Externally validated model predicting gait independence after stroke showed fair performance and improved after updating

Anthonia J. Langerak [1,2,a], Alana B. McCambridge [1], Peter W. Stubbs [1], Jesper Fabricius [3], Kris Rogers [1], Camila Quel de Oliveira [1], Jørgen F. Nielsen [3], Arianne P. Verhagen [1,*]

[1] University of Technology Sydney, Graduate School of Health, Discipline of Physiotherapy, Sydney, Australia
[2] Utrecht University, University Medical Center Utrecht, Physical Therapy Sciences, program in Clinical Health Sciences, Utrecht, the Netherlands
[3] Hammel Neurorehabilitation Centre and University Research Clinic, Aarhus University, Hammel, Denmark

Abstract

Objective: To externally validate recent prognostic models that predict independent gait following stroke.

Study Design and Setting: A systematic search identified recent models (<10 years) that predicted independent gait in adult stroke patients, using easily obtainable predictors. Predictors from the original models were assigned proxies when required, and model performance was evaluated in the validation cohort (n = 957). Models were updated to determine if performance could be improved.

Results: Three prognostic models met our criteria, all with high Risk of Bias. Validation data was only available for the Australian model. This model used National Institute of Health Stroke Scale (NIHSS) and age to predict independent gait, using Motor Assessment Scale (MAS) walking item. For validation, Scandinavian Stroke Scale (SSS) was a proxy for NIHSS, and Functional Independence Measure (FIM) locomotion item was a proxy for MAS. The Area Under the Curve was 0.77 (0.74–0.80) and had good calibration in the validation dataset. Adjustment of the intercept and regression coefficients slightly improved discrimination. By adding paretic leg strength, the model further improved (AUC 0.82).

Conclusion: External validation of the Australian model with proxies showed fair discrimination and good calibration. Updating the model by adding paretic leg strength further improved model performance. © 2021 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Keywords: External validation; Stroke; Prognostic model; Gait; Prediction; Model performance

1. Introduction

Stroke is the leading cause of adult disability and the ability to walk is often affected [1,2]. Recovery of independent walking is a common goal of stroke rehabilitation [3]. Although individual recovery patterns vary between individuals [4], studies suggest that recovery of gait can be predicted in the first days after stroke [5–8]. Early and accurate predictions about a patient's recovery can provide clinicians, patients and relatives, with useful information to set realistic rehabilitation goals, prioritize treatments, and facilitate discharge planning [9].

Currently several prognostic models are published that predict gait recovery after stroke, though few are routinely used in clinical practice [9–11]. Prior to a prognostic model being recommended for clinical use, it is necessary for the model to undergo internal and external validation, and assessed for clinical impact [12]. The lack of validation and impact evaluation studies are limiting factors in the translation of prognostic models into clinical practice. Other potential barriers for use in clinical practice may be the complexity of models. For example, models that include predictors not routinely available (e.g., neuroimaging data), or lack tools for clinician use (e.g., decision tree, online calculator), may be more challenging to adopt clinically.

A recent narrative review identified five multivariable prognostic models that predict walking post stroke [9]. Only one model, predicting 10-meter walking speed, had undergone external validation [11]. External validation can be difficult to perform due to the need for a sufficiently large, separate, dataset with similar predictors and outcome measures. However, without external validation the predictive performance of a prognostic model can be overly opti-

---

mistic, due to over-fitting of the data from which the model was derived [12]. External validation is therefore essential to ensure the generalizability of a prognostic model, and validation studies are needed to facilitate the translation of prognostic models from the literature and into clinical practice. During external validation it is common for a model's performance to be poorer than the development study [12]. To build upon previously developed prognostic models, during external validation, it is recommended that prognostic models are updated and/or recalibrated [13].

Therefore, we aimed to externally validate recent (<10 years) prognostic models that predict independent gait in adults with a stroke, using predictors easily obtained in clinical practice. Furthermore, the prognostic models were updated if their performance in the validation cohort was poorer compared to the development study.

2. Patients and methods

2.1. Prognostic model selection

A systematic search was performed to identify recently published prognostic models for post stroke walking (Appendix 1). The search was restricted to April 1, 2009–April 1, 2019. The search strategy was based on a recommended search for prognostic models [14]. Studies were included if they were related to the development and/or validation of a prognostic model/algorithm/nomogram, in adults (≥18 years) with stroke, predicted walking/ambulation/gait (with differentiation between dependent or independent gait), outcome measured >30 days post stroke, and used predictors assessed within 7 days of stroke onset. We defined independent gait as the ability to walk (with walking aid if needed) but without assistance of another person over short distances. Studies were excluded if the model used predictors that were not easily obtained in clinical practice (e.g., imaging, neurophysiological data).

Two authors (PS, AL) independently screened titles and abstracts of articles and disagreements discussed until consensus. Full texts were screened by two authors (AM, AL), and a third (AV) consulted for disagreements. Data extracted (AL) included study design, participant baseline characteristics, setting, location, sample size, time points, outcome measurements, predictors used in the analysis, developed model, calibration and discrimination measures. We contacted authors if articles did not provide sufficient information. Two assessors (AL, AV) independently assessed risk of bias (RoB) using the Prediction model Risk Of Bias Assessment Tool (PROBAST) [15].

2.2. Validation cohort

Baseline data from the Danish Stroke Registry (DSR) [16] were obtained and matched to discharge data from patients with stroke from Hammel Neurorehabilitation Centre and University Research Clinic (HNRC) in Denmark be-

tween July 6, 2011 and November 30, 2018. HNRC is a rehabilitation hospital for patients with moderate to severe acquired brain injury. Baseline data from the DSR obtained within the first few days post stroke were merged with discharge data from HNRC to create the validation cohort. Data handling (Danish Patient Safety Authority, ID3-3013-2831/1), storage (Data Protection Agency Central Region of Denmark, ID1-16-02-734-18), and ethical approval were obtained (UTS-HREC:ETH19-4073N).

