JAMDA xxx (2023) 1–6

JAMDA

journal homepage: www.jamda.com

Original Study

Validation of the ADFICE_IT Models for Predicting Falls and Recurrent Falls in Geriatric Outpatients

Bob van de Loo MSc [a,b,d,*], Martijn W. Heymans PhD [a,d], Stephanie Medlock DVM, PhD [c,d], Nicole D.A. Boyé MD, PhD [e,f], Tischa J.M. van der Cammen MD, PhD [f,g], Klaas A. Hartholt MD, PhD [f,h], Marielle H. Emmelot-Vonk MD, PhD [i], Francesco U.S. Mattace-Raso MD, PhD [f], Ameen Abu-Hanna PhD [c,d], Nathalie van der Velde MD, PhD [b,d,f], Natasja M. van Schoor PhD [a,d]

[a] *Epidemiology and Data Science, Amsterdam UMC location Vrije Universiteit Amsterdam, Amsterdam, the Netherlands*
[b] *Internal Medicine, Section of Geriatric Medicine, Amsterdam UMC location University of Amsterdam, Amsterdam, the Netherlands*
[c] *Department of Medical Informatics, Amsterdam UMC location University of Amsterdam, Amsterdam, the Netherlands*
[d] *Amsterdam Public Health research institute, Amsterdam, the Netherlands*
[e] *Department of General Surgery, Curaçao Medical Center, Willemstad, Curaçao*
[f] *Section of Geriatric Medicine, Department of Internal Medicine, Erasmus MC, University Medical Center, Rotterdam, the Netherlands*
[g] *Department of Human-Centred Design, Faculty of Industrial Design Engineering, Delft University of Technology, Delft, the Netherlands*
[h] *Department of Surgery-Traumatology, Reinier de Graaf Gasthuis, Delft, the Netherlands*
[i] *Department of Geriatric Medicine, University Medical Centre Utrecht, Utrecht University, Utrecht, the Netherlands*

Keywords:
Accidental falls
Prognosis
decision curve analysis
Individual participant data

A B S T R A C T

Objectives: Before being used in clinical practice, a prediction model should be tested in patients whose data were not used in model development. Previously, we developed the ADFICE_IT models for predicting any fall and recurrent falls, referred as Any_fall and Recur_fall. In this study, we externally validated the models and compared their clinical value to a practical screening strategy where patients are screened for falls history alone.
Design: Retrospective, combined analysis of 2 prospective cohorts.
Setting and Participants: Data were included of 1125 patients (aged ≥65 years) who visited the geriatrics department or the emergency department.
Methods: We evaluated the models' discrimination using the C-statistic. Models were updated using logistic regression if calibration intercept or slope values deviated significantly from their ideal values. Decision curve analysis was applied to compare the models' clinical value (ie, net benefit) against that of falls history for different decision thresholds.
Results: During the 1-year follow-up, 428 participants (42.7%) endured 1 or more falls, and 224 participants (23.1%) endured a recurrent fall (≥2 falls). C-statistic values were 0.66 (95% CI 0.63-0.69) and 0.69 (95% CI 0.65-0.72) for the Any_fall and Recur_fall models, respectively. Any_fall overestimated the fall risk and we therefore updated only its intercept whereas Recur_fall showed good calibration and required no update. Compared with falls history, Any_fall and Recur_fall showed greater net benefit for decision thresholds of 35% to 60% and 15% to 45%, respectively.
Conclusions and Implications: The models performed similarly in this data set of geriatric outpatients as in the development sample. This suggests that fall-risk assessment tools that were developed in community-dwelling older adults may perform well in geriatric outpatients. We found that in geriatric

Funding sources: The ADFICE_IT project was supported by funding from the Netherlands Organization for Health Research and Development (ZonMw, grant 848017004), the Hague. The IMPROveFALL study was supported by ZonMw, grant 170.885.607. The sponsor played no part in the design, execution, analysis, and interpretation of data or writing of the study.

The authors declare no conflicts of interest.

* Address correspondence to Bob van de Loo, MSc, Amsterdam UMC location Vrije Universiteit Amsterdam, Epidemiology and Data Science, De Boelelaan 1117, Amsterdam, the Netherlands.
E-mail address: r.vandeloo@amsterdamumc.nl (B. van de Loo).

https://doi.org/10.1016/j.jamda.2023.04.021

outpatients the models have greater clinical value across a wide range of decision thresholds compared with screening for falls history alone.

