

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Predicting falls in community-dwelling older adults: a systematic review of prognostic models
AUTHORS	Gade, Gustav; Jørgensen, Martin; Ryg, Jesper; Riis, Johannes; Thomsen, Katja; Masud, Tahir; Andersen, Stig

VERSION 1 – REVIEW

REVIEWER	James Frith Newcastle University UK
REVIEW RETURNED	14-Oct-2020

GENERAL COMMENTS	<p>An important topic. Also a confusing area with many conflicting risk tools available. It is very useful to see a good quality systematic review evaluating existing tools. It is long over due in the field of falls.</p> <p>A few comments to address:</p> <p>Abstract</p> <ul style="list-style-type: none">- I would like to see the AUC for the validated models in the results section. It is an important finding that this is much lower than the development models. <p>Inclusion criteria</p> <ul style="list-style-type: none">- did all included participants have to be aged over 60, or did the median age of the study have to be over 60?- do you think it would be more appropriate to limit included studies to the most recent 5 or ten years? For example you include a study from 1994 when population, conditions and treatments were very different to how they are now (e.g. obesity, treatment of Cardiovascular diseases, increased multi morbidity and increased numbers of oldest old). Is this study relevant to today? Perhaps this could be a discussion point? <p>Search strategy</p> <ul style="list-style-type: none">- the variety of databases is robust- the search terms are appropriate- the overall search strategy is robust and sensitive- page 11 line 32 should use the word 'inclusion' in full <p>Results</p> <ul style="list-style-type: none">- in the characteristics of included studies table the term 'falls history' could be changed for extra clarity. Perhaps a more appropriate term would be 'Prior fall'. <p>Discussion</p> <ul style="list-style-type: none">- your scoring work around the term 'aged' feels out of place in the
-------------------------	---

	<p>discussion. It would be better placed in the methods. - can you re-phrase the opening sentence of the conclusion? It is not clear</p> <p>Other The pdf I have to review is 7750 pages long! There is a formatting issue from page 82 onwards</p>
--	---

REVIEWER	David Colquhoun University College London
REVIEW RETURNED	22-Nov-2020

GENERAL COMMENTS	<p>It is profoundly disappointing that out of 11,789 studies, only 30 were eligible for inclusion. It's also disappointing that, after 11,789 studies, all that can be concluded is</p> <p>"All models exhibited a high risk of bias rendering them unreliable for prediction in clinical practice."</p> <p>Nonetheless, it's important that this should be known and I therefore recommend acceptance.</p> <p>It might be worth adding a comment concerning the quality of research in this area, and perhaps some speculation as to the consequences of not including 99.7% of the studies that have been done. How much do the results depend on the very strong selection that has been done? That said, I'm not suggesting that the authors should repeat their study with different selection criteria, It seems very unlikely that increasing or decreasing the number of studies included would change the conclusions but it might be worth mentioning.</p> <p>I notice too that RCTs were excluded, It might be worth a comment in the discussion about the results from RCTs, if it's possible to come to any conclusions from them.</p>
-------------------------	---

VERSION 1 – AUTHOR RESPONSE

Reviewer comments:

Reviewer 1:

1. Abstract: I would like to see the AUC for the validated models in the results section. It is an important finding that this is much lower than the development models.
 - a. **Answer:**
 - i. We acknowledge this comment since validated models provide the most robust evidence on the predictive performance. The AUCs for the validated models have now been added to the abstract (results section) which now states: *"The area under the curve was available for 40 (55.6%) models, ranging from 0.49-0.87. Validated models' AUCs ranged from 0.62-0.69"*.

2. Inclusion criteria: Did all included participants have to be aged over 60, or did the median age of the study have to be over 60?
 - a. **Answer:**
 - i. We apologise if this was unclear. All participants had to be aged 60 years or above, as mentioned in the study protocol (appendix 1): *"Participants: Only studies with all participants aged 60 years or older will be included. This cut-off was chosen in order to encompass studies using different age cut-offs for*

being an older adult. Thus, studies with total age ranges extending below 60 years will be excluded. Also, mean age subtracted by 2 standard deviations must not extend below 60 years unless inclusion criteria specifically stated a lower age limit of 60 years or above". We have now further specified this in the manuscript (Eligibility criteria: Participants and setting), which now states: "All participants had to be community-dwelling, 60 years of age or older, and be recruited from a general population setting".