2.2.1. Participants

Data from 1,489 patients were obtained, termed the full dataset. To establish the validation cohort, patients from the full dataset were excluded when they were diagnosed with transient ischemic attack (TIA), undergoing outpatient rehabilitation, or able to walk independently at baseline (Scandinavian Stroke Scale (SSS) - gait score <9 in acute care).

2.2.2. Predictors

The DSR contains a range of patient variables from acute care, including baseline demographic characteristics and the SSS [16]. The SSS is a valid and reliable assessment tool commonly used in Scandinavian countries, which measures the severity of stroke impairments [17]. The SSS consists of nine items including consciousness, eye movement, arm motor power, hand motor power, leg motor power, orientation, speech, facial palsy, and gait. Each item consists of scoring categories, ranging from 0 to 12 points, which are scored by the physician or specialized therapist. The maximum possible score is 58, ranging from 0 "unconscious" to 58 "no neurological deficits."

Based on the prognostic models that met inclusion criteria, predictors were matched to variables available in the validation cohort. For variables that did not have a direct match, authors discussed suitable proxies based on published literature until consensus (AL, AM, CQ, PS, AV) [18].

2.2.3. Outcome measures

At HNRC, inpatients are assessed using the Functional Independence Measure (FIM) at admission, every fourth week, and at discharge. The FIM is a valid and reliable tool to measure activity limitations in patients with stroke [19,20]. The FIM is 18 items with motor and cognition subscales, and each item is scored 1–7 [21]. In the present study, gait independence was determined using the "Locomotion: Walk, Wheelchair" item of the FIM. The locomotion item consists of two parts; a categorical score "wheelchair, " "walking," or "both" and a nominal score of independence ranging from 1 to 7 [21].

2.3. Analysis

We summarized the characteristics of the validation cohort with mean and standard deviation (SD), or median

and interquartile range (IQR) for continuous variables (as appropriate) and number and percent (%) of total for categorical variables. Available case analyses were performed as missing data were <10% [22].

To evaluate model performance of the selected prognostic models in the validation cohort we obtained discrimination, calibration, specificity, and sensitivity values. The area under the receiver-operating characteristic curve (AUC) was calculated to infer discriminative power. Calibration was assessed by visual inspection of the observed probabilities plotted against the predicted probabilities. To evaluate a suitable cut-off for independent gait on the FIM locomotion, we investigated model discrimination and calibration using the categories "walking" and "both" and evaluated two FIM cut-off scores (≥5 and ≥6), as both descriptions could potentially fit the outcome measures in the included models (Supplementary 1).

If after validation, model performance did not reach sufficient calibration [23] and comparable discrimination as reported during development, the model was updated in the validation cohort, using the following steps; (1) re-estimation of the intercept, (2) re-estimation of the regression coefficients of the predictors, and (3) re-estimation of the intercept and regression coefficients with new predictor(s) added to the model [13].

Selection of potential variables was based on the literature [5,9] and variable availability our dataset. We used logistic regression to re-estimate the intercept and regression coefficients and to add new variables [24]. Correlations between potential new variables were analyzed and the standard errors of the regression coefficients were reviewed to diagnose multicollinearity between variables. If variables were highly correlated (r≥0.8), one variable was included. For the final model we used backwards stepwise regression analysis, with $P = 0.10$ threshold to remove variables from the model. Bootstrapping (b = 300) was performed to shrink coefficients and prevent overfitting. Discrimination and calibration curves for the model and updates were reported.

IBM SPSS Statistics (v25) was used to calculate summary statistics and *R* (R Core Team (2019) R Foundation for Statistical Computing*)* with packages *"tidyverse,"* [25] *"proc"* [26] *and "rms"* [27] for model validation and adjustments. All data reported according to the reporting guideline [24].

3. Results

3.1. Prognostic model selection

Of 2,728 articles retrieved, three prognostic models met the inclusion criteria: the Australian model, [6] the EPOS model [7] and the TWIST model [8] (Fig. 1). Studies collected hospital inpatient data, with baseline assessments predominantly taken within 7 days post stroke, and outcomes predicted at six months [6,7] or 6- and 12-weeks

post stroke [8] (Table 1). Two studies included patients with both hemorrhagic and ischemic stroke, [6,8] while one study included ischemic patients only [7]. Calibration curves and AUC were not reported in two studies [7,8]. Authors were contacted, and TWIST model data were unavailable due to the statistical approach used (i.e., Classification and regression tree analysis [8]), and no correspondence was obtained from EPOS authors. Overall, all models demonstrated high RoB due to the analysis (Table 2).

