

RESEARCH

Open Access



Prediction models for functional status in community dwelling older adults: a systematic review

Bastiaan Van Grootven and Theo van Achterberg*

Abstract

Background: Disability poses a burden for older persons, and is associated with poor outcomes and high societal costs. Prediction models could potentially identify persons who are at risk for disability. An up to date review of such models is missing.

Objective: To identify models developed for the prediction of functional status in community dwelling older persons.

Methods: A systematic review was performed including studies of older persons that developed and/or validated prediction models for the outcome functional status. Medline and EMBASE were searched, and reference lists and prospective citations were screened for additional references. Risk of bias was assessed using the PROBAST-tool. The performance of models was described and summarized, and the use of predictors was collated using the bag-of-words text mining procedure.

Results: Forty-three studies were included and reported 167 evaluations of prediction models. The median c-statistic values for the multivariable development models ranged between 0.65 and 0.76 (minimum = 0.58, maximum = 0.90), and were consistently higher than the values of the validation models for which median c-statistic values ranged between 0.6 and 0.68 (minimum = 0.50, maximum = 0.81). A total of 559 predictors were used in the models. The five predictors most frequently used were gait speed (n = 47), age (n = 38), cognition (n = 27), frailty (n = 24), and gender (n = 22).

Conclusions: No model can be recommended for implementation in practice. However, frailty models appear to be the most promising, because frailty components (e.g. gait speed) and frailty indexes demonstrated good to excellent predictive performance. However, the risk of study bias was high. Substantial improvements can be made in the methodology.

Keywords: Prediction, Disability, Functional status, Community, Aged, ADL

Introduction

Disability is a key outcome for public health [1]. It is associated with a decrease in quality of life, increased use of healthcare resources, institutionalization and

mortality in community dwelling older persons [2], and is therefore generally considered a primary outcome for intervention (e.g. strength and mobility programs). The high prevalence of multimorbidity and the onset of functional impairments in older persons, make this population particularly vulnerable for the development of disability in their instrumental and basic activities of daily living (ADL), e.g. shopping, walking, washing [3–5].

*Correspondence: theo.vanachterberg@kuleuven.be

Department of Public Health and Primary Care, KU Leuven, Leuven, Belgium



Numbers vary substantially, but the incidence of disability in instrumental and basic ADL in community dwelling older persons ranges between 5% and 59% [6, 7]. The incidence is typically higher for instrumental ADL than for basic ADL [8], and also increases with age [9].

Decades of research have focused on identifying risk factors for disability in ADL in community dwelling older persons. Key risk factors include, among others, multimorbidity, frailty, cognition, depression, body mass index, physical activity, and sensory and physical impairments [10–14]. If modifiable, these factors can be the focus of interventions. Alternatively, they can be used to identify an individual's risk for disability or predict a score on a disability scale when incorporated in a prediction model. Prediction models can inform persons about their individual prognosis (risk), can support older persons and healthcare professionals in the decision-making process, and can inform research designed to explore subgroups that respond better (or worse) to interventions that aim to improve functional status or prevent disability.

One systematic review, published in 2015, has previously investigated the utility of clinical prediction models for the outcome functional decline [15]. This review included 16 models in the evaluation and observed areas under the curve ranging between 0.63 and 0.78, thus indicating poor to moderate predictive performance of the models. However, this review had some limitations, e.g. only including short case finding instruments, only including instruments with validation, only considering decline in ADL as outcome. Also, a large number of studies have since been published. A second review investigated the association between frailty indicators and disability, and observed an important association between gait speed and disability [9].

We therefore performed a systematic review to identify models developed for the prediction of functional status in community dwelling older persons. We investigated the types of predictors that were included in the models, and how well the models predicted functional status.

Methods

A protocol for the systematic review was drafted before the start of the study (see appendix 1) and the PRISMA statement was used to structure the report of the study. [16]

Eligibility criteria

Studies had to include older persons, indicated by a mean sample age of 65 years or older, with the majority of the sample living at home.

Prediction models described in the studies had to predict the outcome functional status, defined as the ability to perform instrumental or basic ADL, and could

include ADL scales (e.g. Katz scale) or specific aspects of ADL (e.g. washing or mobility). Physical performance outcomes were considered relevant if the reported data related to daily activities, e.g. ability to mobilise. Physical performance related to strength or speed was not considered for inclusion.

Predictions could include the prediction of ADL (scale or item score), or decline, maintenance, recovery or improvement in ADL (scale or item). The prediction model had to measure a single characteristic (univariable model) or a set of characteristics (multivariable model) to estimate a person's individual prognosis and could include patient, care outcome and care process factors. The models could be presented in any format, e.g. as a statistical model, regression formula with coefficients, web or electronic application, nomogram, or score chart.

Studies with binary or survival outcomes had to report the concordance (c) statistic (which is equal to the area under the curve). Studies with continuous outcomes had to report the R2 statistic.

We included nested case-control studies, prospective and retrospective cohort studies (including database and registry studies), and secondary analyses of trials.

Information sources

We searched the Medline and EMBASE databases from inception up to March 2022 for eligible studies. After selecting the full text manuscripts, we screened references lists and prospective citations (using Google Scholar) for eligible studies. Lastly, the ICTPR portal was searched for protocols and Web of Science for conference proceedings in order to track full text manuscripts.

Search

A search string was drafted using a combination of free text words and MeSH terms. The search terms were grouped according to outcome, prediction models, and setting. We used a validated search string for the terms related to the prediction models [17]. The terms related to the outcome and setting were derived from other systematic reviews, entry terms related to MeSH term, thesaurus searches for synonyms, and key words from published manuscripts. The final search string was adapted to the EMBASE database. No limits were used for the search. The final search string is available in appendix 2.

Study selection

Identified records were collected in an Endnote database. One author screened all titles and abstracts for inclusion in two stages: screening titles and abstracts, and reading full text manuscripts. The second author verified the final inclusion using a standardised checklist.

The final selection was based on a consensus decision. Authors were not blinded to the manuscripts' citation information.

Data collection process

Data were collected in an Excel database. The data collection was first piloted in order to standardise the collection process and define uniform terms. The data collection was performed by one author. Data were verified a second time by the same author.

Data items

The following information was collected: study citation, country, design, setting, sample characteristics, sample size, outcome definitions and measurements, statistical analyses, definition of predictors, purpose and design of prediction models, and the evaluation and performance of prediction models.

Risk of bias

The PROBAST tool (Prediction model Risk Of Bias ASsessment Tool) was used to assess the risk of bias for the participants, predictors, outcome and analysis for each model [18]. A standardised questionnaire was used to rate the risk of bias as 'yes', 'probably yes', 'no', 'probably no', or 'no information'. An overall judgement was made as either low risk of bias, high risk of bias or unclear risk of bias. One author assessed the risk of bias.

Summary measures

We investigated the discrimination and calibration for models with a binary and survival outcome, and the R2 measure and calibration for models with a continuous outcome [19]. The discrimination was expressed using the concordance (c) statistic. The c statistic is equivalent to the area under the curve (for binary outcomes), and the following values were used in the interpretation of the performance: < 0.7 is poor, 0.7 – 0.8 is moderate, 0.8 – 0.9 is good, > 0.9 is very good [20]. The performance of linear models was measured using R2. The calibration could be measured using different statistics: observed versus expected events, calibration slope, calibration in large, calibration plot, or the Hosmer-Lemeshow test.