Optimal use of fall-preventive measures requires identification of patients who are at an increased risk of falling. To this end, a growing number of fall-risk assessment tools have been developed for use in older adults.[1,2] These include prediction models that combine data of patient characteristics as predictors to estimate the probability of a future fall. Prediction models for falls are commonly developed using data from community-dwelling older adults, which means their generalizability to geriatric outpatients may be limited. Geriatric outpatients have been characterized as being more frail and as having a higher risk of falling.[3] Yet, only a few studies have validated prediction models for falls in geriatric outpatients.[4-7]

After developing a prediction model, it is recommended to assess its predictive performance in patients whose data were not used in model development.[8] Predictive performance is usually evaluated in terms of discrimination and calibration. Discrimination reflects the model's ability to differentiate between those who do or do not experience the outcome. Calibration refers to the agreement between the model's predictions and observed outcomes. However, measures for discrimination and calibration alone may not be sufficient to determine a model's clinical value as they offer little insight into the clinical consequences of using a prediction model in practice.[9] Various decision-analytical methods have been developed for weighing the clinical value of prediction models over competing screening methods. To our knowledge, previous validation studies have not compared the clinical value of prediction models for falls to other screening methods that are commonly used.

As part of the AD*F*ICE_IT (Alerting on adverse Drug reactions: Falls prevention Improvement through developing a Computerized clinical support system: Effectiveness of Individualized medicaTion withdrawal) project, we developed models for predicting falls and recurrent falls in community-dwelling older adults.[10] This study aimed to externally validate and update these models in the mixed population of geriatric outpatients as this is the population in which the models are intended to be used. Accordingly, we assessed the performance of the models in terms of discrimination, calibration, and clinical value using a combined data set of 2 cohorts of geriatric outpatients.

Methods

In the present study, we externally validated the AD*F*ICE_IT models, which provide the probability of enduring any fall or a recurrent fall in the next 12 months. The development of the models has been described in detail elsewhere.[10] In brief, the models were developed using harmonized data from 3 European cohort studies of community-dwelling older adults. Logistic regression with backward variable elimination was used to develop the models. The final model for predicting any fall included verbal fluency score as a predictor that may not be feasible to collect in routine clinical practice. Therefore, as described in van de Loo et al,[10] a second version of the model for predicting any fall was developed using a subset of the candidate predictors that were considered easily obtainable in clinical practice. Table 1 presents an overview of the predictors in the models.

In the main analyses of the present study, we externally validated the 2 models with predictors that are easily obtainable in clinical practice, namely, (1) the model for predicting any fall (Any_fall) and (2) the model for predicting recurrent falls (Recur_fall).[10] In an additional analysis, we also assessed the performance of the original version of the Any_fall model, which includes verbal fluency as a predictor.

Study Population

The present study was a retrospective analysis of 2 prospective cohort studies, namely, the IMPROveFALL study[11] and the Utrecht Cardiovascular Cohort.[12] Data from both cohorts were combined, resulting in a total sample of 1125 participants. The intended target population of the models is a mixed population of older persons who visited a geriatric outpatient clinic or were eligible to visit a geriatric outpatient clinic. We combined the cohorts in order to assess the performance of the models in a more heterogeneous sample that more closely resembles the mixed population of geriatric outpatients. Both studies were performed in accordance with the Declaration of Helsinki. All participants provided informed consent. The medical research ethics committees of both participating hospitals approved the protocol for IMPROveFALL. The Medical Ethics Committee of the University Medical Center Utrecht granted a waiver for the Utrecht Cardiovascular Cohort study.

The IMPROveFALL study was a randomized, multicenter trial that investigated the effect of fall-risk-increasing drug withdrawal vs care as usual.[11] Results showed no effect of the intervention on falls and we therefore treated IMPROveFALL as a prospective cohort study.[13] Community-dwelling patients aged 65 years or older that visited the emergency department because of a fall and that used 1 or more fall-risk–increasing drugs were eligible for inclusion. Dutch and international clinical guidelines stipulate that a fall-related emergency department visit warrants a multifactorial fall risk assessment, which are typically conducted at a geriatric outpatient (falls) clinic.[14,15] We therefore considered the IMPROveFALL participants as geriatric outpatients in the context of the present study. In total, 616 participants were recruited across 2 hospitals in the Netherlands between 2008 and 2012. Of these, we included 580 participants for whom follow-up data were available. Additionally, we derived data from the Utrecht Cardiovascular Cohort study, a prospective cohort study that was set up to optimize uniform registration of cardiovascular information in routine clinical practice at University Medical Center Utrecht, the Netherlands.[12] We included data from 545 participants who visited one of the geriatric outpatient clinics at the geriatrics department, who were aged 65 years and older, and for whom follow-up data were available. Participants were recruited between 2011 and 2014.