3.2. Validation cohort

3.2.1. Participants

From the full dataset (n = 1489) we excluded 532 patients because of: a diagnosis of TIA (n = 11), undergoing outpatient rehabilitation (n = 91), walking independently at baseline (n = 388), and missing baseline walking information (n = 16). In addition, we excluded 26 patients because the time between presentation to acute care and baseline measurement was deemed unrealistic (e.g., values ranging from −1,849 to −10 days or +21 to +760 days). Identification of unrealistic values was based on visual inspection of the values matched to a normal distribution. In the full dataset there were 1.67% missing values, mostly at discharge (i.e., FIM scores at discharge were missing mostly due to the 91 outpatients and 11 TIA patients of which the FIM measurements were not part of the routine assessments). In the validation dataset (n = 957) there were 0.93% missing data overall, with 3.76% missing FIM scores at discharge. The time until baseline measurements was within 7 days post stroke for 95.0% of the included patients. For baseline characteristics of the full dataset, validation cohort, and included studies see Table 3. Patients in the validation cohort had lower age compared to the development cohorts, included both ischemic and hemorrhagic stroke, and included mild (31.0%), moderate (37.2%), and severely (31.8%) affected patients at baseline based on the SSS [28]. In the Australian model development cohort, 43% of the patients were classified as "mild" on the NIHSS at baseline [6], whereas patients included in the TWIST model, were mostly (83%) classified as "moderate to severe" on the NIHSS [8]. The EPOS model development cohort included only ischemic stroke, with the median Barthel Index indicating that patients were mostly highly to totally dependent (median, IQR: 6, 2–10) at baseline [7]. At discharge, 62.0% (567/914 available cases) of the validation cohort achieved independent gait at a median of 72 (IQR: 45–115) days post stroke, compared to 63–79% in the development cohorts at respectively 3- and 6 months post stroke.

3.2.2. Predictors and proxy selection

The predictors (Table 1) and proxy comparison (Supplementary 1) are reported. For the EPOS model, the
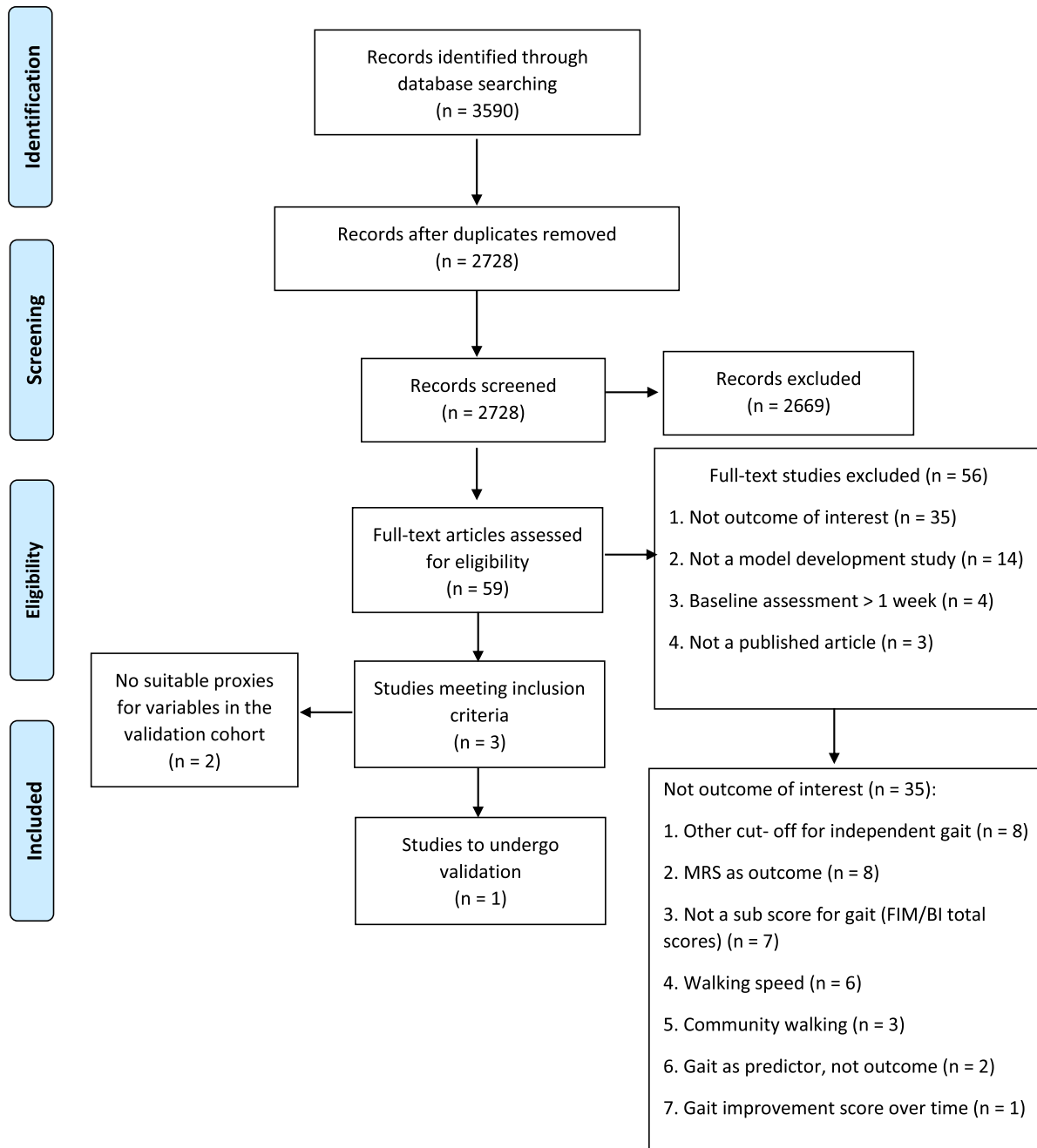
Fig. 1. Flow diagram of the search and selection of studies from the literature. MRS; Modified Rankin Scale. FIM, functional independence measure; FIM, functional independence measure; BI, Barthel index.

description of Motricity Index item lower limb, and for the TWIST model, the Trunk Control Test did not match any variables in the validation dataset (Supplementary 1). Therefore, EPOS and TWIST models were unable to be validated in our dataset.

For the Australian model, we selected SSS total as a proxy for NIHSS total. The SSS score was converted to an NIHSS score using a formula (Supplementary 2) [29] Potential predictors to update the model were selected from the available data and include gender, type of stroke *("is-*

chemic," "hemorrhage," or "other"), mobility *(SSS-gait)* and leg strength *(SSS-leg).*

3.2.3. Outcome

The FIM locomotion item was used as a proxy for the Motor Assessment Scale (MAS) item "walking," which was the outcome measure in the Australian model. The model performed slightly better with FIM cut-off ≥ 6 instead of ≥ 5 as proxy for MAS walking ≥ 3 and was used for all further analyses (Supplementary 3).