Synthesis of results

Summary tables were drafted to describe the study characteristics, risk of bias and findings of the studies. We originally planned to construct funnel plots to visualize the different performance measures (R2, discrimination, calibration). However, the majority of the studies did not report standard errors or sufficient information to construct confidence intervals. We changed our strategy and described the distribution using scatter plots to visualize

the discrimination and R2 of the individual studies, and constructed box plots to describe the central tendency (median) and spread of the data (interquartile range). We further observed substantial differences between studies and models and therefore decided to describe the results within subgroups based on the 1) definition of the outcome (ADL, IADL, ADL or IADL, mobility or disability), 2) type of summary measure (c statistic or R2), 3) type of model (univariable or multivariable), 4) type of model evaluation (development or validation). We used the 'bag of words' text mining procedure to summarise the clinical predictors included in the models. This procedure was used to create a word cloud to describe the frequencies of the predictors. Analyses were performed using R studio (using the 'dplyr', 'ggplot2', 'Hmisc', 'tm' and 'wordcloud' packages).

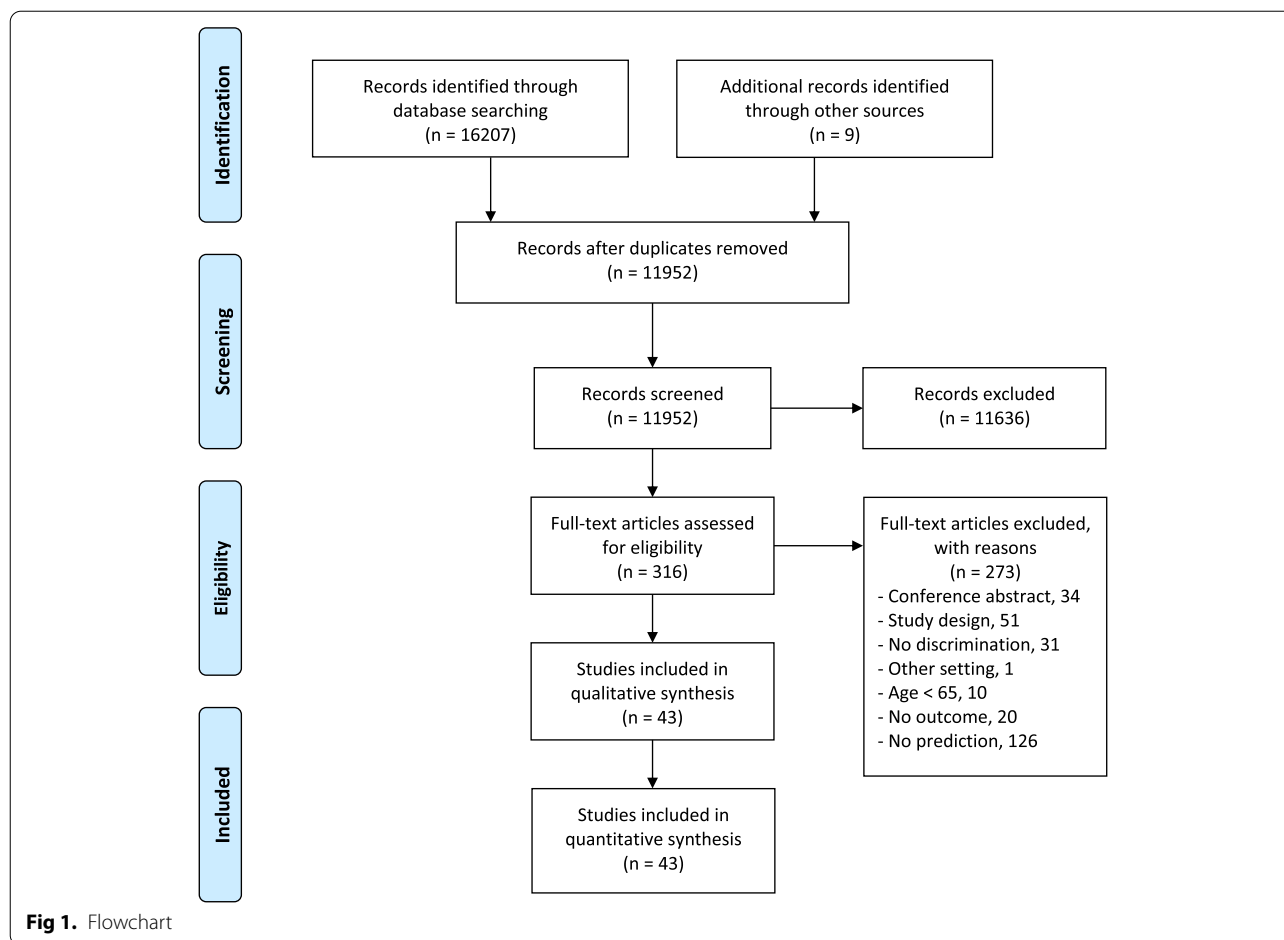
In addition, we described individual models that demonstrated a good performance in a validation cohort (c statistic > 0.8, R2 > 0.5) separately using a narrative synthesis.

Results

A total of 11952 titles and abstracts, and 316 full text manuscripts were screened for inclusion. A total of 34 studies were retained for inclusion. [6, 7, 21–52] An additional nine studies were identified through secondary sources [53–61], resulting in a final sample of 43 studies (see Figure 1).

Study characteristics

The majority of studies originated from North America ($n = 19$) or Europe ($n = 15$), and to a lesser extent from Asia ($n = 7$), or South America ($n = 1$) or Africa ($n = 1$). All but one study used data from a prospective cohort study; the one study using data from a randomized controlled trial. The median age across the samples was 76 years, with a minimum of 66 and a maximum of 85. Eighteen (43%) of the studies recruited persons who were independent in ADL or mobility at baseline. Basic ADL was the most prevalent outcome ($n = 15$), followed by Instrumental ADL (IADL, $n = 8$), mobility ($n = 5$), disability ($n = 4$), or a composite outcome of ADL and IADL ($n = 4$). The remaining studies had multiple outcome measurements of ADL, IADL, mobility, or bathing and dressing. Almost all studies evaluated predictions of functional decline ($n = 39$), two studies predicted a change in outcome score, and two studies predicted the outcome score on a continuous scale. The median incidence of functional decline across the studies was 20%, with a minimum of 5% and a maximum of 59%. The median time to follow-up across the studies was two years, with a minimum of half a year and a maximum of nine years. The 43



studies included 167 evaluations of prediction models. The results will be reported at the level of the 167 evaluations of prediction models. The characteristics are described in table 1.

Risk of bias

There was a low risk of bias in 13 model evaluations, high risk of bias in 135 evaluations, and there was uncertainty in the remaining 19 evaluations (see Figure 2). The majority of the model evaluations had a low risk of bias on the domains related to measuring predictors ($n = 157$) and outcomes ($n = 130$). However, the majority of evaluations had a high risk of bias on the domains related to recruiting participants ($n = 86$) and analyses ($n = 158$). A clear description of the recruitment was often missing or only participants with complete data were selected from the cohort for analysis. Predictors were often selected based on p-values without accounting for overfitting or optimism in the performance, and the influence of right censoring (e.g. due to death) was not estimated.