Outcome

Outcome data on any fall and recurrent falls (2 or more falls) over a follow-up period of 1 year were available. In IMPROveFALL, falls were ascertained prospectively using fall calendars. Participants were asked to record falls on a weekly basis and to return the calendars every 3 months. In IMPROveFALL, a fall was defined as unintentionally coming to rest on the ground or a lower level with or without loss of consciousness, but not as a result of an acute medical condition, such as a stroke, or an exogenous factor, such as a traffic accident. In the Utrecht Cardiovascular Cohort study, falls were measured retrospectively with a questionnaire, which was administered 12 months after baseline. In this questionnaire, falls were ascertained using the

Table 1
Predictors in the ADFICE_IT Models for Predicting Any Fall and Recurrent Falls

Predictor	Model		
	Any_fall*	Original Any_fall†	Recur_fall
Educational status	+	+	+
Standardized depressive symptoms score	+	+	
Standardized verbal fluency score		+	
Visual impairment			+
BMI	+	+	+
Number of functional limitations	+	+	+
Standardized gait speed	+		
Standardized grip strength	+	+	
Urinary incontinence			+
Systolic blood pressure (mmHg)	+	+	
History of 1 or more falls in the previous 12 mo	+	+	+
History of 2 or more falls in the previous 12 mo	+	+	+
Fear of falling	+		
Smoking status	+	+	
Use of calcium channel blockers	+		
Use of anti-Parkinson drugs			+
Use of antiepileptics	+	+	
Use of drugs for urinary frequency and incontinence	+	+	+
Use of antihistamines			+

The "+" indicates the predictor was included in the final model.
Any_fall, model for predicting any fall; BMI, body mass index; Recur_fall, model for predicting recurrent falls.
*The Any_fall model was developed using a subset of predictors that were considered to be easily obtainable in practice.
†The original Any_fall model was developed using all predictors and includes verbal fluency as a predictor, which may not be feasible to collect in practice.

Table 2
Baseline Characteristics for the Development and Validation Cohorts

	Development Cohorts (n = 5722)	Validation Cohorts (n = 1125)
Age (y)	74 [69, 79]	77 [72, 83]
Sex (female)	2785 (48.7)	702 (62.4)
Educational status (middle or high)	1589 (27.8)	611 (57.9)
Fluency test score*	17 [12, 23]	13 [9, 17]
Visual impairment	303 (10.9)	307 (28.3)
BMI	27.19 ± 4.08	26.9 ± 4.7
Grip strength (kg)	31.91 ± 10.93	25.1 ± 9.4
Gait speed (m/s)*	0.9 ± 0.3	1.0 ± 0.4
Urinary incontinence†	870 (31.2)	85 (14.7)
Systolic blood pressure (mm Hg)	148.2 ± 19.7	150.5 ± 24.4
History of 1 or more falls in the previous 12 mo	1664 (33.0)	894 (80.7)
History of 2 or more falls in the previous 12 mo	640 (12.7)	484 (44.2)
Fear of falling (somewhat or very afraid of falling)	1294 (46.3)	481 (43.6)
Current smoker	634 (11.1)	136 (12.3)
Number of medications	3 [1, 5]	6 [4, 8]
Calcium channel blockers	699 (12.2)	213 (19.3)
Anti-Parkinson drugs	68 (1.2)	32 (2.9)
Antiepileptics	108 (1.9)	48 (4.3)
Drugs for urinary frequency and incontinence	76 (1.3)	44 (4.0)
Antihistamines	116 (2.0)	30 (2.7)

BMI, body mass index.
Data are presented as mean ± SD, n (%), or median [interquartile range]. Proportions are provided as valid percentages, meaning missing data were not included in their calculations.
*These variables were not measured in IMPROveFALL.
†This variable was not measured in the Utrecht Cardiovascular Cohort.

questions "Did you endure a fall in the past year?" and "During the past year, how often did you fall?"

Predictors and Harmonization

Data for predictors were collected at baseline. We developed harmonization algorithms for the model predictors. We refer to the harmonization guide for more information regarding the harmonization procedures as well as the measurements and definitions of predictors (Supplementary Materials 1 and 2). For IMPROveFALL, there were no data on gait speed, depressive symptoms, and verbal fluency. We used data from the Timed Up and Go test as a proxy for gait speed and data from the EuroQol-5 Dimensions as a proxy for depressive symptoms.[16,17] Values for these predictors were harmonized using Z score transformations. Two predictors were systematically missing, meaning that there were no data or proxy data available for these predictors in one of the cohorts (ie, verbal fluency in IMPROveFALL and urinary incontinence in the Utrecht Cardiovascular Cohort study). Systematically missing data for these variables were multiply imputed (see subheading "Missing data").

