Journal of Clinical Epidemiology 156 (2023) 53–65

Journal of Clinical Epidemiology

REVIEW ARTICLE

Regression-based prognostic models for functional independence after postacute brain injury rehabilitation are not transportable: a systematic review

Uwe M. Pommerich[a,*], Peter W. Stubbs[b], Peter Preben Eggertsen[a], Jesper Fabricius[a], Jørgen Feldbæk Nielsen[a]

[a]*Hammel Neurorehabilitation Centre and University Research Clinic, Department of Clinical Medicine, Aarhus University, Hammel, Denmark*
[b]*Discipline of Physiotherapy, Graduate School of Health, University of Technology Sydney, Ultimo 2007, Australia*

Accepted 2 February 2023; Published online 9 February 2023

Abstract

Background and Objectives: To identify and summarize validated multivariable prognostic models for the Functional Independence Measure® (FIM®) at discharge from post-acute inpatient rehabilitation in adults with acquired brain injury (ABI).

Methods: This review was conducted based on the recommendations of the Cochrane Prognosis Methods Group and adheres to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines. Three databases were systematically searched in May 2021 and updated in April 2022. Main inclusion criteria were: a) adult patients with ABI, b) validated multivariable prognostic model, c) time of prognostication within 1-week of admission to post-acute rehabilitation, and d) outcome was the FIM® at discharge from post-acute rehabilitation.

Results: The search yielded 3,169 unique articles. Three articles fulfilled the inclusion criteria, accounting for $n = 6$ internally and $n = 2$ externally validated prognostic models. Discrimination was estimated as an area under the curve between 0.76 and 0.89. Calibration was deemed to be assessed insufficiently. The included models were judged to be of high risk of bias.

Conclusion: Current prognostic models for the FIM® in post-acute rehabilitation for patients with ABI lack the methodological rigor to support clinical use outside the development setting. Future studies addressing functional independence should ensure appropriate model validation and conform to uniform reporting standards for prognosis research.

Keywords: Acquired brain injury; Cerebrovascular disorder; Rehabilitation; Prognosis; Prognostic prediction model; Systematic review

1. Introduction

Rehabilitation aims to assist individuals to regain a meaningful life and maintain the maximum level of functional independence and societal participation post-injury [1]. Worldwide, acquired brain injury (ABI) contributes to disability-adjusted life years and a loss of productivity [2]. Post-acute ABI rehabilitation is usually provided during hospitalization in inpatient rehabilitation facilities, depending on the severity of the ABI [3]. Here, functional independence is an important outcome and often considered the definition of 'successful rehabilitation' as it is associated with long-term functional ability, improved quality of life, and reduced caregiver burden [4–7]. Indirectly, lack of functional independence affects healthcare spending [8]. Purchasing power parity adjusted estimates from Europe in 2010 indicate direct and indirect healthcare

Funding: The present research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. U.M.P. has received a grant from *Helsefonden* partially covering the salary during his PhD education, which this study represents a part of. Helsefonden had no influence on the preparation or execution of the present systematic review.

Conflict of interest: P.W.S., P.P.E., J.F., J.F.N.: None.

Author Contributions: U.M.: conceptualization, methodology, project administration, investigation, formal analysis, writing-original draft, review & editing; PW Stubbs: conceptualization, methodology, investigation, supervision, writing-original draft, review & editing; PP Eggertsen: investigation, writing–review & editing; J Fabricius: investigation, writing–review & editing; JF Nielsen: conceptualization, supervision, writing–review & editing; All listed authors meet the ICMJE criteria for authorship and have approved the final.

* Corresponding author. Hammel Neurorehabilitation Centre and University Research Clinic, Department of Clinical Medicine, Aarhus University, Voldbyvej 15, 8450 Hammel, Denmark. Tel.: Phone +45-78419090; fax +45-78419682.

E-mail address: uwepom@clin.au.dk (U.M. Pommerich).

https://doi.org/10.1016/j.jclinepi.2023.02.009

What is new?

Key findings

- Six internally validated prognostic models for the prediction of functional independence in postacute brain injury rehabilitation were identified; which all investigated patients with stroke.
- The identified prognostic models were judged to be of high risk of bias.
- Two models were externally validated for transportability, yet these models used unclear procedures.

What this adds to what is known?

- The identified prognostic models lack the methodological rigor to be useful in clinical practice.
- No appropriate prognostic models for functional independence in the post-acute rehabilitation setting for non-stroke patients with acquired brain injury were identified.

What is the implication, what should change now?

- Novel or updated prognostic models for functional independence should be developed with internal and external validation based on existing methodological guidelines.
- Although all included prognostic models focused on patients with stroke, these models may also be relevant for persons with other brain injuries, such as traumatic brain injury or anoxic brain injury.

and rehabilitation costs for stroke (≈€64.0 Billion) and traumatic brain injury (TBI) (≈€33.0 Billion) [9]. For example, in Denmark the direct costs of stroke and TBI were ≈€1.0 Billion and ≈€500 Million in 2015, respectively [8,10]. In addition, the incidence of stroke is projected to increase and stroke mortality is projected to decrease, putting more people in need of rehabilitation [11]. To limit rising healthcare costs the Danish healthcare system, similar to other countries [12], is transforming to value-based healthcare, that is, achieving the best outcome at the lowest cost [13,14]. In this context, prognosis research in rehabilitation is particularly important. Empirically derived prognoses for meaningful rehabilitation outcomes (e.g., functional independence) at discharge from inpatient rehabilitation may provide information on predicted recovery potential to staff, patients and families to aid patient-centered rehabilitation [15] and clinical decision making [16]. Furthermore, a prognosis of the most likely functional level may allow patients and their families to plan/prepare for a life after hospitalization, for example, whether a return to previous living arrangements appears feasible. Given this, measuring functional independence is of importance. Two frequently used measures of functional independence are the Functional Independence Measure® (FIM®) [17] and the Barthel Index [18]. The FIM® is frequently used in research and post-acute rehabilitation settings and generates comprehensive outcome information based on clinically relevant items [19]. The advantage of the FIM® over the Barthel Index is a cognitive domain which is highly relevant in individuals with an ABI [20,21].

Existing prognosis research in rehabilitation is sparse and considered at high risk of bias with few meaningful patient-centered rehabilitation outcomes [21–26]. Findings from previous, seemingly related, systematic reviews [21,24–28] are not applicable here due to differences in the objectives and eligibility criteria related to the following: a) included populations (e.g., including children), b) time of prognostication (e.g., predicted in the acute settings), c) predicted outcomes (e.g., mortality), and d) the type of prognostic model (i.e., non-internally validated). Additionally, as most of these reviews were performed some years ago, there is a current knowledge gap of recently validated multivariable prognostic models for rehabilitation prognosis in patients with ABI.

The objective of the present study was to identify validated multivariable prognostic models for the prediction of the FIM® score [17] at discharge from post-acute inpatient rehabilitation in adult patients with ABI.

2. Methods

The present review was conducted in accordance with the Prognosis Research Strategy (PROGRESS) framework [29]. Guidelines from the updated Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [30] and the Cochrane Prognosis Methods Group [31–33] were used. The CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS) and the PICOTS (Population, Index, Comparator, Outcome, Timing, Setting) acronym [32] (see Fig. 1) were also used to determine the inclusion criteria and search strategy. The review protocol is registered in PROSPERO (https://www.crd.york.ac.uk/prospero/): CRD42021257098.