Prediction models

Of the 167 evaluations, 62 were univariable model evaluations, of which 58 estimated the risk for functional decline, and four validated previously determined cut-off criteria. Sixty-seven evaluations of multivariable models were performed, of which twelve were also validated by the authors who developed the model. Thirty-eight multivariable model evaluations were external validations by an independent research team. The median sample size across the models was 1198, with a minimum of 83 and a maximum of 27220.

Performance of models

In the evaluations (see Figure 3), the median c statistic values for the multivariable development models ranged between 0.65 and 0.76 (minimum = 0.58, maximum = 0.90), and were consistently higher than the values of the validation models for which median c statistic values ranged between 0.6 and 0.68 (minimum = 0.50, maximum = 0.81). The values of the univariable models tended to be similar to those of the multivariable models

Table 1. Study characteristics

| Author | Year | Country | Design | Population | Age ^a | Aim | n | Outcome, FU ^b |
|------------------|------|-------------|--|--|------------------|---------------------------------|-----------|---|
| Adachi [17] | 2018 | Japan | Prospective cohort | Older people walking independently | 79 [76 - 82] | Evaluate factor | 518 | Mobility limitation, 2 |
| Adachi [18] | 2018 | Japan | Prospective cohort | Women 75 or older walking independently | 79 [77 - 82] | Evaluate factor | 330 | Mobility limitation, 2 |
| Aliberti [19] | 2018 | USA | Prospective cohort | 65 or older without dementia or ADL dependence | 74.4 (7.0) | Develop model Validate index | 7388 7388 | ADL dependence, 2, |
| Arnau [20] | 2016 | Spain | Prospective cohort | 75 or older without severe dependence | 81.7 (4.6) | Develop model | 252 | IADL or ADL decline, 1 |
| Ben-Shalom [21] | 2016 | USA | Prospective cohort linked with administrative database | Medicare beneficiaries 65 or older | | Develop model | 10057 | ADL dependence, 4 |
| Bongue [48] | 2017 | Canada | Prospective cohort | 65 or older | 78.7 (7.9) | Validate index | 1224 | Disability, 2 |
| Brach [22] | 2012 | USA | Prospective cohort | 65 or older walking independently | 79.4 (4.1) | Evaluate factor | 339 | Mobility limitation, 1 |
| Carrière [49] | 2005 | France | Prospective cohort | 75 or older independent in IADL | 79 [76 - 81] | Develop model Validate model | 545 807 | IADL dependence, 7 |
| Clark [50] | 2012 | USA | Prospective cohort | 65 or older independent in ADL | 74.4 (7.2) | Develop model Validate model | 6233 3213 | ADL dependence, 2 |
| Clark [6] | 2015 | USA | Prospective cohort | 65 or older independent in ADL | 74.1 (6.7) | Develop model Validate model | 5332 2763 | ADL dependence, 2 |
| Classon [23] | 2016 | Sweden | Prospective cohort | 85 or older | | Evaluate factor | 83 | IADL decline, 5 |
| Covinsky [24] | 2006 | USA | Prospective cohort | 70 or older independent in ADL | 76.8 (5.3) | Develop model Validate model | 3245 1994 | ADL dependence, 2 |
| Deckx [7] | 2015 | Netherlands | Prospective cohort | 70 or older with and without cancer | 78.1 (5.5) | Validate index | 134 220 | ADL decline, 1 |
| Dixon [48] | 2021 | USA | Prospective cohort | 65 or older | 75.4 (6.1) | Develop model | 93 | IADL dependence, 1.5 |
| Donoghue [25] | 2014 | Ireland | Prospective cohort | 50 or older | 72.8 (6.1) | Evaluate factor | 1664 | ADL dependence, 2 IADL dependence, 2 |
| Ensrud [51] | 2009 | USA | Prospective cohort | 67 or older men walking independently | 76.4 (5.6) | Validate index | 3132 | IADL dependence, 3.2 |
| Faurot [26] | 2015 | USA | Prospective cohort linked with administrative database | 65 or older | | Develop model | 6391 | ADL dependence, 4 |
| Gill [27] | 1997 | USA | Prospective cohort | 72 or older independent in ADL | 78.5 (5.2) | Validate index | 1813 | ADL dependence, 1 |
| Gobbens [52] | 2012 | Netherlands | Prospective cohort | 75 or older | 80.3 (3.8) | Update index | 479 | Disability, 2 |
| Guralnik [53] | 2000 | USA | Prospective cohort | 65 or older without disability | | Evaluate factor | 6534 | Disability, 1-4 |
| Hegendorfer [28] | 2019 | Belgium | Prospective cohort | 80 or older | 84.7 (3.7) | Validate index | 560 | ADL decline, 1.7 |
| Hong [29] | 2016 | South Korea | Prospective cohort | 60 or older | 72.5 (5.5) | Evaluate factor | 8000 | IADL decline, 3 |
| Ishimoto [30] | 2010 | Japan | Prospective cohort | 65 or older | 75.5 (5.9) | Validate index | 518 | ADL decline, 1 |

Table 1. (continued)

| Author | Year | Country | Design | Population | Age ^a | Aim | n | Outcome, FU ^b |
|-------------------|------|---|---------------------------|---|------------------|--|------------|---|
| Jonkman [31] | 2019 | Germany, United Kingdom, Italy, Netherlands | 4 Prospective cohorts | 65 - 75y at baseline without limitations | 69.7 (3.0) | Develop model Validate model | 2560 2560 | ADL dependence, 3 |
| Jonkman [32] | 2018 | Italy, Netherlands | 2 Prospective cohorts | 60 - 70y at baseline | 67.5 (2.1) | Develop model | 312 | IADL or ADL decline, 9 |
| Lam [33] | 2020 | Hong Kong | Prospective cohort | 65 or older walking independently | 72.5 (5.3) | Validate index | 1566 | Physical limitations, 4 |
| Lin [34] | 2004 | Taiwan | Prospective cohort | 65 or older | 73.4 | Evaluate factor | 1200 | ADL decline, 1 |
| McClintock [35] | 2018 | United States | Prospective cohort | Medicare beneficiaries 65 or older | | Develop model Validate model | 12758 8506 | IADL decline, 2 |
| Nuesch [36] | 2015 | UK | 2 Prospective cohorts | Older adults without locomotor disability | 61 - 68 | Develop model Validate model | 2377 3194 | Locomotor disability, 7-8 |
| Onder [54] | 2005 | USA | Prospective cohort | Older women with functional disability | 78.7 (8.0) | Evaluate factor | 484 684 | ADL dependence, 3 Mobility limitation, 3 |
| Op het Veld [55] | 2019 | Netherlands | Prospective cohort | 65 or older who are (pre)frail | 76.3 (6.6) | Validate index | 2420 | IADL or ADL decline, 2 |
| Papacristou [37] | 2017 | UK | Prospective cohort | Men between 71 - 92 | 8.0 (4.4) | Evaluate factor | 1198 | Mobility limitation, 3 |
| Perera [38] | 2015 | International | 7 Prospective cohorts | Older persons | 73.9 (5.3) | Evaluate factor | 27220 | Bathing/dressing dependency, 3 Mobility limitation, 3 |
| Saraiva [39] | 2020 | Brazil | Prospective cohort | 60 or older | 80 (12) | Validate index | 317 | ADL decline, 1 |
| Sarkisian [40] | 2000 | USA | Prospective cohort | Women 65 or older | 73 | Develop model Validate model | 4421 2211 | IADL decline, 4 |
| Spalter [41] | 2013 | Israel | Prospective cohort | 60 or older | 70.9 | Develop model | 982 | Change in mobility, 5 |
| Studenski [56] | 2003 | USA | 3 Prospective cohorts | 65 or older | 74.1 (5.7) | Evaluate factor Validate index Develop model | 974 | IADL or ADL decline, 1 |
| Suijker [42] | 2014 | Netherlands | Prospective cohort | 70 or older | 76.1 [8.4] | Update index Validate index | 644 2085 | ADL decline, 1 |
| Tas [43] | 2011 | Netherlands | Prospective cohort | 55 or older | > 65 | Develop model | 5027 | Mild disability, 6 |
| Teo [44] | 2017 | Signapore | Prospective cohort | 55 or older independent in ADL | 66.1 (7.6) | Validate index | 2406 | IADL dependence, 2 ADL dependence, 2 |
| Terhorst [45] | 2017 | USA | Prospective cohort | Women 70 or older | 78.9 (5.1) | Develop model | 256 | Mobility limitation, 0.5 ADL dependence, 0.5 IADL dependence, 0.5 |
| Wennie Huang [46] | 2010 | USA | Prospective cohort | 65 or older independent in ADL | 80.3 (7.0) | Evaluate factor | 110 | ADL dependence, 0.5 - 1.5 |
| Yam [47] | 2013 | USA | Secondary analysis of RCT | 65 or older | - | Develop model | 582 | IADL function, 5 |