Sample Size

An adequate sample size is needed to estimate the predictive performance of a prediction model with sufficient precision. We applied the sample size calculations of Riley et al[18] to assess whether the combined sample size of the 2 cohorts was adequate for assessing the discriminative performance of the models. We anticipated C-statistic values of 0.65 for Any_fall and 0.70 for Recur_fall.[10] Given the outcome proportion in the nonimputed data, we required at least 486 participants for Any_fall and 652 participants for Recur_fall to obtain C-statistic estimates with a targeted CI width of 0.1. Given our sample size, we would be able to estimate the respective slope values for the Any_fall and Recur_fall models with CIs of 0.43 and 0.40. We assumed alpha = 0 and beta = 1, as recommended by Riley et al.[18]

Missing Data

Excluding systematically missing variables, the median percentage of missing values across the predictors was 0% (IQR 0%-1%) in IMPROveFALL and 3.9% (IQR 3.5%-4.9%) in the Utrecht Cardiovascular Cohort. Median follow-up time in IMPROveFALL was 52 weeks (IQR 11-52). Values for outcome variables in IMPROveFALL were set to missing if loss to follow-up occurred before the event of interest was reported. Multiple imputation by chained equations was employed to impute missing values for predictors and outcome variables.[19] We imputed 10.9% and 14.1% of values for the any fall and recurrent falls outcomes, respectively. Systematically missing predictors were imputed using data from the development cohorts. Results from a simulation study showed this is the optimal method of handling systematically missing predictor values when validating a prediction model.[20]

Statistical Analysis

Discriminative performance was evaluated using the C-statistic, for which a value of 0.5 indicates no discrimination and a value of 1 indicates perfect discrimination. Calibration was assessed using calibration plots as well as by calculating the calibration intercept and slope. The ideal values for the calibration intercept and slope are 0 and 1, respectively. Deviations from these ideal values with P value <.05 were considered significant (see Supplementary Material 3 for further details regarding further details and explanation regarding the statistical analysis). Based on whether these values were significant, we chose one of the following updating methods: (1) no update if intercept and slope values do not deviate from ideal values; (2) recalibration-in-the-large (ie, reestimation of model intercept) if only the intercept value deviates from the ideal value; and (3) recalibration (ie, reestimation of intercept and slope) if the slope value (and intercept value) deviates from ideal value.[21] Predictive performance of the updated models was reevaluated.

We applied decision curve analysis to evaluate the clinical value of the models.[22] In decision curve analysis, the net benefit of a model and alternative strategies is plotted against a range of decision thresholds. Clinical practice guidelines for falls prevention often recommend using fall history as a starting point for risk stratification and deciding on possible interventions.[23] Therefore, we compared the net benefit of the models against that of a practical strategy in which only those with a history of 1 or more falls in the past year receive treatment or further assessment. Additionally, the net benefit of the models was compared against 2 theoretical strategies in which all patients receive treatment or further assessment ("treat all") and in which none receive treatment or further assessment ("treat none").

Statistical analyses were performed with R (version 4.0.2). We used the mice package for multiple imputation and the psfmi package for pooling performance estimates across imputed data sets.[19,24] Net benefit of the models was computed using the rmda package.[25] The Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) Statement checklist was used as a guideline for reporting (Supplementary Material 4).[26]

Results

A total of 1125 participants was included in the analyses. Baseline characteristics of the development and validation cohorts are presented in Table 2. Participants from the validation cohorts differed from those in the development cohorts with respect to most characteristics. Baseline characteristics of 2 validation cohorts are presented in Supplementary Table 1. Participants from the 2 cohorts were similar with respect to most characteristics. Participants from IMPROveFALL were younger and higher educated than those from the Utrecht Cardiovascular Cohort. In the 2 validation cohorts, a total of 428 participants (42.7%) endured at least 1 fall and a total of 224 participants (23.1%) endured 2 or more falls during the follow-up period of 1 year (see Supplementary Table 2 for complete overview).

We tested the predictive performance of the models in the combined data set of the 2 validation cohorts (see Supplementary Table 3 for complete overview). The C-statistic values for the Any_fall model was 0.66 (95% CI 0.63-0.69) vs 0.69 (95% CI 0.65-0.72) for the Recur_fall model. The calibration plot for the Any_fall models revealed a slight overestimation across the entire range of predicted risks (Figure 1A). The calibration plot for the Recur_fall model showed good agreement between observed and predicted risks (Figure 1B). Calibration slope values were 0.96 (95% CI 0.74-1.19) and 0.84 (95% CI 0.64-1.04) for the Any_fall and Recur_fall models, respectively. The calibration slope values of both models did not deviate from their ideal value of 1 ($P > .05$). The respective values for the calibration intercept

Table 3
Model Coefficients of the Updated Any_fall Model

Predictor	Beta	SE
Intercept*	−0.218	0.067
Educational status		
Middle	0.173	0.105
High	0.324	0.079
Depressive symptoms score†	0.068	0.037
BMI	−0.018	0.008
Number of functional limitations†	0.125	0.054
Grip strength (kg)†	−0.148	0.035
Gait speed (m/s)†	0.088	0.041
Systolic blood pressure (mm Hg)	−0.003	0.002
History of 1 or more falls in the previous 12 mo	0.426	0.074
History of 2 or more falls in the previous 12 mo	0.597	0.100
Fear of falling		
Somewhat afraid of falling	0.199	0.078
Very afraid of falling	0.195	0.127
Current smoker	−0.252	0.103
Use of calcium channel blockers	−0.164	0.093
Use of antiepileptics	0.436	0.223
Use of drugs for urinary frequency and incontinence	0.656	0.277

BMI, body mass index.