2.1. General eligibility criteria

Included studies were: 1) validated multivariable prognostic models [29,32] for individualized outcome prediction developed from longitudinal data, 2) peer-reviewed articles, 3) articles published after the year 2000 due to current advances in rehabilitation for ABI and an increased focus in multivariable prognostic modelling, in terms of guiding principles and improvement of methodological quality [23,34,35], and 4) Articles published in English,

Population	Adults following ABI including: ischaemic or haemorrhagic stroke; subarachnoid haemorrhage; traumatic brain injury; hypoxic or anoxic brain injury; encephalitis or meningitis; primary brain tumours.
Index	Internally or externally validated multivariable prognostic models according to the PROGRESS framework i.e. a combination of multiple variables (predictors) to estimate the likelihood of certain events in the future, for which an individual is at risk. This includes both derivation, updating or sole validation of prognostic models.
Comparison	This is irrelevant as the current study is based on including all eligible prognostic models and not on comparing two distinct models.
Outcome	Seven FIM® dimensions are considered eligible: 1) Total, 2) motor, or 3) cognitive FIM® scores at discharge from post-acute inpatient rehabilitation; 4) Functional Independence Staging system grade at discharge; 5) Total, 6) motor, or 7) cognitive FIM® gain (i.e. change from admission to discharge)
Timing	Time of prognostication: within the first week of admission to post-acute inpatient rehabilitation. Time of outcome: within the last week prior to discharge. Timeframe for prognosis: duration of post-acute rehabilitation.
Setting	Post-acute inpatient rehabilitation i.e. temporary comprehensive rehabilitation provided by a hospital or inpatient rehabilitation facility following discharge from acute care /consultative rehabilitation.

Fig. 1. Detailed eligibility criteria according to the PICOTS (Population, Index, Comparator, Outcome, Timing, Setting) acronym.

Danish, Swedish, Norwegian, or German as these were the languages spoken by the authorship team. The population of interest was adults with ABI in a post-acute comprehensive inpatient rehabilitation setting [36]. The included outcome was the FIM®. The FIM® has adequate psychometric properties in ABI populations [37–39] and has been used in various clinical settings worldwide. The FIM® contains 18 items (motor domain: 13 items and cognitive domain: 5 items) scored on a 7-point scale. FIM® scores are often reported as the total score (18–126 points) or sum of motor domain items (13–91 points) or cognitive domain items (5–35 points); with higher scores indicating a higher functional level. The functional independence staging (FIS) system is another utilization of the FIM® [40]; it compiles FIM®-item scores into hierarchical stages of activity profiles according to the order of expected recovery (see Fig. 1 for further details of the eligibility criteria).

2.2. Exclusion criteria

Excluded studies were: conference abstracts and clinical impact studies, machine learning approaches (i.e., automatic learning from data [41]) such as classification and regression tree (CART) type prediction algorithms (see Section 4.4), and studies measuring the outcomes of interest but not reporting the results for a prognostic model.

2.3. Search methods for identification of studies

Electronic searches were performed in May 2021, in PubMed (open PubMed interface), EMBASE (Ovid interface), and Web of Science (Core collection). Search filters were applied for publication date (after 1st January, 2000) and publication language. The searches were repeated on 22nd April, 2022. Both synonyms and appropriate subject or index headings of relevant search terms were used during electronic searches. Search filters developed for the identification of prediction research [42,43] were included in the search strategy. An information specialist was consulted with the electronic search strategy. The detailed search strategy can be found in the online Supplementary Material. In addition, reference lists of eligible studies, relevant systematic reviews, and the personal collections of the authors were inspected for further potentially eligible studies.

2.4. Data collection

2.4.1. Selection of studies

Identified references were exported into Endnote X8 (Clarivate, Philadelphia, USA) and duplicate references were identified and removed both automatically based on author(s), publication year, and title, and manually by one author (UMP). Screening and selection of studies were managed using Covidence (Covidence Ltd, Melbourne, Australia). Titles and abstracts were screened for eligibility independently by two authors (UMP, PPE). Thereafter, full text articles were obtained and reviewed (UMP, PWS). Disagreements were resolved by consulting a third reviewer (JF).

2.4.2. Data extraction and management

Data were extracted by UMP using Covidence, and verified by JF. Disagreements were resolved by discussion. Extraction forms included items recommended in the CHARMS checklist [32] and the TRIPOD guidelines [44] (see online Supplementary Material).

2.4.3. Assessment of risk of bias in included studies

Two authors (UMP, PWS) independently assessed the risk of bias of each distinct prognostic model, using the Prediction model Risk of Bias ASsessment Tool (PROBAST) [45]. The PROBAST tool provides an assessment for the key domains—participants, predictors, outcome, and analysis. Each domain contains 2 to 9 signalling questions to aid judgement. Risk of bias is rated as low, high or unclear. Overall risk of bias was rated based on PROBAST recommendations [45]. Both review authors discussed the

PROBAST items and calibrated their assessment on 2 studies not included in the present review. All disagreements were resolved by discussion.

2.5. Data synthesis

A planned meta-analysis [46] was deemed unfeasible due to few identified prognostic models (see Section 3). For the same reason, we did not grade the certainty of evidence [47]. Hence, a narrative approach was used to describe the included multivariable prognostic models. Due to lack of reporting of information on model performance we inferred likelihood ratio statistics, coefficients of determination, and optimism of included models using published formula [48].

3. Results

Electronic searches yielded 4,595 articles (Fig. 2). After removal of duplicates, 3,169 articles were screened for inclusion by title and abstract. Finally, 240 full text articles were reviewed, resulting in $n = 3$ articles fulfilling the inclusion criteria [49–51]. These three articles included the