ADL Activities of Daily Living, IADL Instrumental Activities of Daily Living, ^a data is reported as mean (standard deviation) or median [interquartile range]; ^b FU = Follow-up, is reported as number of years.

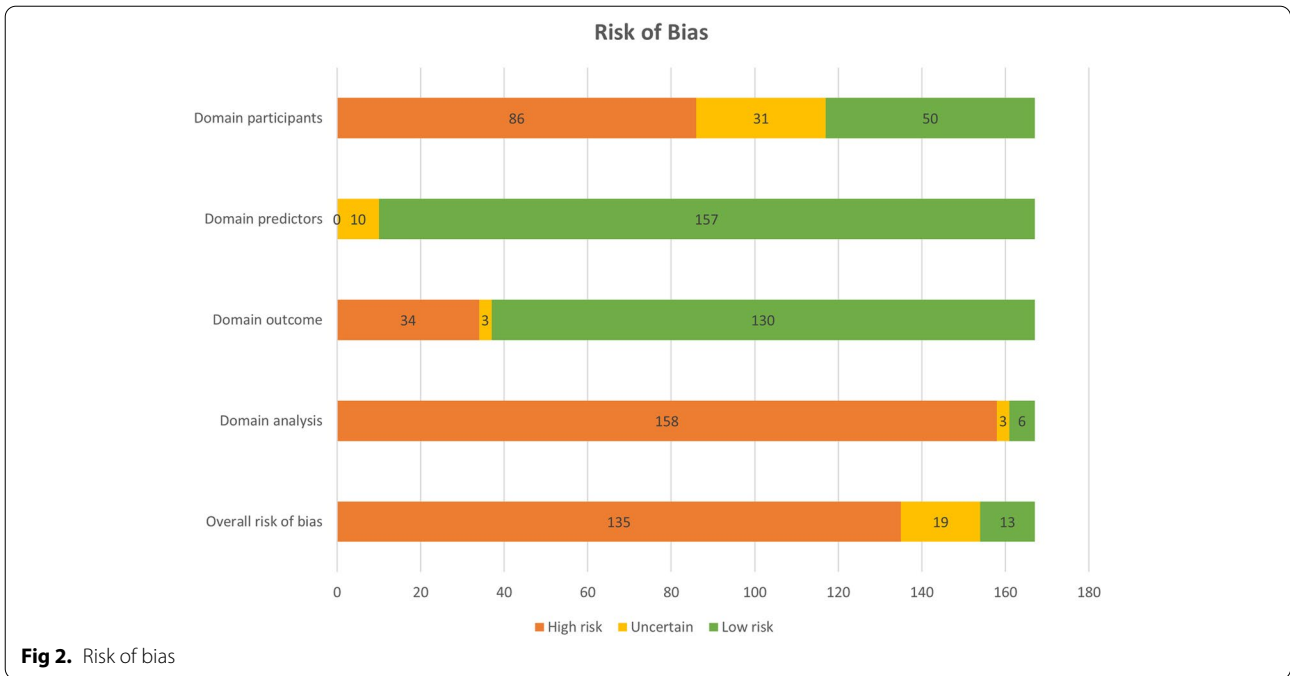


Fig 2. Risk of bias

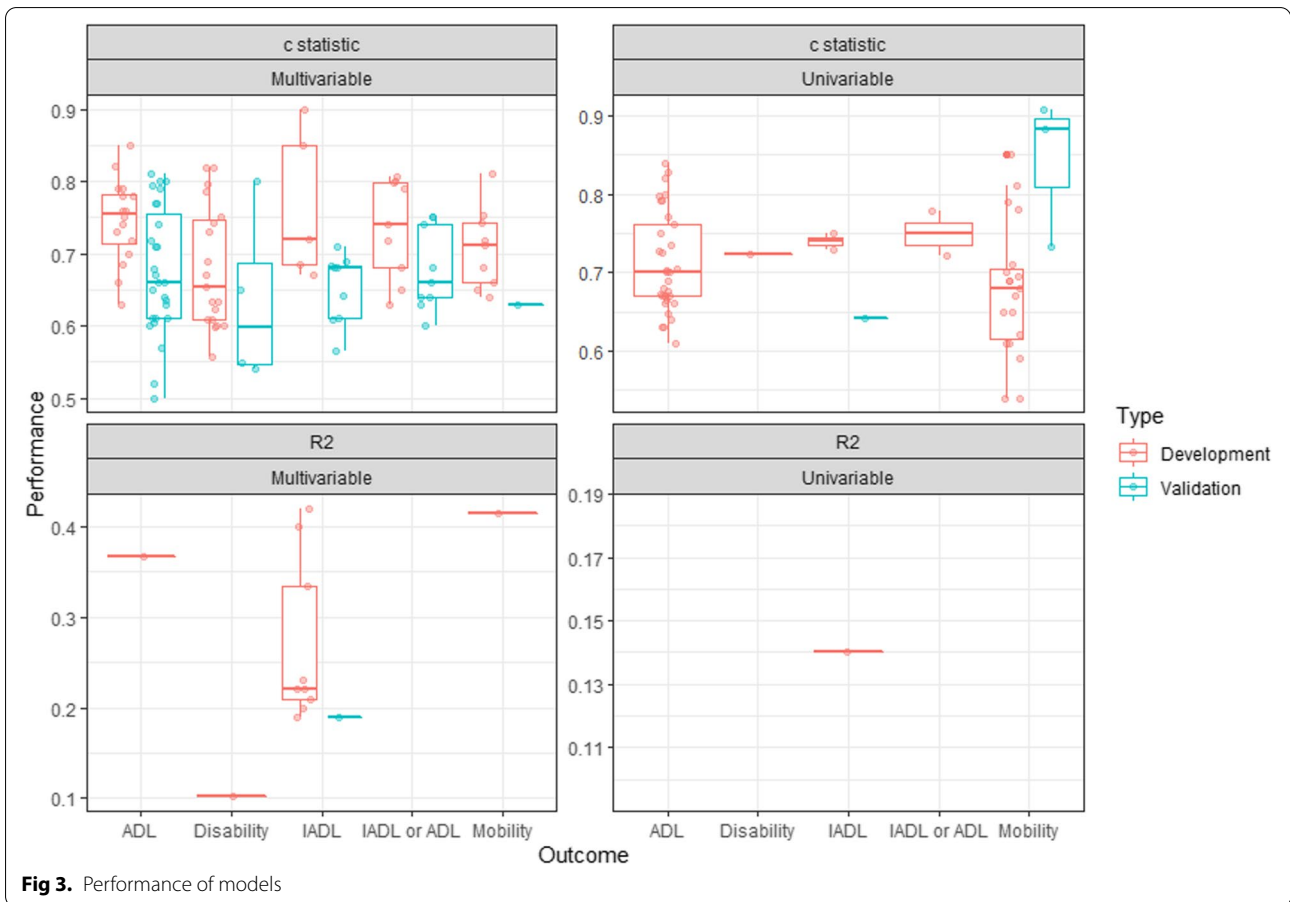


Fig 3. Performance of models

(median values ranging between 0.68 and 0.75 (minimum = 0.54, maximum = 0.85) for development, and ranging between 0.64 and 0.89 (minimum = 0.64, maximum = 0.91) for validation).

The R2 values of the multivariable development models ranged between 0.10 and 0.42, but only the outcome IADL had sufficient observations to estimate the distribution parameters. The performance of the multivariable models appeared to be similar for the different outcomes. The calibration was measured in nine evaluations. Six evaluation reported the expected versus observed events, two used a calibration plot, and one evaluation estimated the calibration slope and intercept.

Four models demonstrated a good performance in a validation cohort. However, the risk of bias was high in the first three models, and uncertain in the last model. Clark et al. included nine baseline predictors in a model to predict ADL dependence and validated the model at two years follow-up (c statistic = 0.80) [6]. Gobbens et al. evaluated the Tilburg Frailty Index (15 predictors) for the prediction of disability measured using the Groningen Activity Restriction index at two years follow-up (c