*Intercept of the model was updated based on the combined data set.

†Beta is based on 1-unit difference in Z score, which can be calculated as follows: Z score depressive symptoms score$_{\text{IMPROveFALL}}$= (EuroQol − 1.312)/0.523, where EuroQol can take on a value of 1, 2, or 3, corresponding with the answers "I am not anxious or depressed," "I am moderately anxious or depressed," and "I am extremely anxious or depressed"; Z score depressive symptoms score$_{\text{Utrecht Cardiovascular Cohort}}$ = (Geriatric Depression Scale score − 1.440)/1.942; Z score functional limitations$_{\text{IMPROveFALL}}$ = (Katz-12 score − 4.714)/3.982; Z score functional limitations$_{\text{Utrecht Cardiovascular Cohort}}$ = (Katz-15 score − 0.530)/1.232; Z score grip strength$_{\text{IMPROveFALL}}$ = [grip strength (kg) − 26.290]/9.316; Z score grip strength$_{\text{Utrecht Cardiovascular Cohort}}$ = [grip strength (kg) − 23.580]/9.850; Z score gait speed$_{\text{IMPROveFALL}}$ = [Timed Up & Go Test score (m/s) − 0.613]/0.258; and Z score gait speed$_{\text{Utrecht Cardiovascular Cohort}}$ = [gait speed (m/s) − 0.943]/0.557.

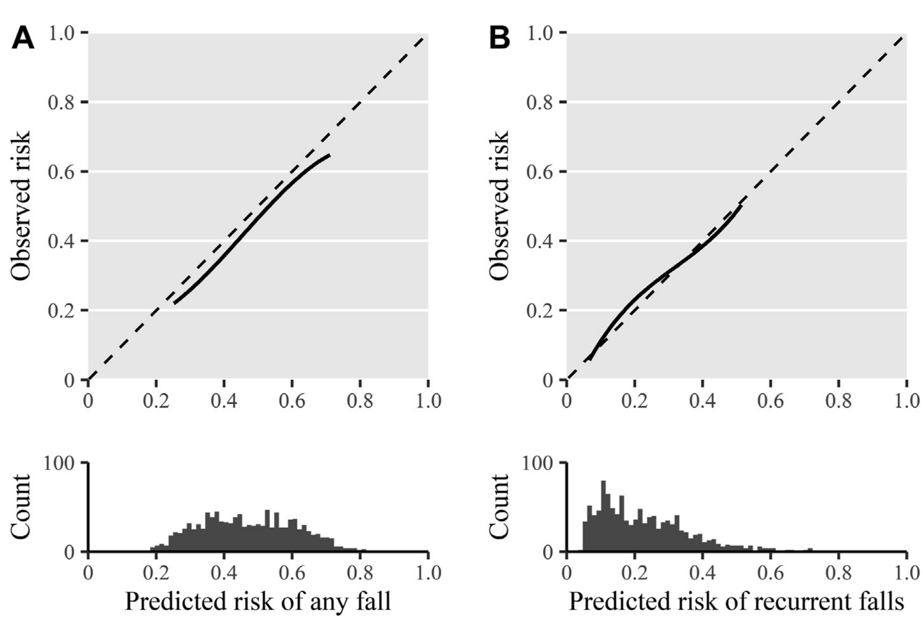

Fig. 1. Calibration plots for (A) the Any_fall model and (B) the Recur_fall model as derived from the combined data set. Perfect calibration is represented by the dashed line. The graphs in the lower row show the distribution of the predicted risks.