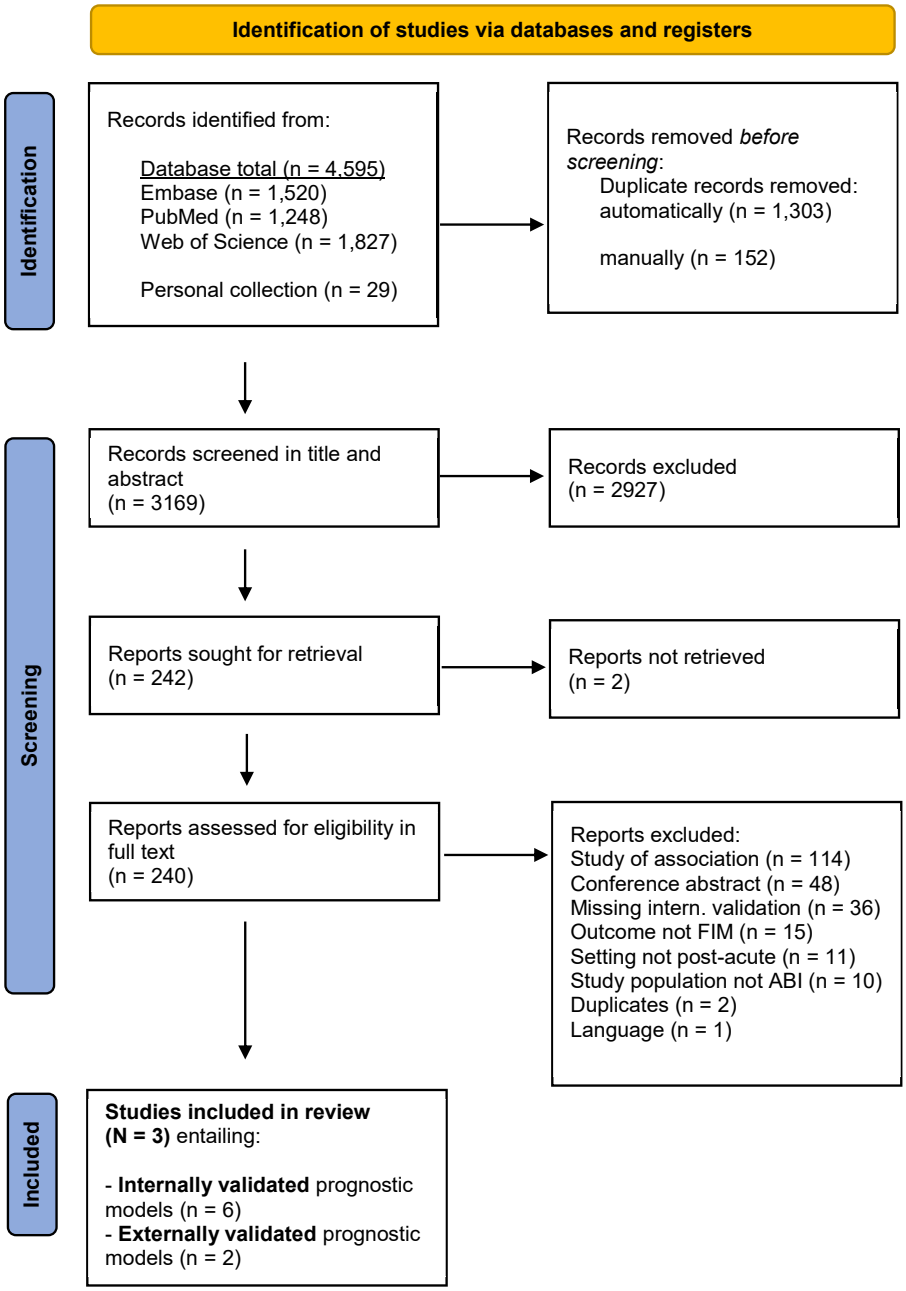

Fig. 2. Flow diagram of study identification and selection.

Table 1. Characteristics of included studies

Reference	Country	*n*	Age	Gender % males	Stroke type %	Rehabilitation admission FIM® score median (IQR)			Neglect[a] %	Days from injury until rehabilitation admission, Median (IQR)	Rehabilitation length of stay (days), mean (SD)
						Total	Motor	Cognitive			
Scrutinio et al. (2017) [50]	Italy	717	Mean (SD) 72 (12)	57.2	Ischemic: 81.2 Hemorrhagic: 18.8	40 (27–54)	19 (14–28)	19 (10–26)	9.9	15 (11–20)	52 (11)
Scrutinio et al. (2019) [49]	Italy	951	Median (IQR) 73 (65–80)	54.8	Ischemic: 82 Hemorrhagic: 18	35 (24–45)	17 (14–23)	16 (9–25)	14	25 (17–38)	56 (14)
García-Rudolph et al. (2021) [51]	Spain	710	Mean (SD) 52 (10)	61.8	Ischemic: 53 Hemorrhagic: 47	47 (30–63)	25 (16–38)	19 (11–27)	35.9	37 (23–56)	75 (31)

Abbreviations: FIM®, Functional Independence Measure®; IQR, interquartile range.
All reported characteristics refer to the derivation cohort, that is, the sample used to develop the prognostic models in.
[a] The definition of neglect and criteria for diagnosis were not specified.

development of $n = 6$ prognostic models and external validation of $n = 2$ prognostic models. Four other articles were discussed for inclusion due to disagreements in the review process but were ultimately excluded due to missing internal validation [52] and the setting not fulfilling the inclusion criteria [53,54]. The fourth article [55] was an external validation of the model presented by Inouye [52]. However, modelling procedures were described insufficiently to assess whether eligibility criteria were fulfilled. The authors did not respond to our inquiry concerning the modelling procedures.

3.1. Included study populations

The participants in two included studies [49,50] were from the same source population and were likely partially overlapping because inclusion criteria were similar. The remaining participants from the third study [51] were recruited from another centre in a different country. All included participants presented with either ischemic or hemorrhagic stroke, and were mostly males. Average age (±SD) differed between studies and ranged from 50 (±10) [51] to 72 (±12) [50]; as did the prevalence of neglect and ischemic stroke. Median total-, motor-, and cognitive FIM® scores at admission ranged from 35 to 47, 17–25, and 16–19, respectively [49–51]. See Table 1 for further details.

3.2. Description of prognostic models

Three prognostic models were developed to predict the motor FIM® score at discharge either as treatment failure (<37 points; model I) [49] or mild stroke impairment (>61 points; model II & III) [50,51], respectively. Two models were developed to predict the achievement of the FIS system grade 5 or higher (model IV & V) [50,51]. Finally, one model was developed predicting the achievement of a clinically important improvement in motor FIM® scores (>25 points; model VI) [50]. All included prognostic models predicted dichotomized outcomes. Three models (II, IV & VI) were internally validated in a different cohort of patients from other centres under the same Rehabilitation Institute [50]. Model I was internally validated using bootstrapping; Models III & V were also presumed to be validated with bootstrap procedures, although not explicitly stated [51]. Two models (VII & VIII) were self-described as externally validating models II & IV geographically, that is, in a different setting from the development setting [51]. The effective sample size ranged from 5 (model IV) to 44 (model I) events-per-candidate-predictor-parameter (EPP; sometimes referred to as Events-per-Variable, EPV). A summary of the characteristics of included studies and prognostic models are shown in Tables 1 and 2.

3.3. Risk of bias

All included models were rated as high risk of bias. Definitions and descriptions of the assessment procedures for candidate predictors were sparse and non-transparent. For example, although neglect (yes/no) was included in five prognostic models (models I, II, II, V, and VI) [49–51], no diagnostic criteria or assessment procedures were provided. Second, suboptimal or unclear modelling and selection procedures may have introduced bias in the analysis. For example, in five prognostic models (models II–VI) variable selection was based on stepwise forward selection. Furthermore, selection procedures were not repeated in any internal validation of the included prognostic models. Figure 3 summarizes the PROBAST assessment in these models.