Table 1. Characteristics of models predicting independent gait in stroke patients

Reference	Australian model [6]	EPOS model [7]	TWIST model [8]
Study type	Model development	Model development	Model development
Population	200 adult stroke patients: hemorrhage, ischemic and SAH (with symptoms still present after 24hrs).	221 ischemic first-ever anterior circulation stroke patients, presenting with hemiparesis (even after rTPa treatment).	41 adult stroke patients: ischemic and intracerebral hemorrhage with new lower limb weakness and required supervision or assistance for walking.
Setting	Hospital, acute care	Hospital, acute care	Hospital, acute care
Country	Australia	The Netherlands	New Zealand
Time of recruitment	2009–2010	2007–2009	2015–2017
Patients included for analysis	Not able to walk independently at baseline (n = 141)	Not able to walk independently at baseline (n = 189)	Not able to walk independently at baseline (n = 41)
Potential predictors	Age, NIHSS (stroke severity), MAS item 4 (standing up ability), premorbid BI score, Tardieu Scale score for spasticity.	Age, social support, comorbidity, hemisphere of stroke, consciousness at onset, days between stroke onset and baseline assessment, type of stroke, extinction or inattention, hemianopia, conjugate deviation, sensory loss, BI urinary incontinence, TCT item 3 (sitting balance), MI leg, MI arm, FIM arm, FIM leg.	Age, sex, stroke classification, NIHSS (stroke severity), stroke type, CCI (comorbidities), initial FAC, MRC grades for hip flexion, extension and abduction, knee flexion and extension and ankle dorsiflexion and plantarflexion; MI leg, TCT, therapy dose, therapy intensity.
Predictors in final model	Age per 10 years (beta: −0.11) NIHSS (beta: −0.24) Constant: −11.03	TCT-3 (beta: 2.69) MI-leg (beta: 2.08) Constant: −0.98	TCT score MRC grade for hip extension strength
Outcome variable	Independent gait (MAS item walking, ≥3)	Independent gait (FAC ≥4)	Independent gait (FAC ≥4)
Follow-up	6 months	6 months	6 weeks, 12 weeks
Independent gait at follow-up (%)	114/141 participants (80.9%)	122/154 participants (79.2%)	6 weeks: 21/41 (51%) 12 weeks: 26/41 (63%)
AUC (95% CI)	0.84 (0.77–092)	-	-
Calibration curve	Yes	-	-
Hosmer-Lemeshow test goodness of fit	0.7	-	-
Specificity (95% CI)/sensitivity (95% CI)	-	63% (43–78%)/93% (86–96%)	100% (90–100%)/80% (28–99%)

SAH, subarachnoid hemorrhage; rTPa, Tissue plasminogen activator; NIHSS, National Institutes of Health stroke scale; TCT, trunk control test; MI, Motricity Index; FIM, Functional Independence Measure; MAS, Motor Assessment Scale; CCI, Charlson Comorbidity Index; MRC, Medical Research Council scale; -, not reported.

Table 2. PROBAST Risk of Bias assessment per domain and overall

Study	Participants	Predictors	Outcome	Analysis	Overall
Australian model [6]	Low	Low	Low	High	High
EPOS model [7]	Low	Unclear	Unclear	High	High
Twist model [8]	Low	Unclear	Unclear	High	High

3.2.4. Model Validation and Updates

As there were less than 10% missing values, we performed an available case analysis (n = 914) to analyze the performance of the Australian model. First, the validation with proxies had an AUC of 0.77 (95% CI: 0.74–0.80; Table 4). The model showed good calibration; however it slightly overestimated the predicted scores (Fig. 2). As discrimination was lower compared to the development study (AUC 0.84), model updates were performed.

Updating the model's intercept alone (Update A), did not improve calibration or discrimination (Table 4, Fig. 2). Calibration improved after re-estimating the intercept and regression coefficients (Update D). The intercept and regression coefficients decreased relative to the original model after re-estimation. Discrimination was improved by re-estimation of all regression coefficients and further improved by updating the model with a new variable. The variables gender, type of stroke (*"ischemic," "hemor-*

Table 3. Baseline characteristics of the validation and development cohorts

	Full dataset (n = 1489)	Validation cohort (n = 957)	Australian model [6] (n = 114)[I]	EPOS model [7] (n = 154)[II]	TWIST model [8] (n = 41)
Age yrs, mean (sd)	62.6 (12.4)	63.9 (12.2)	median 78 (IQR 67 - 83)	67.5 (14.2)	72 (43–96)
Gender male/female, n (%)	954 (64.1)/535	587 (61.3) /370	53 (47.0) /61	61 (39.6) /93	17 (41.0) /24
Type of stroke, n (%)					
Infarction	1083 (72.7)	699 (73.0)	78 (68.0)	154 (100.0)	45 (85)
– *Cerebral infarction*	1064 (71.5)	690 (72.1)		154 (100.0)	
– *Other infarction*	19 (1.3)	9 (0.9)			
Hemorrhage	259 (17.4)	191 (20.0)	25 (22.0)		6 (15)
– *Intracerebral hemorrhage*	237 (15.9)	182 (19.0)	24 (21.0)		
– *Subarachnoid hemorrhage*	17 (1.1)	7 (0.7)	1 (1.0)		
– *Other hemorrhage*	5 (0.3)	2 (0.2)			
Unknown	136 (9.1)	67 (7.0)	11 (10.0)		
TIA	11 (0.8)	0	0		
Stroke impairment, n (%)	*SSS*: (n = 1450) Mild: 717 (49.4) Moderate: 406 (28.0) Severe: 327 (22.6)	*SSS*: (n = 946): Mild: 293 (31.0) Moderate: 352 (37.2) Severe: 301 (31.8)	*NIHSS*: Mild: 49 (43.0) Moderate: 43 (38.0) Severe: 22 (19.0)	BI: 6 (2–10) Median (IQR)	*NIHSS* Mild: 7 (17.0) Moderate/Severe: 34 (83.0)
Days in acute care, median (IQR)	7 (4–17)	7 (4–14)			
Days in rehabilitation, median (IQR)	50 (28–85)	57 (34–96)			32 (13–82)