statistic = 0.8) [57]. Ishimoto et al. evaluated a 21-item fall risk index for the prediction of ADL decline at one year follow-up (c statistic = 0.8) [34]. Teo et al. validated a social and physical frailty index (18 predictors) for the prediction of severe ADL disability at half a year follow-up (c statistic 0.81) [48].

Two univariable models demonstrated a good to excellent performance in a validation cohort. However, the risk of bias was uncertain for both models. Both models were evaluated in the same cohort by Adachi et al., which observed that a gait speed of <0.8m/s had an excellent discrimination (c statistic 0.91), and a gait speed of <1.0m/s had a good discrimination (c statistic 0.89) for the outcome mobility limitations at two years follow-up [22].

Predictors

The median number of predictors in the multivariable models that were evaluated was seven, with a minimum of two and a maximum of 26. Six model evaluations included longitudinal repeated measurements

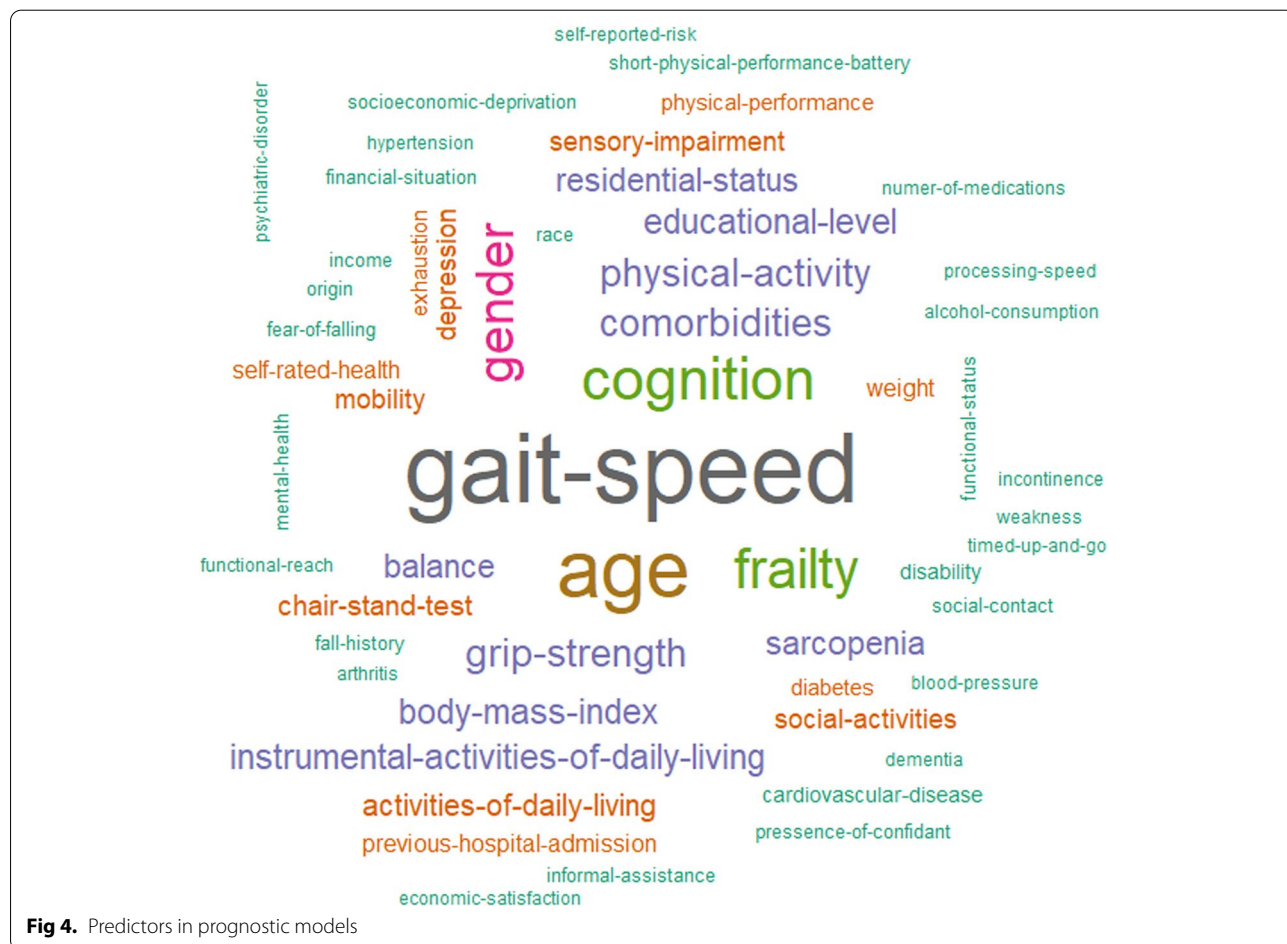


Fig 4. Predictors in prognostic models

of predictors, with the remaining models only included baseline measurements.