3.4. Modelling procedures and model performance

Nineteen (models II, IV & VI) and 16 (models III & V) candidate predictors were tested for inclusion in five prognostic models [50,51] based on forward stepwise selection

Table 2. Overview of included prognostic models

Reference in text model nr	Predicted outcome	*n*	Events (*n*)	Events-per-candidate-predictor-parameter (EPP)	Candidate predictors (*n*)	Predictors included in the final model	Selection procedure
Internally validated prognostic models							
Scrutinio et al. (2019) model I [49]	Discharge motor FIM® score ≤37 points	951	440	44[a]	10[d]	Age, Stroke type, Admission Motor FIM score, Admission cognitive FIM® score, Neglect	Backward stepwise selection With $P > 0.157$ For Exclusion
Scrutinio et al. (2017) model II [50]	Discharge motor FIM® score >61 points	717	206	11[b]	19[e]	Age, Injury-Rehabilitation admission interval, Unilateral neglect, Admission Motor FIM® score, Admission cognitive FIM® score	Forward stepwise with $P < 0.05$ for addition
García-Rudolph et al. (2021) model III [51]	Discharge motor FIM® >61 points	710	302	19[c]	16[f]	Age, onset-rehabilitation admission interval, unilateral neglect, Admission motor FIM® score, aphasia	Forward stepwise selection with $P < 0.05$ for addition
Scrutinio et al. (2017) model IV [50]	Functional Independence Staging Grade ≥5	717	100	5[b]	19[e]	Age, Gender, Injury- rehabilitation admission interval, Admission motor FIM® score, Admission cognitive FIM® score	Forward stepwise with $P < 0.05$ for addition
García-Rudolph et al. (2021) model V [51]	Functional Independence Staging Grade ≥5	710	148	9[c]	16[f]	Age, onset-rehabilitation admission interval, unilateral Neglect, admission motor FIM® score, aphasia	Forward stepwise selection with $P < 0.05$ for addition
Scrutinio et al. (2017) model VI [50]	Motor FIM® gain ≥25 points	NR[g]	309	Unclear	Unclear[g]	Age, injury- rehabilitation admission interval, Side of impairment, Stroke type, unilateral neglect, Admission motor FIM® score, Admission cognitive FIM® score	Forward stepwise with $P < 0.05$ for addition
Externally validated prognostic models							
García-Rudolph et al. (2021) model VII [51]	Discharge motor FIM® >61 points	710	302	Unclear	Irrelevant	Age, onset-rehabilitation admission interval, unilateral neglect, admission motor FIM® score, admission Cognitive FIM® score	Irrelevant
García-Rudolph et al. (2021) model VIII [51]	Functional Independence Staging Grade ≥5	710	148	Unclear	Irrelevant	Age, onset-rehabilitation admission interval, sex, admission motor FIM® score, admission cognitive FIM® score	Irrelevant

Abbreviations: AUC, area under the receiver operated characteristic curve; EPP, Events per candidate predictor parameter (often referred to as EPV, events per variable); FIM®, Functional Independence Measure®; and NR, not reported.

[a] This estimate is likely an overestimation. The functional form of some predictors is unclear. For example, according to the results presented by Scrutinio et al. 2019 (their Table 4), the regression coefficient for age is per 5-year increase, indicating a categorization. It remains unclear if this was the case as the regression coefficient itself is presented as linear increase (i.e., per 1 age-unit difference).

[b] The EPP is realistically lower as an interaction was also tested (year of admission x centre in 4 categories: requiring 3 parameters).

[c] This is a crude estimate. Some variables or their categorization were not clearly defined. It is therefore unclear exactly how many parameters were estimated.

[d] age, sex, marital status, diabetes mellitus, onset-rehabilitation admission interval, stroke type, side of impairment, neglect, admission motor FIM®, and admission cognitive FIM®.

[e] age, sex, marital status, employment status, hypertension, diabetes mellitus, chronic obstructive pulmonary disease (COPD), coronary heart disease, atrial fibrillation, onset-rehabilitation admission interval, stroke type, side of impairment, aphasia, unilateral neglect, admission motor FIM® score, admission cognitive FIM® score, blood urea nitrogen, estimated glomerular filtration rate (eGFR), and haemoglobin.

[f] age, sex, marital status, employment status, hypertension, diabetes mellitus, dyslipidaemia, Body Mass Index, atrial fibrillation, onset-rehabilitation admission interval, stroke type, side of impairment, aphasia, unilateral neglect, admission motor FIM®, admission cognitive FIM®.

[g] The prognostic model is reported in the Supplementary Material to the article; it is not explicitly stated whether exactly the same derivation/validation cohort and candidate predictors were used for this model as for the other 2 reported models (II & IV) or if changes were made.

Derivation			Validation			
Apparent model performance	Discrimination AUC (95% CI)	Calibration	Type of validation	Optimism adjusted model performance	Discrimination AUC (95% CI)	Calibration
Not reported (NR)	0.834 (0.809 –0.859)	Hosmer-Lemeshow χ^2 7.77 ($P = 0.456$)	Resampling 200 bootstrap eplications, calculation of shrinkage factor	NR shrinkage factor 0.965	0.831 (0.812 –0.857)	NR
NR	0.883 (0.858 –0.910)	Hosmer-Lemeshow χ^2 4.12 ($P = 0.249$)	Validation cohort $n = 875$ recruited from 3 centres (incl. development centre) at varying periods	NR	0.886 (0.840 –0.892)	Hosmer-Lemeshow χ^2 8.86 ($P = 0.115$) Calibration plot by quintiles
NR	0.894 (0.857 –0.929)	Hosmer-Lemeshow χ^2 10.40 ($P = 0.23$)	Resampling 2000 bootstrap replications	NR	NR	Calibration plot, unclear categories and smoother function
NR	0.913 (0.884 –0.942)	Hosmer-Lemeshow χ^2 1.20 ($P = 0.754$)	Validation cohort $n = 875$ recruited from 3 centres (incl. development centre) at varying periods	NR	0.850 (0.815 –0.885)	Hosmer-Lemeshow χ^2 34.5 ($P = 0.001$) Calibration plot by quintiles
NR	0.845 (0.789 –0.900)	Hosmer-Lemeshow χ^2 6.94 ($P = 0.54$)	Resampling 2000 bootstrap replications	NR	NR	Calibration plot, unclear categories and smoother function
NR	0.754 (0.718 –0.790)	Hosmer-Lemeshow χ^2 8.94 ($P = 0.111$)	Validation cohort NR[g]	NR	0.757 (0.726 –0.789)	Hosmer-Lemeshow χ^2 4.57 ($P = 0.206$)
Irrelevant	Irrelevant	Irrelevant	External geographical validation: $n = 710$ recruited from 1 centre different than the development cohort	NR	0.873 (0.833 –0.915)	Hosmer- Lemeshow χ^2 6.07 ($P = 0.63$) Calibration plot, unclear categories and smoother function
Irrelevant	Irrelevant	Irrelevant	External geographical validation $n = 710$: recruited from 1 centre different than the development cohort	NR	0.803 (0.749 –0.857)	Hosmer-Lemeshow χ^2 8.91 ($P = 0.34$) Calibration plot, unclear categories and smoother function

Reference	In text model number	Predicted outcome	Participants	Predictors	Outcome	Analysis	Overall
Internally validated prognostic models							
Scrutinio et al. (2019)	I	Discharge motor FIM® ≤37 points	?	?	-	-	-
Scrutinio et al. (2017)	II	Discharge motor FIM® >61 points	+	-	+	-	-
García-Rudolph et al. (2021)	III	Discharge motor FIM® >61 points	+	-	+	-	-
Scrutinio et al. (2017)	IV	Functional Independence Staging Grade ≥5	+	-	+	-	-
García-Rudolph et al. (2021)	V	Functional Independence Staging Grade ≥5	+	-	+	-	-
Scrutinio et al. (2017)	VI	Motor FIM® gain ≥25	+	?	-	-	-
Externally validated prognostic models							
García-Rudolph et al. (2021)	VII	Discharge motor FIM® >61 points	+	-	+	-	-
García-Rudolph et al. (2021)	VIII	Functional Independence Staging Grade ≥5	+	-	+	-	-

FIM® = Functional Independence Measure, + = low risk of bias, - = high risk of bias, ? = unclear risk of bias

Fig. 3. Prediction model risk of bias assessment tool (PROBAST) summary for included prognostic models.

with a significance level of $P < 0.05$ for adding the predictor. An interaction between year of admission and rehabilitation centre was tested in three models (II, IV & VI). Univariable analysis and the order of addition of variables to the models were not reported. In the remaining model (model I) [49], ten candidate predictors, identified through a literature review, were tested for inclusion based on backward stepwise selection with $P > 0.157$ for exclusion. The functional form, for example, potential nonlinearity, of some the candidate predictors entered into the model remains unclear (e.g., age). The external validation procedure of two models (VII & VIII) was not described clearly. It appears that the original models were refitted in the validation data using cross-validation; the assessment of the performance of the original model (i.e., the linear predictor with its original coefficients) in the validation data was not reported.