sd, standard deviation; TIA, transient ischemic attack; IQR, interquartile range; SSS, Scandinavian Stroke Scale; NIHSS, National Institutes of Health Stroke Scale; BI, Barthel index.

[I] One hundred fourteen patients out of the initial included 200 patients were not able to walk independently at baseline.

[II] One hundred fifty-four patient out of the initial included 221 patients were not able to walk independently at baseline.

Table 4. Discrimination, specificity (spec) and sensitivity (sens) of original and updated Australian model with proxies in the validation cohort

Model	Equation	AUC (95% CI)	Spec/Sens
Original (Cut-off ≥ 6)	$1/(1+e^{-(11.0285\ -\ 0.1053\ age-0.2436\ proxyNIHSS)})$	0.77 (0.74 - 0.80)	0.45/0.9
Update A *Intercept*	$1/(1+e^{-(10.41\ -\ 0.1053\ age-0.2436\ proxyNIHSS)})$	0.77 (0.74 - 0.80)	0.57/0.81
Update B *Intercept, regression coefficient age*	$1/(1+e^{-(5.60717\ -0.03188\ age-0.2436\ proxyNIHSS)})$	0.8 (0.77 - 0.83)	0.61/0.84
Update C *Intercept, regression coefficient proxyNIHSS*	$1/(1+e^{-(10.6146\ -0.1053\ age\ -0.2596proxyNIHSS)})$	0.78 (0.75 - 0.81)	0.59/0.81
Update D *Intercept, regression coefficients age and proxyNIHSS*	$1/(1+e^{-(4.80918\ -\ 0.02765age\ -0.20291proxyNIHSS)})$	0.80 (0.77 - 0.83)	0.59/0.85
Update E *Intercept, all regression coefficients, new variable added*	$1/(1+e^{(2.4685\ -0.0330\ *\ age\ -0.0891\ *\ proxyNIHSS\ +\ 0.3766\ *\ sss_{leg}=2\ -\ 0.2368\ *\ sss_{leg}=4\ +\ 0.3673\ *\ sss_{leg}=5\ -0.2853\ *\ sss_{leg}=6)})*$	0.83 (0.80 - 0.85) 0.82[I]	0.68/0.85

Note: For example, Update E predicts that an 85-year-old patient with SSS total of 35 and the ability to raise a straight leg with reduced strength (SSSleg = 5) at baseline would not walk independently at discharge from inpatient rehabilitation (predicted = 0).
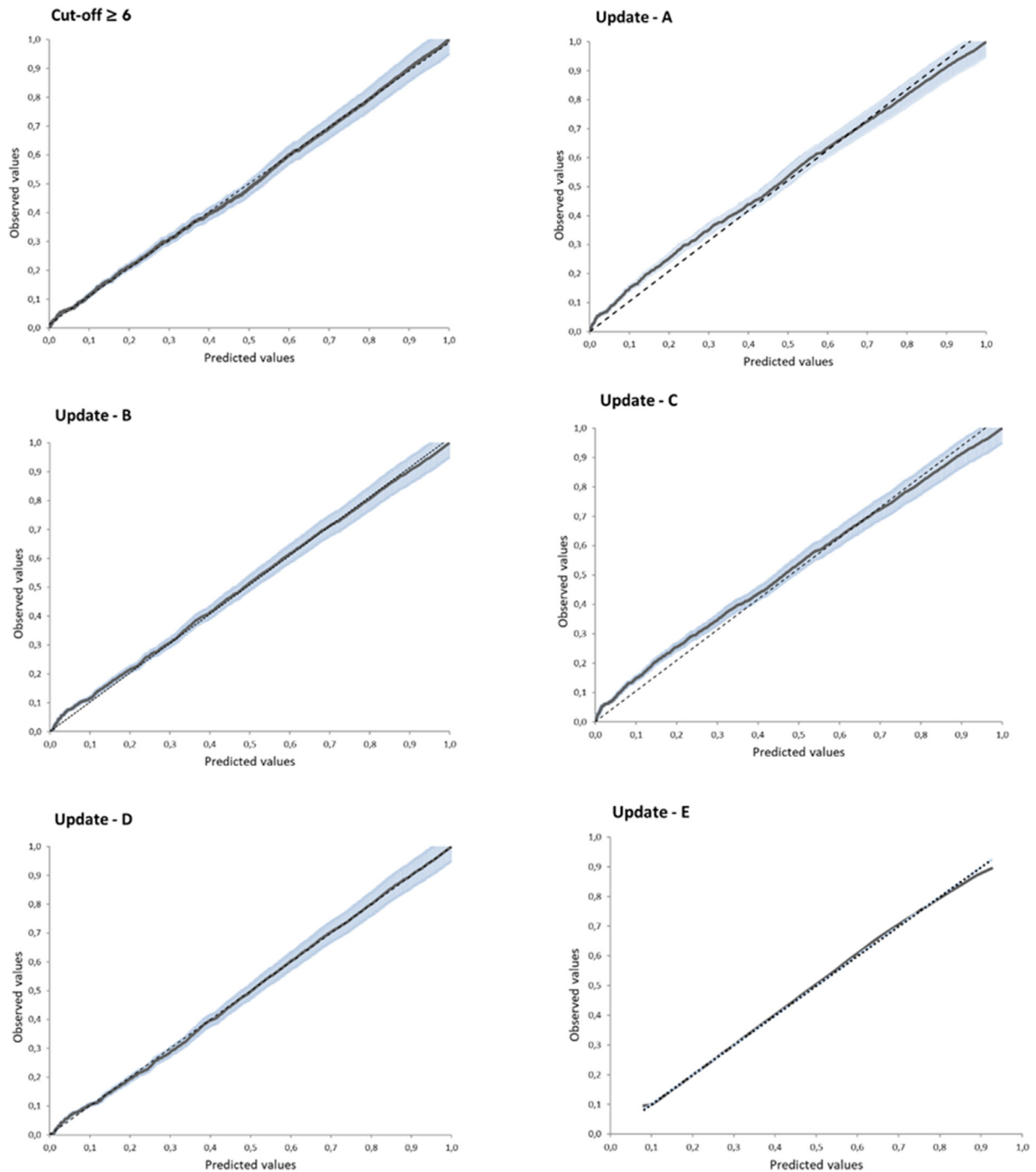
[I] denotes after bootstrapping.