A total of 559 predictors were recorded, with 55 predictors being used in at least three evaluations (see Figure 4; note that the larger words indicate a higher frequency of use of the predictor in the models). The fifteen predictors most frequently used were gait speed ($n = 47$), age ($n = 38$), cognition ($n = 27$), frailty ($n = 24$), gender ($n = 22$), comorbidity ($n = 15$), grip strength ($n = 15$), physical activity ($n = 15$), body mass index ($n = 15$), IADL ($n = 13$), balance ($n = 12$), educational level ($n = 12$), residential status ($n = 12$), sarcopenia ($n = 12$), and ADL ($n = 10$).

Discussion

This review evaluated the state of the art for models that aim to predict future functional status in older persons. We identified the models, and summarised their performance, as well as the predictors that were used. Self-reported (in)dependence on activities of daily living was the most prevalent outcome for the prediction of functional status. The performance of the models varied substantially, ranging from poor to very good, but was moderate to low on average. The performance of the prediction models was generally lower in validation cohorts.

Multivariable models appeared to be slightly better for the prediction of future functional status than univariable models, but this was difficult to assess because most univariable models lacked external validation. Gait speed was the most prevalent predictor, in particular for univariable predictions, and demonstrated a moderate predictive performance (median c statistic = 0.70, minimum = 0.64, maximum = 0.91, data not shown). The majority of studies (95%) did not evaluate the calibration of the model.

Our results are in line with the previous review on prediction models for community dwelling older adults [9, 13], that also observed a poor to moderate performance of prediction models for the outcome functional decline. Our review included more and different outcomes, but conclusions are the same across the different outcome measurements. We also included substantially more and more recent models, but the state of the art does not appear to have improved. To date, no model appears to be ready for implementation in clinical practice. However, frailty models and gait speed measurements appear to be the most promising. Nonetheless, the high risk of bias for many studies, including those on frailty models, is a particular concern. Most studies ignored missing data and censoring of outcomes, which can bias the regression coefficients and ultimately lead to poorly calibrated models at the population level. Furthermore, calibration was only investigated in 5% of the models. Nonetheless, this

is a key measure for prediction models as it assesses if the predicted outcome (probability or score) corresponds to the observed outcome. Poorly calibrated models over or under predict the probability and can potentially harm the subsequent decision-making process [62].

Specific suggestions can be made to improve the body of evidence on the prediction of functional status. The majority of studies relied on statistical significance to select predictors, which can result in overfitting and overly optimistic performance measures for the developed model [63]. This optimism is often not detected because most models are never validated. This could be mitigated by finding consensus on a core set of prognostic factors for functional status. For example, stacked regression could be used to derive predictors from a combination of different prediction models, i.e. the analysis would find the ideal combination of predictors across the models [64]. If new predictors are tested, and if testing relies on statistical selection, than penalised regression models, e.g. lasso-regression, is preferred; or a shrinkage factor should be applied to minimise the optimism as a result of predictor selection. These methods were not used in any of the included studies. An important discussion should also be the appropriate selection of statistical methods. The predominant method, logistic regression, does not account for censored observations (e.g. due to death) and disregards the missing data. Furthermore, the design of cohort studies have an inherent risk for attrition bias, i.e. that patients who experience disability or die are also more likely to have missing values because they drop-out of the study. The association between disability and death is worrisome, because it is conceivable that persons who have died would have experienced a different disability trajectory than persons who survived. A model that ignores this will have biased coefficients and therefore predictions [65]. These 'informative right censoring' assumptions should at least be tested in samples with loss to follow-up, e.g. using a joint model strategy. Lastly, clinical usefulness should also be part of the evaluation of a model that is well calibrated. Classification plots with the area under the curve and classification measures (e.g. sensitivity, specificity) for different potential cut-off values should be preferred over ROC curves [66].

The results of this review are somewhat discouraging. Nonetheless, the burden of disability remains high and will be an important driver for increasing long-term healthcare costs [67]. We believe that the identification of persons who could benefit from interventions designed to prevent disability therefore remains a worthwhile public health strategy.

Lastly, it is important that the impact of models is evaluated in practice. We have additionally searched for studies that implemented disability prediction models

in the community setting, but we could not identify any references. However, evaluating if the introduction of a prediction model changes care, e.g. increases physical therapy interventions, and improves outcome, e.g. reduced incidence of disability, will be an important future investigation.

Limitations

Some limitations should be noted. Although both authors screened the included studies independently, titles and abstracts were only screened by one author. However, we included a large number of studies and model which makes it likely that we have a sample that is representative from population of studies.. One author collected the data, and although this process was double-checked, some data abstraction mistakes may have been made. Further, we inferred the performance of the models based on the distribution (median, interquartile range and range), but these should not be considered pooled results. The models, and their evaluation, differed substantially from each other and we did not consider it relevant nor appropriate to perform a meta-analysis.

Conclusion

Currently available models for the prediction of future functional status in older persons have a low to moderate performance on average. Though multivariable models perform slightly better than univariable models, high risks of bias in the evaluations of prediction models do not allow for any firm conclusions. There is currently no model that can be recommended for implementation in practice, but frailty models appear to be the most promising. Substantial improvements can be made in the methodology of developing and validating prediction models for disability in the community setting.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12877-022-03156-7>.

Additional File 1. Prediction Models for Functional Status in Community Dwelling Older Adults: Review Protocol, Search Strategy and Prisma Checklist.

Acknowledgements

Not applicable

Authors' contributions

All authors designed the study and wrote the protocol. All authors were involved in the study selection, and BVG collected and analysed the data. All authors interpreted the data. BVG wrote the manuscript and TvA critically revised the manuscript. All authors read and approved the final manuscript.

Funding

The systematic review was performed without project funding.

Availability of data and materials

The dataset supporting the conclusions of this article is available upon request. Please contact bastiaan.vangrootven@kuleuven.be.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests

Received: 13 October 2021 Accepted: 12 May 2022

Published online: 30 May 2022

References

- Guralnik JM, Fried LP, Salive ME. Disability as a public health outcome in the aging population. *Annu Rev Public Health*. 1996;17:25–46.
- Fried LP, Guralnik JM. Disability in older adults: evidence regarding significance, etiology, and risk. *J Am Geriatr Soc*. 1997;45(1):92–100.
- Ofori-Asenso R, Chin KL, Curtis AJ, Zomer E, Zoungas S, Liew D. Recent Patterns of Multimorbidity Among Older Adults in High-Income Countries. *Popul Health Manag*. 2019;22(2):127–37.
- Jindai K, Nielson CM, Vorderstrasse BA, Quiñones AR. Multimorbidity and Functional Limitations Among Adults 65 or Older, NHANES 2005–2012. *Prev Chronic Dis*. 2016;13:E151.
- Verbrugge LM, Jette AM. The disablement process. *Soc Sci Med*. 1994;38(1):1–14.
- Clark DO, Stump TE, Tu W, Miller DK. Improving the validity of activity of daily living dependency risk assessment. *J Appl Gerontol*. 2015;34(3):329–42.
- Deckx L, van den Akker M, Daniels L, De Jonge ET, Bulens P, Tjan-Heijnen VC, et al. Geriatric screening tools are of limited value to predict decline in functional status and quality of life: results of a cohort study. *BMC Fam Pract*. 2015;16:30.
- Rajan KB, Hebert LE, Scherr PA, Mendes de Leon CF, Evans DA. Disability in Basic and Instrumental Activities of Daily Living is Associated with Faster Rate of Decline in Cognitive Function of Older Adults. *The Journals of Gerontology: Series A*. 2012;68(5):624–30.
- Sjölund B-M, Wimo A, Engström M, von Strauss E. Incidence of ADL Disability in Older Persons, Physical Activities as a Protective Factor and the Need for Informal and Formal Care-Results from the SNAC-N Project. *PLoS ONE*. 2015;10(9):e0138901.
- Stuck AE, Walthert JM, Nikolaus T, Büla CJ, Hohmann C, Beck JC. Risk factors for functional status decline in community-living elderly people: a systematic literature review. *Soc Sci Med*. 1999;48(4):445–69.
- Vermeulen J, Neyens JC, van Rossum E, Spreeuwenberg MD, de Witte LP. Predicting ADL disability in community-dwelling elderly people using physical frailty indicators: a systematic review. *BMC geriatr*. 2011;11:33.
- van der Vorst A, Zijlstra GA, Witte N, Duppen D, Stuck AE, Kempen GI, et al. Limitations in Activities of Daily Living in Community-Dwelling People Aged 75 and Over: A Systematic Literature Review of Risk and Protective Factors. *PLoS ONE*. 2016;11(10): e0165127.
- Ryan A, Wallace E, O'Hara P, Smith SM. Multimorbidity and functional decline in community-dwelling adults: a systematic review. *Health Qual Life Outcomes*. 2015;13:168.
- Beaton K, McEvoy C, Grimmer K. Identifying indicators of early functional decline in community-dwelling older people: a review. *Geriatr Gerontol Int*. 2015;15(2):133–40.
- O'Caomh R, Cornally N, Weathers E, O'Sullivan R, Fitzgerald C, Orfila F, et al. Risk prediction in the community: A systematic review of case-finding instruments that predict adverse healthcare outcomes in community-dwelling older adults. *Maturitas*. 2015;82(1):3–21.

16. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*. 2009;339: b2535.
17. Geersing GJ, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M, Moons KG. Search filters for finding prognostic and diagnostic prediction studies in Medline to enhance systematic reviews. *PLoS ONE*. 2012;7(2): e32844.
18. Wolff RF, Moons KGM, Riley RD, Whiting PF, Westwood M, Collins GS, 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.
19. Steyerberg EW, Vickers AJ, Cook NR, Gerdts T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology*. 2010;21(1):128–38.
20. Polo TCF, Miot HA. Use of ROC curves in clinical and experimental studies. *J Vasc Bras*. 2020;19:e20200186.
21. Adachi T, Kamiya K, Kono Y, Iwatsu K, Shimizu Y, Honda I, et al. Predicting the Future Need of Walking Device or Assistance by Moderate to Vigorous Physical Activity: A 2-Year Prospective Study of Women Aged 75 Years and Above. *Biomed Res Int*. 2018;2018:1340479.
22. Adachi T, Kamiya K, Kono Y, Iwatsu K, Shimizu Y, Honda I, et al. Estimation of reduced walking speed using simple measurements of physical and psychophysiological function in community-dwelling elderly people: a cross-sectional and longitudinal study. *Aging Clin Exp Res*. 2019;31(1):59–66.
23. Aliberti MJR, Cenzer IS, Smith AK, Lee SJ, Yaffe K, Covinsky KE. Assessing Risk for Adverse Outcomes in Older Adults: The Need to Include Both Physical Frailty and Cognition. *J Am Geriatr Soc*. 2019;67(3):477–83.
24. Arnau A, Espauella J, Serrarols M, Canudas J, Formiga F, Ferrer M. Risk factors for functional decline in a population aged 75 years and older without total dependence: A one-year follow-up. *Arch Gerontol Geriatr*. 2016;65:239–47.
25. Ben-Shalom Y, Stapleton DC. Predicting disability among community-dwelling medicare beneficiaries using claims-based indicators. *Health Serv Res*. 2016;51(1):262–81.
26. Brach JS, Wert D, VanSwearingen JM, Newman AB, Studenski SA. Use of stance time variability for predicting mobility disability in community-dwelling older persons: a prospective study. *J Geriatr Phys Ther*. 2012;35(3):112–7.
27. Classon E, Fallman K, Wressle E, Marcusson J. Relations between Concurrent Longitudinal Changes in Cognition, Depressive Symptoms, Self-Rated Health and Everyday Function in Normally Aging Octogenarians. *PLoS ONE*. 2016;11(8): e0160742.
28. Covinsky KE, Hilton J, Lindquist K, Dudley RA. Development and validation of an index to predict activity of daily living dependence in community-dwelling elders. *Med Care*. 2006;44(2):149–57.
29. Donoghue OA, Savva GM, Cronin H, Kenny RA, Horgan NF. Using timed up and go and usual gait speed to predict incident disability in daily activities among community-dwelling adults aged 65 and older. *Arch Phys Med Rehabil*. 2014;95(10):1954–61.
30. Faurot KR, Jonsson Funk M, Pate V, Brookhart MA, Patrick A, Hanson LC, et al. Using claims data to predict dependency in activities of daily living as a proxy for frailty. *Pharmacoevidemiol Drug Saf*. 2015;24(1):59–66.
31. Gill TM, Williams CS, Richardson ED, Berkman LF, Tinetti ME. A predictive model for ADL dependence in community-living older adults based on a reduced set of cognitive status items. *J Am Geriatr Soc*. 1997;45(4):441–5.
32. Hegendörfer E, Vaes B, Van Pottelbergh G, Matheï C, Verbakel J, Degryse JM. Predictive Accuracy of Frailty Tools for Adverse Outcomes in a Cohort of Adults 80 Years and Older: A Decision Curve Analysis. *J Am Med Dir Assoc*. 2020;21(3):440.e1–440.e8.
33. Hong S, Kim S, Yoo J, Kim BS, Choi HR, Choi SE, et al. Slower gait speed predicts decline in Instrumental Activities of Daily Living in community-dwelling elderly: 3-year prospective finding from Living Profiles of Older People Survey in Korea. *Journal of Clinical Gerontology and Geriatrics*. 2016;7(4):141–5.
34. Ishimoto Y, Wada T, Kasahara Y, Kimura Y, Fukutomi E, Chen W, et al. Fall Risk Index predicts functional decline regardless of fall experiences among community-dwelling elderly. *Geriatr Gerontol Int*. 2012;12(4):659–66.
35. Jonkman NH, Colpo M, Klenk J, Todd C, Hoekstra T, Del Panta V, et al. Development of a clinical prediction model for the onset of functional decline in people aged 65–75 years: pooled analysis of four European cohort studies. *BMC geriatr*. 2019;19(1):179.
36. Jonkman NH, Del Panta V, Hoekstra T, Colpo M, van Schoor NM, Bandinelli S, et al. Predicting Trajectories of Functional Decline in 60- to 70-Year-Old People. *Gerontology*. 2018;64(3):212–21.
37. Lam FMH, Su Y, Lu ZH, Yu R, Leung JCS, Kwok TCY. Cumulative and Incremental Value of Sarcopenia Components on Predicting Adverse Outcomes. *J Am Med Dir Assoc*. 2020;21(10):1481–9.
38. Lin MR, Hwang HF, Hu MH, Wu HD, Wang YW, Huang FC. Psychometric comparisons of the timed up and go, one-leg stand, functional reach, and Tinetti balance measures in community-dwelling older people. *J Am Geriatr Soc*. 2004;52(8):1343–8.
39. McClintock HF, Kurichi JE, Kwong PL, Xie D, Stineman MG, Bogner HR. A Risk Scoring System for the Prediction of Functional Deterioration, Institutionalization, and Mortality Among Medicare Beneficiaries. *Am J Phys Med Rehabil*. 2018;97(10):698–707.
40. Nuesch E, Pablo P, Dale CE, Prieto-Merino D, Kumari M, Bowling A, et al. Incident disability in older adults: prediction models based on two British prospective cohort studies. *Age Ageing*. 2015;44(2):275–82.
41. Papachristou E, Wannamethee SG, Lennon LT, Papacosta O, Whincup PH, Iliffe S, et al. Ability of Self-Reported Frailty Components to Predict Incident Disability, Falls, and All-Cause Mortality: Results From a Population-Based Study of Older British Men. *J Am Med Dir Assoc*. 2017;18(2):152–7.
42. Perera S, Patel KV, Rosano C, Rubin SM, Satterfield S, Harris T, et al. Gait Speed Predicts Incident Disability: A Pooled Analysis. *J Gerontol A Biol Sci Med Sci*. 2016;71(1):63–71.
43. Saraiva MD, Venys AL, Abdalla FLP, Fernandes MS, Pisoli PH, Sousa DMDRV, et al. AMPI-AB validity and reliability: a multidimensional tool in resource-limited primary care settings. *BMC geriatr*. 2020;2(1):124.
44. Sarkisian CA, Liu H, Gutierrez PR, Luxenberg JS. A clinical prediction tool that includes modifiable risk factors predicted functional decline in elderly women. *Evid Based Med*. 2000;5(5):160.
45. Spalter T, Brodsky J, Shnoor Y. Improvements and decline in the physical functioning of Israeli older adults. *Gerontologist*. 2014;54(6):919–29.
46. Suijker JJ, Buurman BM, van Rijn M, van Dalen MT, ter Riet G, van Geloven N, et al. A simple validated questionnaire predicted functional decline in community-dwelling older persons: prospective cohort studies. *J Clin Epidemiol*. 2014;67(10):1121–30.
47. Taş U, Steyerberg EW, Bierma-Zeinstra SM, Hofman A, Koes BW, Verhagen AP. Age, gender and disability predict future disability in older people: the Rotterdam Study. *BMC geriatr*. 2011;11:22.
48. Teo N, Gao Q, Nyunt MSZ, Wee SL, Ng TP. Social Frailty and Functional Disability: Findings From the Singapore Longitudinal Ageing Studies. *J Am Med Dir Assoc*. 2017;18(7):637.e13–637.e19.
49. Terhorst L, Holm MB, Toto PE, Rogers JC. Performance-Based Impairment Measures as Predictors of Early-Stage Activity Limitations in Community-Dwelling Older Adults. *J Aging Health*. 2017;29(5):880–92.
50. Wennie Huang WN, Perera S, VanSwearingen J, Studenski S. Performance measures predict onset of activity of daily living difficulty in community-dwelling older adults. *J Am Geriatr Soc*. 2010;58(5):844–52.
51. Yam A, Marsiske M. Cognitive longitudinal predictors of older adults' self-reported IADL function. 2013. p. 1635–855.
52. Dixon A, Porter S, Suhrie K, Hammers D, Duff K. Predicting daily functioning with the modified Telephone Interview for Cognitive Status. *Aging Clin Exp Res*. 2022. online ahead of print. <https://doi.org/10.1007/s40520-022-02081-4>.
53. Bongue B, Buisson A, Dupre C, Beland F, Gonthier R, Crawford-Achour É. Predictive performance of four frailty screening tools in community-dwelling elderly. *BMC geriatr*. 2017;17(1):262.
54. Carrière I, Colvez A, Favier F, Jeandel C, Blain H. Hierarchical components of physical frailty predicted incidence of dependency in a cohort of elderly women. *J Clin Epidemiol*. 2005;58(11):1180–7.
55. Clark DO, Stump TE, Tu W, Miller DK. A comparison and cross-validation of models to predict basic activity of daily living dependency in older adults. *Med Care*. 2012;50(6):534–9.
56. Ensrud KE, Ewing SK, Taylor BC, Fink HA, Cawthon PM, Stone KL, et al. Comparison of 2 frailty indexes for prediction of falls, disability, fractures, and death in older women. *Arch Intern Med*. 2008;168(4):382–9.
57. Gobbens RJ, van Assen MA, Luijckx KG, Schols JM. The predictive validity of the Tilburg Frailty Indicator: disability, health care utilization, and quality of life in a population at risk. *Gerontologist*. 2012;52(5):619–31.
58. Guralnik JM, Ferrucci L, Pieper CF, Leveille SG, Markides KS, Ostir GV, et al. Lower extremity function and subsequent disability: consistency across

- studies, predictive models, and value of gait speed alone compared with the short physical performance battery. *J Gerontol A Biol Sci Med Sci*. 2000;55(4):M221–31.
59. Onder G, Penninx BW, Ferrucci L, Fried LP, Guralnik JM, Pahor M. Measures of physical performance and risk for progressive and catastrophic disability: results from the Women's Health and Aging Study. *J Gerontol A Biol Sci Med Sci*. 2005;60(1):74–9.
 60. Op Het Veld LPM, Beurskens A, de Vet HCW, van Kuijk SMJ, Hajema K, Kempen G, et al. The ability of four frailty screening instruments to predict mortality, hospitalization and dependency in (instrumental) activities of daily living. *European journal of ageing*. 2019;16(3):387–94.
 61. Studenski S, Perera S, Wallace D, Chandler JM, Duncan PW, Rooney E, et al. Physical performance measures in the clinical setting. *J Am Geriatr Soc*. 2003;51(3):314–22.
 62. Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. *BMC Med*. 2019;17(1):230.
 63. Derksen S, Keselman HJ. Backward, forward and stepwise automated subset selection algorithms: Frequency of obtaining authentic and noise variables. *Br J Math Stat Psychol*. 1992;45(2):265–82.
 64. Su TL, Jaki T, Hickey GL, Buchan I, Sperrin M. A review of statistical updating methods for clinical prediction models. *Stat Methods Med Res*. 2018;27(1):185–97.
 65. Long JD, Mills JA. Joint modeling of multivariate longitudinal data and survival data in several observational studies of Huntington's disease. *BMC Med Res Methodol*. 2018;18(1):138.
 66. Verbakel JY, Steyerberg EW, Uno H, De Cock B, Wynants L, Collins GS, et al. ROC curves for clinical prediction models part 1. ROC plots showed no added value above the AUC when evaluating the performance of clinical prediction models. *J Clin Epidemiol*. 2020;126:207–16.
 67. European Commission Economic and Financial Affairs. The 2018 Aging report. Economic & budgetary projections for the 28 EU member States (2016 - 2070). Institutional Paper 079. Brussels; 2018.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