3.4.1. Overall model performance

Overall performance measures such as the coefficient of determination (R^2) were not reported. Inferring the apparent model performance from the reported area under the receiver operated characteristic curve (AUC) estimates [48] yielded the following estimates of the apparent Cox-Snell R^2: 0.39 (model I); 0.32 (model II); 0.44 (model III); 0.19 (model IV); 0.21 (model V); and 0.25 (model VI). Shrinkage (i.e., optimism adjustment) was reported for model I [49] and estimated as 0.965. The precise estimation method was not reported. For the remaining prognostic models, shrinkage was not reported. Inferring the global shrinkage factor through the closed form provided by Van Houwelingen and Le Cessie [48,56] yielded the following estimates: 0.92 (model II); 0.96 (model III); 0.85 (model IV); 0.90 (model V); and 0.89 (model VI).

3.4.2. Discrimination

Discrimination, using AUC was reported for all included prognostic models. The AUC in the validation samples was lowest for model IV and estimated as 0.76 (95% CI 0.73–0.79); the highest observed AUC was in model II and estimated as 0.89 (95% CI 0.84–0.89). It is unclear if the AUC estimates reported for models III & V are for the apparent or optimism-adjusted models. Further details are reported in Table 2.

3.4.3. Calibration

Calibration was assessed in all included prognostic models [49–51] with estimation based on the Hosmer-Lemeshow test (Table 2). Calibration plots were presented for models II–V [50,51] indicating potential miscalibration for models II & IV in the validation cohort on visual inspection. A moderate overestimation can be detected for model II although both underestimation and overestimation of extreme values can be detected for model IV. The interpretation of the calibration plots for models III, V, VII & VIII [51] is hindered by unclear optimism-adjustment procedures. A potential miscalibration appears for both the newly developed model V and the externally validated model VIII.

4. Discussion

Valid and reliable clinical prognostic models are important for informed decision-making and may aid

patient-centered planning, ultimately benefiting both resource allocation and patient satisfaction [15,29]. In the setting of post-acute rehabilitation, meaningful patient-centered rehabilitation outcomes should be the outcome of prognosis to inform rehabilitation goals and the planning and preparation of life after discharge [1,16]. We identified three eligible articles, including the development and external validation of eight relevant prognostic models, all of which were deemed to be high risk of bias. Likewise, although discrimination appeared adequate, calibration was assessed or reported insufficiently, and optimism adjustment remained unclear.

4.1. Identified prognostic models

The FIS system was applied as a prognostic outcome in two prognostic models [50,51]; one of which may have been externally validated [51]. While the interpretation of FIM® total or domain scores may be considered relatively arbitrary to clinical staff, patients, and relatives, the FIS system provides FIM® scores as hierarchical activity profiles based on expected recovery [40]. Hence, these models are important to patients and are potentially a valuable prognostic tool in clinical practice [40,53], through provision of a tangible outcome (i.e., achievement of grade 5: requiring only supervision in most FIM® items). However, most patients similar to those under investigation appear to not reach grade 5 during rehabilitation [50,51]. This dichotomous classification is problematic and achievement of function for patients at lower grades cannot be estimated. Hence, a multinomial logistic regression predicting the likelihood of achieving any of the 7 grades may be worthwhile in the future [57,58]. Furthermore, model I was transformed into a nomogram [49] which may be easily implemented into daily clinical practice by yielding a visual impression of the relevance of predictors [59]. We did not include this nomogram [49] as a separate prognostic model, as it was unclear whether it was converted from the apparent or optimism-adjusted coefficients. There was some uncertainty whether a validation in a non-random split-sample (models II, IV, & VI) [50] may be considered an external geographical validation with some authors arguing for [60] and some against [[44,61] pp. 159 & 166] this. Here, we consider these models as internally validated for two reasons: a) the validation addresses reproducibility rather than transportability [62] as development and validation samples appear very similar and b) it was not investigated if the models needed updating or recalibration [63]. Model VI used the improvement on the motor FIM® as an outcome [50]. This model should be used extremely cautiously as the admission motor FIM® score was included as a predictor. Hence, a predictor defines the outcome (improvement = admission score - discharge score), causing mathematical coupling and voiding the independence assumption underlying regression analysis [64].

4.2. Methodological limitations of prognostic models

Most candidate predictors selected into the final prognostic models (Table 2) have previously been identified to be associated with functional independence in post-stroke rehabilitation [21]. Nevertheless, modelling choices along with sparse and non-transparent definitions of candidate predictors increases model uncertainty and reduces the interpretability and potential for further validation of the included models [44,65]. External validation addressing model transportability [62], that is, in a sample other than from the setting the model was developed and internally validated in, is required before any prognostic model should be applied in clinical practice [44,59−61]. Here, we identified the external validation of two prognostic models [51]. The intention of external validation is to test the prognostic performance of the original model in a different but similar sample than the original development sample [62]. This is one rationale behind prognostic models, that is, the application in new individuals. The two externally validated models (VII & VIII) identified in the present study did not appear to assess the performance of the original prognostic models (i.e., the linear predictor with its regression coefficients) in the validation cohort. Instead, the models appear to have been refitted in the validation cohort using unclear procedures; although both cross-validation and bootstrapping methods are mentioned. Although this approach may be necessary sometimes, for example, when case-mix or predictors effects are very different; it is usually not the preferred first step in an external validation [[59] p. 399ff.]; that is, the displayed models (VII & VIII) yield no information on the performance of the original model in a new sample. In addition, differences in case-mix or predictors effects were not investigated. Hence, these models should be seen as newly developed models with a priori defined predictors, in our opinion [[51,59] p. 382ff.]. It has been proposed that validation of existing prognostic models may be discouraged due to scarce and non-uniform reporting and high risk of bias, which was also observed in the prognostic models identified in the current study. Thus, developing a new prognostic model is often preferred and likely required [22,66]. Although there has been an increased advocacy for methodological prognosis frameworks in recent years [15,29,35], suboptimal statistical procedures with high risk of bias are still frequently observed in prognostic models proposed in healthcare [26,67−70].