Fig. 2. Calibration plots for independent gait in stroke patients, of the Australian model applied to the validation cohort. Calibration of the original model equation is shown and model adjustments (Update A-E). Calibration for re-estimation of the intercept (Update A), re-estimation of the intercept and regression coefficients of the predictor 'age' (Update B), re-estimation of the intercept and predictor 'NIHSS' (Update C) and finally re-estimation for the intercept and both predictors (Update D) are shown in separate panels. Calibration for 'Update E' shows re-estimation of the intercept and regression coefficients with a new predictor added to the model. The dotted 45° line from zero denotes ideal calibration (slope = 1, intercept = 0) and the solid lines represent the smoothed calibration for each model update. The error bars represent the 5% error margin. For Update E, the grey line represents the original values and the dark line shows the corrected calibration after bootstrapping.

rhage," or "other") and leg strength (SSS-leg) were combined with the original predictors, age and stroke severity (proxy-NIHSS), in a multivariate analysis. SSS-gait showed a standard error of 1.2 and high correlation (r = 0.8) with SSS-total and was excluded from the multivariate analysis. Backwards stepwise selection showed that age, SSS-total, and SSS-leg were significant predictors and were therefore included in the final model. After bootstrapping, performance analysis of the final model showed AUC 0.82 (Table 4) and almost perfect calibration (Fig. 2, Update E). The final updated model showed good sensitivity (0.85) and specificity improved (0.68).

4. Discussion

Three prognostic models met inclusion criteria, though only one could be validated in our dataset. The Australian model, with proxies, for predicting independent gait at a median of 72 days post stroke showed good calibration and fair discrimination in the validation cohort, compared with good calibration and discrimination reported in the development study [6]. By updating the intercept and regression coefficients, calibration and discrimination improved, and the addition of a new predictor variable further improved discrimination of the model. The final updated model exhibited good discrimination and calibration, with good sensitivity and moderate specificity.

External validation typically yields poorer predictive performance of a prognostic model than the development study [12]. The Australian model was not internally validated, therefore predictive performance was likely optimistic which may explain why the model showed poorer discrimination in our validation dataset. In addition, predictive performance in the validation may have been influenced by differences in geographical setting and cohort characteristics [30,31]. In the Australian model development cohort, patients were generally younger and had less severe stroke symptoms than the Danish validation cohort [32]. Moreover, the proportion of patients achieving independent gait in the development cohort was higher than the validation cohort, possibly due to a longer follow-up period (6 months), and/or use of a proxy outcome. The MAS walking item ≥3 requires a patient to walk a minimum of 3 meters without assistance, whereas the FIM requires longer distances [21,33]. It is possible the predictive performance of the Australian model was influenced by these differences during validation.

Rather than rejecting a model showing unfavorable results after external validation, researchers are encouraged to recalibrate or update an already developed model [13,24]. Recalibration builds upon information gained during development, thereby improving the generalizability of a model by incorporating characteristics of a new cohort [13,34]. Updating models prevents researchers from again developing new models with the same clinical goals, which may further delay translation to practice [35]. Pre-

dictive performance can be further improved by adding new predictors that yield independent effects, not already accounted for. There are several promising prognostic factors or biomarkers for stroke recovery reported in the literature that may yield additional prognostic ability [36,37]. Though consideration of the practical and cost implications of implementing a model that includes more complex biomarkers [36–38] is advised, and should be examined when testing the clinical impact of a model.

4.1. Strengths and limitations

The present study is the first external validation and recalibration of an existing prognostic model for post stroke gait. A major strength was the use of a large validation cohort with over 200 events, reducing the risk of overfitting [39]. Another strength was the inclusion of only predictors that were already routinely used by clinicians, however, measurements used to obtain predictors in clinical practice may vary between countries.

Limitations were the use of proxy variables and that two models could not be validated due unsuitable proxies. Although the use of proxies is common in external validation [18,40], this may have reduced the predictive performance of the Australian model. We used proxies for stroke severity (predictor) and independent gait (outcome). For stroke severity we used a conversion factor previously published in the literature to convert NIHSS scores into SSS scores, suggesting a strong correlation between the two scales when converted 3 months after stroke ($R^2 = 0.80$) [29,41]. However, the correlation is weaker ($R^2 = 0.60$) in the acute phase post stroke [29]. For the outcome, the FIM locomotion item was chosen as a proxy for the MAS walking item based on variable descriptions of gait independence. To our knowledge, the relationship between both is yet unknown. In addition, it was not clear if dichotomizing the FIM at ≥5 or ≥6 was the most comparable to the dichotomization of the MAS walking item of ≥3 hence we investigated the predictive performance of the model with both cut-offs. Future research should consider developing an international core outcome set for recovery of gait post stroke to ensure comparable outcomes are used internationally, both clinically and for research [42].