3. Inclusion criteria: Do you think it would be more appropriate to limit included studies to the most recent 5 or ten years? For example, you include a study from 1994 when population, conditions and treatments were very different to how they are now (e.g. obesity, treatment of Cardiovascular diseases, increased multi morbidity and increased numbers of oldest old). Is this study relevant to today? Perhaps this could be a discussion point?
 - a. **Answer:**
 - i. We agree that it is in general relevant to consider whether a limitation is necessary on study year. However, as the reviewer mentions, this is an area with many conflicting risk tools available. For this reason, we chose not to have any restrictions on study year to have a complete overview of the available evidence. We do agree that the study from 1994 (Maki et al.) may represent models more applicable in a time where the population, conditions, and treatments probably were different from today. However, when comparing the force platform predictors of this study to others, four more studies in our review (Boulgarides et al., Delbaere et al., Brauer et al., and Stalenhoef et al.) all included these, or similar, predictors. These studies were published from 2000 to 2010, which may support Maki et al. in 1994. None of the models in the abovementioned studies have been externally validated, for which reason we do not know whether these findings can provide reliable predictions today. As this is in line with our review's overall conclusion, we have not made changes to the manuscript regarding this point.
4. Search strategy: page 11 line 32 should use the word 'inclusion' in full
 - a. **Answer:**
 - i. We apologise for this abbreviation. This has now been rewritten in the manuscript (Data items) which now states: *"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 receiver-operating-characteristic Curve (AUC), internal and external validation procedures, final model presentation, and sources of funding."*
5. Results: in the characteristics of included studies table the term 'falls history' could be changed for extra clarity. Perhaps a more appropriate term would be 'Prior fall'.
 - a. **Answer:**
 - i. We agree, and the manuscript and Supplemental Table 1 (Characteristics of included studies) have now been rewritten using the term "prior falls" instead of "falls history". Thus, the abstract (Results) section now states: *"Most frequent falls predictors were prior falls, age, sex, measures of gait, balance and strength, along with vision and disability"*. Furthermore, the results (Index/Model) section now states: *"The most frequently applied predictors were prior falls, age, sex, measures of gait, balance, and strength, along with vision and disability"*.
6. Discussion: your scoring work around the term 'aged' feels out of place in the discussion. It would be better placed in the methods.
 - a. **Answer:**
 - i. We acknowledge this and have therefore moved this to the methods section (Search) which now states: *"We used a validated search string for prediction models. With the help of a research librarian in health science, we added the following terms to the search string: independent living, aged, and accidental falls. Details on the search string are available in data supplements (Appendix 3). No search filters were applied. We included "Aged" as a search term in the search string. Since this would restrict the number of search hits*

and thus the sensitivity of the search string, we pre-tested the search string without "Aged" in all databases before commencing the review. From this, the first 3,000 hits were screened independently and in duplicate, and we did not find studies not identified by the final search string. Thus, we believe this had a limited influence on the sensitivity of the search string."

7. Can you re-phrase the opening sentence of the conclusion? It is not clear
 - a. **Answer:**
 - i. We apologise for this and have now rewritten the opening sentence of the conclusion, which now states: *"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"*.
8. The pdf I have to review is 7750 pages long! There is a formatting issue from page 82 onwards
 - a. **Answer:**
 - i. We apologise for the inconvenience this has caused. This may be due to the appendices from page 82 and onwards have been uploaded as Excel-files, the intended format. However, ScholarOne has converted this to a PDF which may lead to the formatting issues mentioned. We have now formatted these appendices to fit a PDF format.

Reviewer: 2

9. It is profoundly disappointing that out of 11,789 studies, only 30 were eligible for inclusion. It's also disappointing that, after 11,789 studies, all that can be concluded is "All models exhibited a high risk of bias rendering them unreliable for prediction in clinical practice." Nonetheless, it's important that this should be known, and I therefore recommend acceptance. It might be worth adding a comment concerning the quality of research in this area, and perhaps some speculation as to the consequences of not including 99.7% of the studies that have been done. How much do the results depend on the very strong selection that has been done? That said, I'm not suggesting that the authors should repeat their study with different selection criteria. It seems very unlikely that increasing or decreasing the number of studies included would change the conclusions, but it might be worth mentioning.
 - a. **Answer:**
 - i. We agree that these findings are disappointing but relevant for researchers to consider when designing falls prediction studies. We were only able to include 30 out of 11,789 studies which may seem like strong selection. However, the number of search hits are a result of the search string's sensitivity, which has previously been estimated to be 95% (94-97%) for prediction modelling studies (Geersing et al., 2012, doi: 10.1371/journal.pone.0032844), along with the number of databases searched. We have provided rationales for each eligibility criterion in the study protocol (appendix 1) based on the current evidence on the prediction modelling methodology. Thus, we agree that this review's results depend on the selection process, but choices regarding these are justified. The included studies have been assessed for risk of bias and discussed, for which reason we consider the quality of this research area for prospective cohort studies to have been commented on in the discussion section of the paper. We acknowledge that other models may have been available based on other study designs. This aspect has been added to the Discussion (Limitations) section which now states: *"Furthermore, we excluded randomised controlled trials and retrospective cohort studies. Consequently, we were only able to include 0.3 percent (30/11,789) of studies screened even though other models, based on other study designs, may had 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 one specific study design, but we consider this exclusion of the other studies to be justified."*

10. I notice too that RCTs were excluded. It might be worth a comment in the discussion about the results from RCTs, if it's possible to come to any conclusions from them.

a. **Answer:**

i. Unfortunately, we are not able to comment on models derived from RCTs based on this systematic review. As prespecified in the study protocol (appendix 1), we did not include randomised controlled trials since these tend to have narrow predictor distributions resulting in poor discriminatory performance. Furthermore, predictive performance may also be influenced by treatment effects in the design (Moons et al. 2014, doi: 10.1371/journal.pmed.1001744), and generalisation to the target population may be compromised due to strict eligibility criteria (Moons et al., 2019, doi: 10.7326/M18-1377). Thus, we have not made changes to the manuscript regarding this point.

VERSION 2 – REVIEW

REVIEWER	James Frith Newcastle University UK
REVIEW RETURNED	08-Mar-2021
GENERAL COMMENTS	The few comments that I made have been addressed