4.3. Clinical implications

In all included prognostic models [49−51] calibration was sparsely discussed and discrimination was emphasized. Yet, most prognostic models should not be evaluated solely based on discrimination. Calibration was primarily assessed using the Hosmer−Lemeshow test, which is considered inadequate to effectively detect miscalibration [45,59,71]. Although calibration plots were presented for some models [50,51], their interpretation was hampered by a sparse

description of assessment procedures (e.g., applied grouping and smoother) and graphical presentation. Calibration-in-the-large or calibration slopes were not reported in the external validation [71]. This observation is similar to prognostic model studies early after TBI [26]. Calibration plays a substantial role in the performance of prognostic models by estimating the reliability of risk estimates [71,72]. Model calibration is substantially influenced by overfitting [71], which may be especially problematic here because of the following reasons: a) optimism (i.e., overfitting) was only assessed in one model [49], which did not elaborate on the estimation method, b) five included prognostic models [50,51] used stepwise forward selection methods, which are prone to overfitting and instability of selection [44,45,59], c) limited samples based on the effective sample size (i.e., EPP) [48,61], and d) non-repetition of selection procedures in the internal validation [60,73]. We have estimated global shrinkage factors for five models [50,51] based on published mathematical formula [48,56]. The global shrinkage estimates indicate some overfitting, likely related to testimation bias [59] caused by forward selection procedures. Yet, possibly due to the similarity between the split-samples [50], the overfitting does not appear severe. Importantly, poor calibration may affect the clinical utility of prognostic models and may lead to harmful decision making in some circumstances [72]. Although the models may perform well in the local context and may already be used to provide local evidence, prognostic models usually perform more poorly with actual external validation [66,70]. We disagree with the authors [51] on the conclusion that the newly developed model III outperformed model II (i.e., the original model) based on a crude comparison of the AUC estimates. The differences in AUC may be attributable to differences in selection procedures, case-mix or predictor effects which the authors did not formally assess [[59] p. 382ff.]. Some even advocate that prognostic models should be treated as health technology and be investigated as such before clinical application [61]. Hence, we deem a widespread application of the identified models unfeasible, based on the models uncertain calibration alone. We want to emphasize the need for uniform reporting and valid methodological approaches in prognosis research in postacute rehabilitation [74]. Future studies should draw upon recently published guidelines [29,33,44,45,59,75] to increase the transparency of prognosis research so it can be properly externally validated and safely implemented into routine clinical practice. Thus, aiding rehabilitation and benefiting both patients and healthcare providers.

4.4. Strengths and limitations of the present systematic review

We used a strict methodological approach based on acknowledged guidelines to plan and conduct the present review and deliberately applied a very broad search strategy as we expected non-uniform reporting and terminology [22]. In addition, we included a broad definition of the ABI population which, besides major subgroups, also included anoxic brain injuries; encephalitis and primary brain tumours. We applied the minimum inclusion criteria for valid prognostic model research [44,59], and filtered studies that had the potential for low risk of bias. Furthermore, we assessed the risk of bias in included prognostic models with the PROBAST tool [45], which was specifically developed for this purpose.

Reporting was very poor and the use of shared terminology (e.g., based on the PROGRESS framework or its precursors) [15,29,35] was rare. For example, one of the included studies was identified only through the ongoing updating of our personal collections. This sometimes made it difficult to ascertain what a study had actually done. In addition, our strict inclusion criteria, filtering only validated prognostic models for the FIM®, may be seen as a limitation as we might have excluded potentially relevant studies by design (i.e., missing validation). Yet, our intention was to quantify the performance of prognostic models satisfying minimum-criteria for prognostic model research (i.e., internal validation). A meta-analysis of the performance of non-validated prognostic models is not recommended due to high risk of bias and overestimation [31,76]; instead a meta-analysis of individual-patient data may be a better approach [77]. This approach however would require data sharing, which may be difficult in some jurisdictions and circumstances [78]. Although the best measure for prognosis of meaningful rehabilitation outcomes might be debatable [18,20,21], our decision to include only the FIM® as the eligible outcome is based on superior face validity, from our clinical perspective. Although both the Barthel Index and the FIM® are the most frequently used measures of disability/burden of care [21,79], only the FIM® incorporates a cognitive domain, which we deem of paramount importance to reflect functional independence in our study population. In contrast, the Barthel Index only contains motor items and is less frequently used in distinct post-acute rehabilitation therapy settings [20,21]. Instead of including other outcome measures we included seven possible FIM® dimensions as eligible outcomes, for example, the cognitive domain improvement or the FIS system [40]. Furthermore, we acknowledge that the definition of the time of prognostication and setting are grounded in a location (Western-European/Australian) understanding of post-acute ABI rehabilitation; another definition may be more meaningful in other healthcare contexts where access to these services may differ or rehabilitation is provided in other settings [3,36,80]. This definition lead to the exclusion of two otherwise qualified studies [53,54] which were deemed to provide mostly consultative rehabilitation in a subacute setting. Similarly, we excluded machine learning type prognostic models. Although machine learning type algorithms yield potential in prognostic model development in some circumstances (e.g., large datasets, real-time data

or competing risks), disadvantages in health research include reduced interpretability, a lacking framework for prognosis [75] and large required samples, ideally featuring high signal: noise ratios, seldom found in health research [81,82]. Reduced interpretability may be particularly problematic in clinical prognostic models which are intended for decision making and treatment planning in actual patients. Clinicians may not rely on a 'black box' type prognostic model, that is, those lacking transparency or clinical credibility, as they would be held accountable for potential mistreatment or suboptimal decision making [83]. In addition, it appears that current machine learning based prognostic models do not display superior performance over traditional regression based models [82,84].

5. Conclusion

The lack of prognostic models for functional independence in post-acute rehabilitation does not match the impact ABI has on individuals and healthcare spending. Although applying a broad search strategy, including diverse diagnoses of ABI, we identified only a few validated prognostic models, including only people with stroke. These models, although internally validated, were insufficiently reported, at high risk of bias, and should be considered local evidence not yet fit for widespread application. Future studies must acknowledge the importance of appropriate and robust model development, including internal validation, and conform to uniform reporting standards.

Acknowledgments

We would like to thank Henrik Sehested Laursen, Medical Librarian at Regional Hospital Central Jutland for his aid in drafting, testing and translating the electronic search strategy.

Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jclinepi.2023.02.009.