4.2.Future. implications

Before implementing prognostic models in clinical practice, both external validation and impact studies should be conducted to evaluate validity and effectiveness [34]. To our knowledge, there are currently no externally validated prognostic models for gait independence post stroke that have been evaluated for clinical impact. Researchers are encouraged to externally validate already developed models and test the clinical impact of promising models to facilitate the translation from research into clinical practice [24,34,43]. Promising prognostic models should be made

accessible for clinicians via apps or nomograms and the usability of the tool should be evaluated by clinicians. To our knowledge, prognostic models that predict gait independence after stroke are not yet able to be recommended for clinical use.

In conclusion, our final updated model used simple predictors easily obtained in routine practice, demonstrating good calibration and discrimination. The final model was corrected for overfitting by internal validation and should now undergo validation before being tested for clinical impact.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

What is new?

Many prognostic models for gait independence post stroke have been published in the literature.

None had been externally validated and evaluated for clinical impact, therefore limiting translation into clinical practice.

We systematically searched for models predicting independent gait post stroke, appraised the level of bias, and summarized their predictive performance.

The present study is the first external validation of a prognostic model obtained for post stroke gait independence using a large Danish cohort. To improve performance of the model, recalibration and updates were performed and the updated model was corrected for overfitting by internal validation.

The updated model uses easy to obtain parameters and should now undergo validation in a separate dataset before being tested for clinical impact.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.clinepi.2021.03.022.

CRediT authorship contribution statement

Anthonia J. Langerak: Conceptualization, Methodology, Formal analysis, Visualization, Writing - original draft, Writing - review & editing. **Alana B. McCambridge:** Conceptualization, Methodology, Writing - original draft, Writing - review & editing. **Peter W. Stubbs:** Conceptualization, Methodology, Writing - original draft. **Jesper Fabricius:** Resources, Data curation, Writing - review & editing. **Kris Rogers:** Formal analysis, Writing - review & editing. **Camila Quel de Oliveira:** Conceptualization, Writing - review & editing. **Jørgen F. Nielsen:**

Resources, Writing - review & editing. **Arianne P. Verhagen:** Conceptualization, Methodology, Writing - review & editing, Supervision, Project administration.

References

[1] Feigin VL, Abajobir AA, Abate KH, Abd-Allah F, Abdulle AM, Abera SF, et al. Global, regional, and national burden of neurological disorders during 1990–2015: a systematic analysis for the Global Burden of Disease. Study 2015 2017;16:877–97.

[2] Li S, Francisco GE, Zhou P. Post-stroke hemiplegic gait: new perspective and insights. Front Physiol 2018;9 1021. 2018/08/22. doi:10.3389/fphys.2018.01021.

[3] Dobkin BH. Clinical practice. Rehabilitation after stroke. N Engl J Med 2005;352:1677–84 2005/04/22. doi:10.1056/NEJMcp043511.

[4] Kwakkel G, Kollen B, Lindeman E. Understanding the pattern of functional recovery after stroke: facts and theories 2004;22:281–99.

[5] Craig LE, Wu O, Bernhardt J, Langhorne P. Predictors of poststroke mobility: systematic review. Int J Stroke 2011;6:321–7 2011/07/13. doi:10.1111/j.1747-4949.2011.00621.x.

[6] Kwah LK, Harvey LA, Diong J, Herbert RD. Models containing age and NIHSS predict recovery of ambulation and upper limb function six months after stroke: an observational study. J Physiother 2013;59:189–97. doi:10.1016/s1836-9553(13)70183-8.

[7] Veerbeek JM, Van Wegen EE, Harmeling-Van der Wel BC, Kwakkel G. Is accurate prediction of gait in nonambulatory stroke patients possible within 72 hours poststroke? The EPOS study. Neurorehabil Neural Repair 2011;25:268–74 2010/12/28. doi:10.1177/1545968310384271.

[8] Smith MC, Barber PA, Stinear CM. The TWIST algorithm predicts time to walking independently after stroke. Neurorehabil Neural Repair 2017;31:955–64 2017/11/02. doi:10.1177/1545968317736820.

[9] Stinear CM, Smith MC, Byblow WD. Prediction tools for stroke rehabilitation. Stroke 2019;50:3314–22 2019/10/16. doi:10.1161/STROKEAHA.119.025696.

[10] Kwah LK, Herbert RD. Prediction of walking and arm recovery after stroke: a critical review. Brain Sci 2016 6 2016/11/10. doi:10.3390/brainsci6040053.

[11] Bland MD, Sturmoski A, Whitson M, Connor LT, Fucetola R, Huskey T, et al. Prediction of discharge walking ability from initial assessment in a stroke inpatient rehabilitation facility population. Arch Phys Med Rehabil 2012;93:1441–7 2012/03/27. doi:10.1016/j.apmr.2012.02.029.

[12] Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model. research 2013;10:e1001381.

[13] Moons KG, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: II. External validation, model updating, and impact assessment. Heart 2012;98:691–8 2012/03/09. doi:10.1136/heartjnl-2011-301247.

[14] 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:e32844 2012/03/07. doi:10.1371/journal.pone.0032844.

[15] 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;170:51–8 2019/01/01. doi:10.7326/M18-1376.