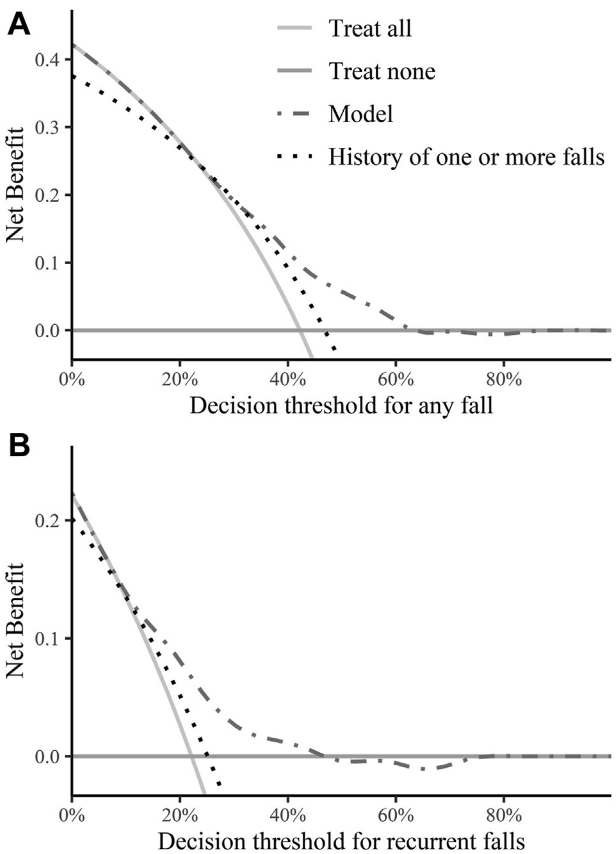

Fig. 2. Decision curves for (A) the Any_fall model and (B) the Recur_fall model as derived from the combined data set. The decision threshold is the probability above which a patient or physician judges the benefit of treatment to outweigh the potential harm of unnecessary treatment. The net benefit of a model is the difference between the proportion of true positives and the proportion of false positives as weighted by the odds of a given decision threshold.

were −0.17 (95% CI −0.31 to −0.04, $P = .01$) and 0.07 (95% CI −0.09 to 0.23, $P = .37$), with the calibration intercept for the Any_fall model deviating significantly from its ideal value of 0. We therefore performed recalibration-in-the-large for the Any_fall model by updating its intercepts based on the combined data set. The updated model is presented in Table 3. Predictive performance of the updated model is presented in Supplementary Table 4 and Supplementary Figures 1 and 2.

We plotted decision curves for each model to compare the net benefit of the models against that of 3 alternative strategies for deciding on treatment or further assessment: treat those with a history of 1 or more falls, treat all, and treat none. In comparison with the alternative strategies, application of the Any_fall model yields a higher net benefit when the decision threshold for any fall lies between 35% and 60% (Figure 2A). The Recur_fall model yields a higher net benefit than alternative strategies when the decision threshold for recurrent falls lies between 15% and 45% (Figure 2B).

As an additional analysis, we externally validated the original Any_fall model, which includes verbal fluency as a predictor. The C-statistic of the model was 0.66 (95% CI 0.62-0.69). The calibration plot for the model revealed a slight overestimation of predicted risks (Supplementary Figure 3). The calibration slope of the original Any_fall model did not deviate significantly from its ideal value ($P = .92$) but the calibration intercept did ($P = .01$) and we therefore performed recalibration-in-the-large (Supplementary Table 5). In comparison with the alternative strategies, application of the original Any_fall model yields a higher net benefit when the decision threshold for any fall lies between 35% and 60% (Supplementary Figure 4).

Discussion

In this study, we externally validated the ADFICE_IT models using data of geriatric outpatients. Of the validated models, discriminative performance was best for the Recur_fall model. Discriminative performance of the 2 Any_fall models was similar. To our knowledge, this was the first study to compare the clinical utility of prediction models for falls against a screening strategy in which patients are screened for falls history alone.

Predictive performance of the models is comparable with that of other prediction models for falls that have been validated in geriatric outpatients. Peeters et al validated the LASA fall risk profile for identifying recurrent fallers in older adults who consulted their general practitioner or emergency department after a fall.[4] The authors obtained a C-statistic of 0.65 and showed the LASA fall risk profile to be well calibrated.[4] Other studies have presented tools for assessing a patient's risk of enduring 1 or more falls following an emergency department visit, of which the Two-Item Screening Tool and the FROP-Com have been validated.[27] Studies that validated these 2 tools have reported mixed results with regard to their predictive performance.[5-7] One study reported poor discriminative performance, with C-statistic values of 0.54 and 0.57 for the Two-Item Screening Tool and the FROP-Com, respectively.[5] One study that validated the Two-Item Screening Tool showed it to be well calibrated.[7] Calibration was not assessed in the other 2 studies that validated the Two-Item Screening Tool and FROP-Com.[5,6]

Our results suggest that fall-risk assessment tools that were developed in community-dwelling older adults may be generalized to geriatric outpatients. Indeed, the discriminative performance of the models in this sample of geriatric outpatients was similar to that in the development cohorts, which consisted of community-dwelling older adults.[10] However, the Any_fall model was found to overestimate the general risk of any fall and we therefore updated its intercept. The Recur_fall model for predicting recurrent falls required no update as the model was well calibrated, meaning its predictions were in line with the observed number of recurrent falls. The application of fall prevention strategies to participants in the validation cohorts may have resulted in a lower number of falls than would have occurred otherwise, which may have affected the calibration of the Any_fall model.