References

[1] United Nations. Convention on the Rights of Persons with Disabilities (CRPD). New York, NY: United Nations; 2008.
[2] 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. Lancet Neurol 2017;16:877−97.
[3] Buntin MB. Access to postacute rehabilitation. Arch Phys Med Rehabil 2007;88:1488−93.
[4] Rigby H, Gubitz G, Phillips S. A systematic review of caregiver burden following stroke. Int J Stroke 2009;4:285−92.
[5] Verdugo MA, Fernandez M, Gomez LE, Amor AM, Aza A. Predictive factors of quality of life in acquired brain injury. Int J Clin Health Psychol 2019;19:189−97.
[6] Kohnen R, Lavrijsen J, Smals O, Gerritsen D, Koopmans R. Prevalence and characteristics of neuropsychiatric symptoms, quality of life and psychotropics in people with acquired brain injury in long-term care. J Adv Nurs 2019;75:3715−25.
[7] Doser K, Norup A. Caregiver burden in Danish family members of patients with severe brain injury: the chronic phase. Brain Inj 2016;30:334−42.
[8] Jennum P, Iversen HK, Ibsen R, Kjellberg J. Cost of stroke: a controlled national study evaluating societal effects on patients and their partners. BMC Health Serv Res 2015;15:466.
[9] Olesen J, Gustavsson A, Svensson M, Wittchen H-U, Jönsson B, CDBE2010 Study Group, et al. The economic cost of brain disorders in Europe. Eur J Neurol 2012;19:155−62.
[10] Vestergaard SV, Rasmussen TB, Stallknecht S, Olsen J, Skipper N, Sørensen HT, et al. Occurrence, mortality and cost of brain disorders in Denmark: a population-based cohort study. BMJ Open 2020;10:e037564.
[11] Wafa HA, Wolfe CDA, Emmett E, Roth GA, Johnson CO, Wang Y. Burden of stroke in Europe. Stroke 2020;51:2418−27.
[12] Stabile M, Thomson S, Allin S, Boyle S, Busse R, Chevreul K, et al. Health care cost containment strategies used in four other high-income countries hold lessons for the United States. Health Aff (Millwood) 2013;32:643−52.
[13] Bonde M, Bossen C, Danholt P. Translating value-based health care: an experiment into healthcare governance and dialogical accountability. Sociol Health Illn 2018;40:1113−26.
[14] Porter ME, Lee TH. The strategy that will fix health care. Harv Bus Rev 2013;91:50−70.
[15] Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? BMJ 2009;338:b375.
[16] Wade D. Rehabilitation − a new approach. Part three: the implications of the theories. Clin Rehabil 2016;30:3−10.
[17] Linacre JM, Heinemann AW, Wright BD, Granger CV, Hamilton BB. The structure and stability of the functional independence measure. Arch Phys Med Rehabil 1994;75:127−32.
[18] Mahoney FI, Barthel DW. Functional evaluation: the Barthel index. Md State Med J 1965;14:61−5.
[19] Hobart JC, Lamping DL, Freeman JA, Langdon DW, McLellan DL, Greenwood RJ, et al. Evidence-based measurement. Which disability scale for neurologic rehabilitation? Neurology 2001;57:639−44.
[20] Turner-Stokes L, Williams H, Rose H, Harris S, Jackson D. Deriving a Barthel index from the northwick park dependency scale and the functional independence measure: are they equivalent? Clin Rehabil 2010;24:1121−6.
[21] Meyer MJ, Pereira S, McClure A, Teasell R, Thind A, Koval J, et al. A systematic review of studies reporting multivariable models to predict functional outcomes after post-stroke inpatient rehabilitation. Disabil Rehabil 2015;37:1316−23.
[22] Bouwmeester W, Zuithoff NP, Mallett S, Geerlings MI, Vergouwe Y, Steyerberg EW, et al. Reporting and methods in clinical prediction research: a systematic review. PLoS Med 2012;9:1−12.
[23] Counsell C, Dennis M. Systematic review of prognostic models in patients with acute stroke. Cerebrovasc Dis 2001;12:159−70.
[24] Perel P, Edwards P, Wentz R, Roberts I. Systematic review of prognostic models in traumatic brain injury. BMC Med Inform Decis Mak 2006;6:38.
[25] Jaja BNR, Cusimano MD, Etminan N, Hanggi D, Hasan D, Ilodigwe D, et al. Clinical prediction models for aneurysmal subarachnoid hemorrhage: a systematic review. Neurocrit Care 2013;18:143−53.
[26] Dijkland SA, Foks KA, Polinder S, Dippel DWJ, Maas AIR, Lingsma HF, et al. Prognosis in moderate and severe traumatic brain

injury: a systematic review of contemporary models and validation studies. J Neurotrauma 2019;37:1−13.
[27] Veerbeek JM, Kwakkel G, van Wegen EE, Ket JC, Heymans MW. Early prediction of outcome of activities of daily living after stroke: a systematic review. Stroke 2011;42:1482−8.
[28] Hakkennes SJ, Brock K, Hill KD. Selection for inpatient rehabilitation after acute stroke: a systematic review of the literature. Arch Phys Med Rehabil 2011;92:2057−70.
[29] 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 Med 2013;10:e1001381.
[30] Page MJ, Moher D, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. BMJ 2021;372:n160.
[31] Debray TP, Damen JA, Snell KI, Ensor J, Hooft L, Reitsma JB, et al. A guide to systematic review and meta-analysis of prediction model performance. BMJ 2017;356:i6460.
[32] Moons KG, de Groot JA, Bouwmeester W, Vergouwe Y, Mallett S, Altman DG, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. PLoS Med 2014;11:e1001744.
[33] Cochrane methods prognosis. Available at https://methods.cochrane.org/prognosis/. Accessed May 6, 2021.
[34] Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15:361−87.
[35] Altman DG, Royston P. What do we mean by validating a prognostic model? Stat Med 2000;19:453−73.
[36] Stineman MG, Xie D, Kurichi JE, Kwong PL, Vogel WB, Ripley DC, et al. Comprehensive versus consultative rehabilitation services post-acute stroke: outcomes differ. J Rehabil Res Dev 2014;51:1143−54.
[37] Choo SX, Stratford P, Richardson J, Bosch J, Pettit SM, Ansley BJ, et al. Comparison of the sensitivity to change of the functional independence measure with the assessment of motor and process skills within different rehabilitation populations. Disabil Rehabil 2018;40:3177−84.
[38] Pretz CR, Kean J, Heinemann AW, Kozlowski AJ, Bode RK, Gebhardt E. A multidimensional rasch analysis of the functional independence measure based on the national Institute on disability, independent living, and rehabilitation research traumatic brain injury model systems national database. J Neurotrauma 2016;33:1358−62.
[39] Stineman MG, Shea JA, Jette A, Tassoni CJ, Ottenbacher KJ, Fiedler R, et al. The Functional Independence Measure: tests of scaling assumptions, structure, and reliability across 20 diverse impairment categories. Arch Phys Med Rehabil 1996;77:1101−8.
[40] Stineman MG, Ross RN, Fiedler R, Granger CV, Maislin G. Functional independence staging: conceptual foundation, face validity, and empirical derivation. Arch Phys Med Rehabil 2003;84:29−37.
[41] Beam AL, Kohane IS. Big data and machine learning in health care. JAMA 2018;319:1317−8.
[42] Ingui BJ, Rogers MAM. Searching for clinical prediction rules in medline. J Am Med Inform Assoc 2001;8:391−7.
[43] Geersing G-J, Bouwmeester W, Zuithoff P, Spijker R, Leeflang M, Moons K. Search filters for finding prognostic and diagnostic prediction studies in medline to enhance systematic reviews. PLoS One 2012;7:e32844.
[44] Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162:W1−73.
[45] Moons KGM, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. Ann Intern Med 2019;170:W1−33.
[46] Pommerich UM, Stubbs PW, Fabricius JJ, Eggertsen PP, Nielsen JF. Multivariable prognostic models for functional independence at discharge from post-acute inpatient rehabilitation following acquired brain injury − protocol for a systematic review and meta-analysis. PROSPERO. 2021. Available at https://wwwcrdyorkacuk/prospero/display_recordphp?ID=CRD42021257098. Accessed June 30, 2021.
[47] Huguet A, Hayden JA, Stinson J, McGrath PJ, Chambers CT, Tougas ME, et al. Judging the quality of evidence in reviews of prognostic factor research: adapting the GRADE framework. Syst Rev 2013;2:1−12.
[48] Riley RD, Snell KI, Ensor J, Burke DL, Harrell FE, Jr, Moons KG, et al. Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. Stat Med 2019;38:1276−96.
[49] Scrutinio D, Guida P, Lanzillo B, Ferretti C, Loverre A, Montrone N, et al. Rehabilitation outcomes of patients with severe disability poststroke. Arch Phys Med Rehabil 2019;100:520−529.e3.
[50] Scrutinio D, Lanzillo B, Guida P, Mastropasqua F, Monitillo V, Pusineri M, et al. Development and validation of a predictive model for functional outcome after stroke rehabilitation. Stroke 2017;48:3308−15.
[51] García-Rudolph A, Bernabeu M, Cegarra B, Saurí J, Madai VI, Frey D, et al. Predictive models for independence after stroke rehabilitation: maugeri external validation and development of a new model. NeuroRehabilitation 2021;49:415−24.
[52] Inouye M. Predicting outcomes of patients in Japan after first acute stroke using a simple model. Am J Phys Med Rehabil 2001;80:645−9.
[53] Bates BE, Xie D, Kwong PL, Kurichi JE, Cowper Ripley D, Davenport C, et al. Development and validation of prognostic indices for recovery of physical functioning following stroke: part 1. PM R 2015;7:685−98.
[54] Bates BE, Xie D, Kwong PL, Kurichi JE, Ripley DC, Davenport C, et al. Development and validation of prognostic indices for recovery of physical functioning following stroke: part 2. PM R 2015;7:699−710.
[55] Matsugi A, Tani K, Mitani Y, Oku K, Tamaru Y, Nagano K. Revision of the predictive method improves precision in the prediction of stroke outcomes for patients admitted to rehabilitation hospitals. J Phys Ther Sci 2014;26:1429−31.
[56] Van Houwelingen JC, Le Cessie S. Predictive value of statistical models. Stat Med 1990;9:1303−25.
[57] Harrell FE, Jr. Regression modeling strategies with applications to linear models, logistic and ordinal regression, and survival analysis. Cham: Springer International Publishing; 2015.
[58] Edlinger M, van Smeden M, Alber HF, Wanitschek M, Van Calster B. Risk prediction models for discrete ordinal outcomes: calibration and the impact of the proportional odds assumption. Stat Med 2022;41:1334−60.
[59] Steyerberg EW. Clinical prediction models a practical approach to development, validation, and updating. 2nd ed. Cham: Springer Nature; 2019.
[60] Steyerberg EW, Harrell FE, Jr. Prediction models need appropriate internal, internal-external, and external validation. J Clin Epidemiol 2016;69:245−7.
[61] Riley RD, Moons KG, Debray TP, Snell K, Steyerberg EW, Altman DG, et al. Prognostic model research. In: Riley RD, van der Windt DA, Croft P, Moons KG, editors. Prognosis Research in Healthcare Concepts, Methods, and Impact. Oxford: Oxford University Press; 2019:139−87.
[62] 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.
[63] Van Calster B, McLernon DJ, van Smeden M, Wynants L, Steyerberg EW, Bossuyt P, et al. Calibration: the Achilles heel of predictive analytics. BMC Med 2019;17:230.
[64] Hawe RL, Scott SH, Dukelow SP. Taking proportional out of stroke recovery. Stroke 2019;50:204−11.

