restrictions were made by excluding participants with specific diseases or conditions that could increase the risk of falling, such as disability or impaired mobility. Furthermore, studies rated as having unclear applicability concerns did not sufficiently report whether the participants were community-dwellers or whether the setting was the general population rather than, for example, primary or secondary care. Regarding predictors, 28 (93.3%) studies had no concerns. The remaining two studies used specific laboratory measures which may be challenging to apply in a general population setting. Regarding applicability concerns for the outcome, 18 (60%) studies had no concerns since they reported using the falls definition of the review or similar.

DISCUSSION
The current systematic review found 72 prognostic models on falls risk with the area under the curve ranging from 0.49 to 0.87. All models had a high risk of bias mostly due to limitations in statistical methods and outcome assessments, combined with restrictive eligibility criteria. Thus, using the models in clinical practice would entail unreliable predictions. This review provides an extensive overview of prediction models for falls and information for future study methodology.

Strengths
The current review followed guidelines on prediction modelling reviews strictly for search strings, data extraction, risk of bias assessment, along with development and transparent reporting of the review.

Limitations
Review level
We excluded potentially eligible studies without full text (n=14) or published in other languages (n=7) during screening of titles and abstracts. These studies are listed in the data supplement. Furthermore, we excluded randomised controlled trials and retrospective cohort studies. Consequently, we were only able to include 0.25% (30/11 789) of studies screened even though other models, based on other study designs, may have been available. As prespecified in the study protocol, this exclusion criterion was chosen due to limitations with generalisability and missing data when developing or validating prediction models using these designs. Thus, this systematic review only provides an overview of models based on a specific study design, but we consider this exclusion of the other studies to be justified.

Study level
Limitations were found in the studies with a high risk of bias, poor quality of reporting, and finally, a low response rate when contacting authors for retrieval of missing data extraction items.

Risk of bias
We found a high risk of bias within all studies. Hence, the predictive performance may be low, and predictions unreliable when models are used in clinical practice. The bias ratings were primarily based on eligibility criteria, methods of outcome assessments, and statistical analysis. Building prediction models on selected subgroups of the target population can yield biased performance estimates when used in clinical practice on a different population. Thus, study eligibility criteria should be aligned with the research questions, that is, broad and with as few exclusion criteria as possible. In terms of outcome assessments, we found the definition of falls missing for one-third of studies along with varying falls recording methods. These findings are similar to results of a previous review on methodology in falls prevention trials, where only half of the studies provided a falls definition and recording methods varied highly. The problem with not defining a fall is that the notion of falls is taken for granted. As seen in our review, the prevalence of falls differed markedly between studies, which could be due to different understandings of the fall’s definition. Consequently, falls become harder to predict while at the same time, comparing and combining studies in systematic reviews with meta-analyses becomes complicated. To address these issues, a common outcome data set on falls trials is available along with a falls definition and recommendations for falls recording methods. Finally, statistical analysis methods raised concerns for risk of bias. Primarily, this was due to the handling of missing data with most of the studies applying a complete-case analysis method. Significant limitations can arise from the exclusion of participants due to missing data, for example, on a single predictor among many, since otherwise useful predictors on each participant are lost. Consequently, this can lead to low sample sizes and biased model performances. In such cases, imputation methods have proved useful when dealing with missing data. Furthermore, for the majority of developed models, no internal validation procedures were applied. This shortcoming typically causes models’ predictive performance to be optimistic. Finally, the critical appraisal was compromised due to incomplete reporting. We believe that future studies and systematic reviews would benefit from adhering to reporting guidelines for prediction modelling studies.
Implications for clinical practice

Only seven studies could address the review question appropriately, and all of these had a high risk of bias. Consequently, the evidence available to inform health-care professionals is limited and, as mentioned, possibly biased. Thus, no model can currently be recommended for clinical practice.

Implications for research

We recognise that most studies (n=23/30) were conducted before the publishing of prediction modelling guidelines.25 26 Thus, with the benefit of hindsight, studies would be expected to have different shortcomings within their methods and reporting. On the other hand, this also supports the reason for publishing guidelines in the first place. Despite this, the included studies provide valuable information on future candidate predictors. Thus, selecting predictors for prediction models on non-statistical grounds, that is, based on literature and clinical knowledge, is commonly used to avoid predictor selection bias.25 Therefore, future development studies may include the most frequently applied predictors found in this review. Lastly, it is essential to test the generalisability of developed models by performing validation studies to determine which models provide stable predictions across different populations.26

CONCLUSIONS

There are several studies on falls prognostic models intended for a general population setting, but only a few are fully applicable to the heterogeneous population of community-dwelling older adults. Thus, the evidence available to address this is limited. From all included studies, we found an abundance of falls prognostic models available. However, the discriminatory performance of these varied and was only reported for half of the models. Each model had concerns regarding risk of bias mainly due to restrictive eligibility criteria along with methods of statistical analysis and outcome assessments. Consequently, this could give rise to unreliable predictions should the models be used in clinical practice. Future prognostic prediction models should comply with TRIPOD.

TwitterGustav Valentin Gade @gustavvalentin1

AcknowledgementsThe authors wish to thank health science librarian Permiile Skou Gaardsted for assisting with the development of the search strategy.

ContributorsGVG is the guarantor. GVG drafted the manuscript for the protocol and performed preliminary searches and search strategy. GVG, JRY, MGJ, TM, KT and SA developed selection criteria. JRI assisted GVG in screening titles, abstracts and reference lists of papers included after full text reading along with data extraction, assessing risk of bias, presence of meta-bias along with adherence to reporting guidelines. KT assisted GVG in full text reading. and MGJ was arbitrator if agreement could not be reached between reviewers. GVG drafted the manuscript for the paper, while MGJ, JRY, SA, TM, JRI and KT assisted in the interpretation of results, read, provided feedback and approved the final manuscript of the paper. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

FundingThe authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

DisclaimerThe study did not involve participants for whom the study results can be disseminated. The authors intend to distribute the results of the study by engaging with local healthcare providers for the general population setting during 2020 and 2021, along with social media to reach a broader audience.

Competing interestsNone declared.

Patient consent for publicationNot required.

Provenance and peer reviewNot commissioned; externally peer reviewed.

Data availability statementAll data relevant to the study are included in the article or uploaded as online supplemental information. The study protocol is available online at https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=124021. All information extracted for the included studies in the review is available as a data supplement.

Supplemental materialThis content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open accessThis is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs

Gustav Valentin Gade http://orcid.org/0000-0002-7944-9913

Jesper Ryg http://orcid.org/0000-0002-8641-3062

Johannes Riis http://orcid.org/0000-0001-8398-5420

Katja Thomsen http://orcid.org/0000-0002-0454-836X

Stig Andersen http://orcid.org/0000-0003-3632-5213

REFERENCES