[16] Johnsen SP, Ingeman A, Hundborg HH, Schaarup SZ, Gyllenborg J. The Danish Stroke Registry. Clin Epidemiol 2016;8:697–702 2016/11/16. doi:10.2147/CLEP.S103662.

[17] Stubbs PW, Mortensen J. Clinimetrics: The Scandinavian stroke scale. J Physiother 2020;66:132 2019/09/16. doi:10.1016/j.jphys.2019.08.010.

[18] Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating. Springer Science & Business Media; 2008.

[19] Gosman-Hedstrom G, Svensson E. Parallel reliability of the functional independence measure and the Barthel ADL index. Disabil Rehabil 2000;22:702–15 2000/12/16. doi:10.1080/09638280050191972.

[20] Ottenbacher KJ, Hsu Y, Granger CV, Fielder RC. The reliability of the functional independence measure: a quantitative review. Arch Phys Med Rehabil 1996;77:1226–32 1996/12/01. doi:10.1016/s0003-9993(96)90184-7.

[21] UDSfM Rehabilitation. The FIM Instrumen; It's background, structure and usefulness. Buffalo; 2012.

[22] Bennett DA. How can I deal with missing data in my study? Aust N Z J Public Health 2001;25:464–9 2001/11/02.

[23] Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW. Calibration: the Achilles heel of predictive analytics. BMC Med 2019;17:230 2019/12/18. doi:10.1186/s12916-019-1466-7.

[24] Collins GS, Altman DG, Reitsma JB, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162:W1–73 2015/01/07. doi:10.7326/M14-0698.

[25] Wickham H, Averick M, Bryan J, Chang W, McGowan L, Francois R, et al. Welcome to the Tidyverse. J Open Source Softw 2019;4:1686.

[26] Robin X, Turck N, Hainard A, et al. Package 'pROC'. 2020. 2012-09-10 09: 34.

[27] Harrell Jr FE, Harrell Jr MFE, Hmisc D. Package 'rms', 229. Vanderbilt University; 2019.

[28] Govan L, Langhorne P, Weir CJ. Categorizing stroke prognosis using different stroke scales. Stroke 2009;40:3396–9 2009/08/08. doi:10.1161/STROKEAHA.109.557645.

[29] Gray LJ, Ali M, Lyden PD, Bath PM. Interconversion of the National Institutes of Health Stroke Scale and Scandinavian Stroke Scale in acute stroke. J Stroke Cerebrovasc Dis 2009;18:466–8 2009/11/11. doi:10.1016/j.jstrokecerebrovasdis.2009.02.003.

[30] Riley RD, Ensor J, Snell KI, Harrel FE, Martin GP, Reitsma JB, et al. External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. BMJ 2016;353:i3140 2016/06/24. doi:10.1136/bmj.i3140.

[31] Debray TP, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KG. A new framework to enhance the interpretation of external validation studies of clinical prediction models. J Clin Epidemiol 2015;68:279–89 2014/09/03. S0895-4356(14)00275-3 [pii] doi:10.1016/j.jclinepi.2014.06.018.

[32] Vergouwe Y, Moons KG, Steyerberg EW. External validity of risk models: use of benchmark values to disentangle a case-mix effect from incorrect coefficients. Am J Epidemiol 2010;172:971–80 2010/09/03. doi:10.1093/aje/kwq223.

[33] Carr JH, Shepherd RB, Nordholm L, Lynne D. Investigation of a new motor assessment scale for stroke patients. Phys Ther 1985;65:175–80 1985/02/01. doi:10.1093/ptj/65.2.175.

[34] Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. PLoS medicine 2013;10:e1001381.

[35] Cowley LE, Farewell DM, Maguire S, Kemp AM. Methodological standards for the development and evaluation of clinical prediction rules: a review of the literature. Diagnostic Prognostic Res 2019;3:16. doi:10.1186/s41512-019-0060-y.

[36] Boyd LA, Hayward KS, Ward NS, Stinear CM, Rosso C, Fisher RJ, et al. Biomarkers of stroke recovery: consensus-based core recommendations from the stroke recovery and rehabilitation roundtable. Neurorehabil Neural Repair 2017;31:864–76 2017/12/14. doi:10.1177/1545968317732680.

[37] Stinear CM. Prediction of motor recovery after stroke: advances in biomarkers. Lancet Neurol 2017;16:826–36 2017/09/19. doi:10.1016/S1474-4422(17)30283-1.

[38] Connell LA, Smith MC, Byblow WD, Stinear CM. Implementing biomarkers to predict motor recovery after stroke. NeuroRehabilitation 2018;43:41–50 2018/07/30. doi:10.3233/NRE-172395.

[39] Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. Stat Med 2016;35:214–26 2015/11/11. doi:10.1002/sim.6787.

[40] Ettema RG, Peelen LM, Schuurmans MJ, Nierich AP, Kalkman CJ, Moons KG. Prediction models for prolonged intensive care unit stay after cardiac surgery: systematic review and validation study. Circulation 2010;122:682–9 687 p following p 689. 2010/08/04. doi:10.1161/CIRCULATIONAHA.109.926808.

[41] Goldie FC, Fulton RL, Frank B, Lees KR. Interdependence of stroke outcome scales: reliable estimates from the Virtual International Stroke Trials Archive (VISTA). Int J Stroke 2014;9:328–32 2013/11/12. doi:10.1111/ijs.12178.

[42] Kwakkel G, Lannin NA, Borschmann K, English C, Ali M, Churilov L, et al. Standardized measurement of sensorimotor recovery in stroke trials: consensus-based core recommendations from the stroke recovery and rehabilitation roundtable. Neurorehabil Neural Repair 2017;31:784–92 2017/09/25. doi:10.1177/1545968317732662.

[43] Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. BMJ 2009;338:b606 2009/06/09. doi:10.1136/bmj.b606.