[65] Sauerbrei W, Perperoglou A, Schmid M, Abrahamowicz M, Becher H, Binder H, et al. State of the art in selection of variables and functional forms in multivariable analysis—outstanding issues. Diagn Progn Res 2020;4:3.
[66] Siontis GCM, Tzoulaki I, Castaldi PJ, Ioannidis JPA. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. J Clin Epidemiol 2015;68:25−34.
[67] Kaiser I, Diehl K, Heppt MV, Mathes S, Pfahlberg AB, Steeb T, et al. Reporting quality of studies developing and validating melanoma prediction models: an assessment based on the TRIPOD statement. Healthcare 2022;10:238.
[68] Haller MC, Aschauer C, Wallisch C, Leffondré K, van Smeden M, Oberbauer R, et al. Prediction models for living organ transplantation are poorly developed, reported, and validated: a systematic review. J Clin Epidemiol 2022;145:126−35.
[69] Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. BMJ 2020;369: m1328.
[70] Helmrich IRAR, Mikolić A, Kent DM, Lingsma HF, Wynants L, Steyerberg EW, et al. Does poor methodological quality of prediction modeling studies translate to poor model performance? An illustration in traumatic brain injury. Diagn Progn Res 2022;6:8.
[71] Van Calster B, Nieboer D, Vergouwe Y, De Cock B, Pencina MJ, Steyerberg EW. A calibration hierarchy for risk models was defined: from utopia to empirical data. J Clin Epidemiol 2016; 74:167−76.
[72] Van Calster B, Vickers AJ. Calibration of risk prediction models: impact on decision-analytic performance. Med Decis Making 2015; 35:162−9.
[73] Steyerberg EW, Uno H, Ioannidis JPA, van Calster B, Ukaegbu C, Dhingra T, et al. Poor performance of clinical prediction models: the harm of commonly applied methods. J Clin Epidemiol 2018;98: 133−43.
[74] Damen JAAG, Hooft L. The increasing need for systematic reviews of prognosis studies: strategies to facilitate review production and improve quality of primary research. Diagn Progn Res 2019;3:2.
[75] Collins GS, Dhiman P, Andaur Navarro CL, Ma J, Hooft L, Reitsma JB, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. BMJ Open 2021;11:e048008.
[76] Deeks JJ, Higgins JPT, Altman DG. On behalf of the Cochrane statistical methods Group. Chapter 10: analysing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions. Cochrane; 2022. www.training.cochrane.org/handbook. Accessed April 13, 2022.
[77] Debray TPA, de Jong VMT, Moons KGM, Riley RD. Evidence synthesis in prognosis research. Diagn Progn Res 2019;3:13.
[78] Peloquin D, DiMaio M, Bierer B, Barnes M. Disruptive and avoidable: GDPR challenges to secondary research uses of data. Eur J Hum Genet 2020;28:697−705.
[79] Sangha H, Lipson D, Foley N, Salter K, Bhogal S, Pohani G, et al. A comparison of the Barthel Index and the Functional Independence Measure as outcome measures in stroke rehabilitation: patterns of disability scale usage in clinical trials. Int J Rehabil Res 2005;28: 135−9.
[80] Langhammer B, Becker F, Sunnerhagen KS, Zhang T, Du X, Bushnik T, et al. Specialized stroke rehabilitation services in seven countries. Int J Stroke 2015;10:1236−46.
[81] van der Ploeg T, Austin PC, Steyerberg EW. Modern modelling techniques are data hungry: a simulation study for predicting dichotomous endpoints. BMC Med Res Methodol 2014;14:137.
[82] Steyerberg EW, van der Ploeg T, Van Calster B. Risk prediction with machine learning and regression methods. Biom J 2014;56: 601−6.
[83] Wyatt JC, Altman DG. Commentary: prognostic models: clinically useful or quickly forgotten? BMJ 1995;311:1539−41.
[84] Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol 2019;110:12−22.