Although the Any_fall and the Recur_fall models require the use of 10 to 14 predictors, they are easily obtainable and commonly available from a comprehensive geriatric assessment. Nonetheless, more practical fall-risk assessment strategies exist that require data on only 1 or 2 patient characteristics, such as a strategy in which patients are screened for history of falls alone.[28] However, fall-risk assessment tools that cover multiple domains of risk factors are likely to show better predictive performance, given the multifactorial nature of falls. In fact, results from Palumbo et al[29] indicate that a total number of 20 to 30 model predictors may be required to attain the highest predictive performance when estimating fall risk. However, the inclusion of more predictors in the model may limit their usability in clinical practice.

Results of the decision curve analysis demonstrate that for a broad range of decision thresholds, the models yield a higher net benefit than a strategy in which patients are screened for history of falls alone. However, to our knowledge, no study has investigated which decision thresholds patients and physicians prefer for fall prevention interventions. Nonetheless, interventions that pose no risk and require no effort, such as educational interventions, will evidently correspond to a low decision threshold. In cases in which such interventions are under consideration, the models are unlikely to benefit decision making over a treat-all strategy. Indeed, we found that the Any_fall and Recur_fall models benefit decision making in cases for which the decision threshold lies between 35%-60% and 15%-45%, respectively. Therefore, the models may only benefit decision making when the intervention under consideration involves some risk or effort, such as a medication review or an exercise intervention. Still, we were unable to account for fall prevention strategies that were applied to participants, which may have affected the proportion of true positives and false positives for different decision thresholds. Therefore, more research is needed to evaluate the impact of the models on clinical decision making before the models can be recommended outside of the research setting.[30] The Any_fall model was implemented in a clinical decision support system for optimizing deprescribing of fall-risk–increasing drugs in older fallers, which is currently being evaluated in a multicenter trial.[31] Here, the model is used to provide a personalized fall risk estimate, which we hypothesize to support both the patient and physician in engaging in a more informed discussion about treatment options.

A few limitations deserve consideration. There were some differences in how predictors were measured in IMPROveFALL and the Utrecht Cardiovascular Cohort compared with the development cohorts. We used proxy variables for some predictors. For 2 other predictors, there were no proxies available in one of the cohorts. Data for these predictors were multiply imputed, which has been shown to result in accurate model predictions.[20] Additionally, a limitation of this study was that the ascertainment of the outcome differed between the 2 cohorts. In IMPROveFALL, falls were ascertained using weekly falls calendars, the gold standard of falls measurement. In the Utrecht Cardiovascular Cohort, falls were measured retrospectively using a questionnaire, which likely resulted in some misclassification because of recall bias. We expect this misclassification to be mostly random and as such have little effect on the overall discriminative performance of the models. Finally, our sample size calculations suggest the sample size may not have been large enough to estimate the slope with sufficient precision.

Conclusions and Implications

The models can potentially be used in geriatric outpatient settings for opportunistic case finding or as an adjunct to decision making, by supporting the physician and patient in weighing treatment options based on the estimated fall risk. However, further impact assessment is required before the models can be recommended outside the research setting. Our study demonstrates that, in a geriatric outpatient setting, prediction models may outperform a screening strategy in which the single question “Have you fallen in the last 12 months?” is posed. Finally, our results indicate that fall-risk assessment tools that were developed in community-dwelling older adults may be generalized to geriatric outpatients.

Supplementary Data

Supplementary data related to this article can be found online at https://doi.org/10.1016/j.jamda.2023.04.021.

References

1. Park SH. Tools for assessing fall risk in the elderly: a systematic review and meta-analysis. *Aging Clin Exp Res*. 2018;30:1–16.
2. Gade GV, Jørgensen MG, Ryg J, et al. Predicting falls in community-dwelling older adults: a systematic review of prognostic models. *BMJ Open*. 2021;11:e044170.
3. Kikkert LHJ, De Groot MH, Van Campen JP, et al. Gait dynamics to optimize fall risk assessment in geriatric patients admitted to an outpatient diagnostic clinic. *PLoS One*. 2017;12:1–14.
4. Peeters GMEE, Pluijm SMF, Van Schoor NM, Elders PJM, Bouter LM, Lips P. Validation of the LASA fall risk profile for recurrent falling in older recent fallers. *J Clin Epidemiol*. 2010;63:1242–1248.
5. Harper KJ, Barton AD, Arendts G, Edwards DG, Petta AC, Celenza A. Failure of falls risk screening tools to predict outcome: A prospective cohort study. *Emerg Med J*. 2018;35:28–32.
6. Russell MA, Hill KD, Blackberry I, Day LM, Dharmage SC. The reliability and predictive accuracy of the falls risk for older people in the community assessment (FROP-Com) tool. *Age Ageing*. 2008;37:634–639.
7. Tiedemann A, Sherrington C, Orr T, et al. Identifying older people at high risk of future falls: development and validation of a screening tool for use in emergency departments. *Emerg Med J*. 2013 Nov;30:918–922.
8. Moons KGM, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart*. 2012;98:691–698.
9. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. *BMJ*. 2016;352:3–7.
10. van de Loo B, Seppala LJ, van der Velde N, et al. Development of the ADF ICE_IT models for predicting falls and recurrent falls in community-dwelling older adults: pooled analyses of European cohorts with special attention to medication. *J Gerontol A Biol Sci Med Sci*. 2022;77:1446–1454.
11. Hartholt KA, Van Der Velde N, Van Lieshout EM, et al. [Cost] effectiveness of withdrawal of fall-risk increasing drugs versus conservative treatment in older fallers: design of a multicenter randomized controlled trial (IMPROveFALL-study). *BMC Geriatr*. 2011;11:1–8.
12. Goto NA, Hamaker ME, Willems HC, Verhaar MC, Emmelot-Vonk MH. Accidental falling in community-dwelling elderly with chronic kidney disease. *Int Urol Nephrol*. 2019;51:119–127.
13. Boyé NDA, van der Velde N, de Vries OJ, et al. Effectiveness of medication withdrawal in older fallers: results from the improving medication prescribing to reduce Risk Of FALLs (IMPROveFALL) Trial. *Age Ageing*. 2016;46:142–146.
14. Federation of Medical Specialists. *Fall risk assessment in community-dwelling older adults [Federatie Medisch Specialisten, Valrisicobeoordeling thuiswonende ouderen]*. Accessed April 29, 2023. https://richtlijnendatabase.nl/richtlijn/preventie_van_valincidenten_bij_ouderen/valrisicobeoordeling_thuiswonende_ouderen.html?query=valkliniek#considerations
15. Montero-odasso M, van der Velde N, Martin FC, et al. World guidelines for falls prevention and management for older adults: a global initiative. *Age Ageing*. 2022;51:afac205.
16. Podsiadlo D, Richardson S. The timed up and go: a test of basic functional mobility for frail elderly persons. *J Am Geriatr Soc*. 1991;39:142–148.
17. Group TE. EuroQol - a new facility for the measurement of health-related quality of life. *Health Policy (New York)*. 1990;16:199–208.
18. Riley RD, Debray TPA, Collins GS, et al. Minimum sample size for external validation of a clinical prediction model with a binary outcome. *Stat Med*. 2021;40:4230–4251.
19. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. *J Stat Softw*. 2011;45:1–67.
20. Janssen KJM, Vergouwe Y, Donders ART, et al. Dealing with missing predictor values when applying clinical prediction models. *Clin Chem*. 2009;55:994–1001.
21. Steyerberg E, Borsboom G, van Houwelingen H, Eijkemans R, Habbema D. Validation and updating of predictive logistic regression models: a study on sample size and shrinkage. *Stat Med*. 2004;23:2567–2586.
22. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Mak*. 2006;26:565–574.
23. Montero-Odasso MM, Kamkar N, Pieruccini-Faria F, et al. Evaluation of clinical practice guidelines on fall prevention and management for older adults: a systematic review. *JAMA Netw Open*. 2021;4:1–15.
24. Heymans M. psfmi: Prediction Model Selection and Performance Evaluation in Multiple Imputed Datasets. R package version 0.7.1 [Internet]. 2021. Accessed April 29, 2023. https://cran.r-project.org/package=psfmi
25. Brown M. rmda: Risk Model Decision Analysis (version 1.6) [Internet]. 2018. Accessed April 29, 2023. https://cran.r-project.org/web/packages/rmda/index.html
26. Moons KGM, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med*. 2015;162:W1–W73.
27. Carpenter CR, Avidan MS, Wildes T, Stark S, Fowler SA, Lo AX. Predicting geriatric falls following an episode of emergency department care: a systematic review. *Acad Emerg Med*. 2014;21:1069–1082.
28. Meekes WM, Korevaar JC, Leemrijse CJ, van de Goor IA. Practical and validated tool to assess falls risk in the primary care setting: a systematic review. *BMJ Open*. 2021;11:e045431.
29. Palumbo P, Palmerini L, Bandinelli S, Chiari L. Fall risk assessment tools for elderly living in the community: can we do Better? *PLoS One*. 2015;10:e0146247.
30. Kappen TH, van Klei WA, van Wolfswinkel L, Kalkman CJ, Vergouwe Y, Moons KGM. Evaluating the impact of prediction models: lessons learned, challenges, and recommendations. *Diagnostic Progn Res*. 2018;2:1–11.
31. van der Velde N. *A Clinical Decision Support System and Patient Portal for Preventing Medication-related Falls in Older Patients (ADFICE_IT) [Internet]*. Accessed October 20, 2022. https://clinicaltrials.gov/ct2/show/NCT05449470