frontiers
in Neurology

Thinking About the Future: A Review of Prognostic Scales Used in Acute Stroke

Bogna A. Drozdowska, Sarjit Singh and Terence J. Quinn*

Institute of Cardiovascular and Medical Sciences, College of Medical, Veterinary & Life Sciences, University of Glasgow, Glasgow, United Kingdom

Background: There are many prognostic scales that aim to predict functional outcome following acute stroke. Despite considerable research interest, these scales have had limited impact in routine clinical practice. This may be due to perceived problems with internal validity (quality of research), as well as external validity (generalizability of results). We set out to collate information on exemplar stroke prognosis scales, giving particular attention to the scale content, derivation, and validation.

Methods: We performed a focused literature search, designed to return high profile scales that use baseline clinical data to predict mortality or disability. We described prognostic utility and collated information on the content, development and validation of the tools. We critically appraised chosen scales based on the CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modeling Studies (CHARMS).

Results: We chose 10 primary scales that met our inclusion criteria, six of which had revised/modified versions. Most primary scales used 5 input variables (range: 4–13), with substantial overlap in the variables included. All scales included age, eight included a measure of stroke severity, while five scales incorporated pre-stroke level of function (often using modified Rankin Scale), comorbidities and classification of stroke type. Through our critical appraisal, we found issues relating to excluding patients with missing data from derivation studies, and basing the selection of model variable on significance in univariable analysis (in both cases noted for six studies). We identified separate external validation studies for all primary scales but one, with a total of 60 validation studies.

Conclusions: Most acute stroke prognosis scales use similar variables to predict long-term outcomes and most have reasonable prognostic accuracy. While not all published scales followed best practice in development, most have been subsequently validated. Lack of clinical uptake may relate more to practical application of scales rather than validity. Impact studies are now necessary to investigate clinical usefulness of existing scales.

Keywords: stroke, functional outcome, prognostic scale, risk score, validation

INTRODUCTION

Outcomes following a stroke event can range from full recovery, through varying degrees of disability to death. Given the subsequent need for intervention planning, resource use, and lifestyle adjustments, predicting outcome following stroke is of key interest and importance to patients, their families, clinicians, and hospital administrators. Various tools exist to assist in estimating stroke-related prognosis. For example, the ABCD2 score uses clinical features to predict risk of stroke following transient ischemic attack (TIA) (1). Although there are criticisms of ABCD2, it is widely used and included in stroke guidelines (2).

Scales for predicting acute stroke outcomes from baseline features are also described in the scientific literature (3–5). Often prognosis scales report mortality; however, given the disabling nature of stroke, scales predicting death and/or longer-term disability may be more useful in the stroke setting (6). However, these prognostic scales have had limited clinical traction and have not been incorporated into routine clinical practice (3). There are many plausible reasons why these scales have not been adopted by the stroke community (6). In an acute setting, scales may be perceived as being too complex to use or may require information that is not routinely available (for example, sophisticated neuroimaging) (3). Clinicians may moreover be concerned that scales are inherently too generic, and may not provide insight over what the clinician can conclude based on individual patient factors and clinical gestalt (7).

For many scales, clinicians may simply not be convinced of their utility or the rigor of the underpinning science. These points can be addressed by describing the validity of the scales. Issues with validity could relate to the methodological quality of the initial derivation of the scale (internal validity) or the generalizability of a scale to a real-world population (external validity). Robust evidence of validity requires assessment of the scale in cohorts independent of the population used to derive that scale (8). However, In some areas of stroke practice, for example rehabilitation, it has been demonstrated that independent validation studies are lacking for many scales (5).

Collating evidence around the quality of the research that led to development of prognostic scales and also the results of subsequent validation work could be useful for various stakeholders. For clinicians it may convince of the utility, or lack of utility, of certain tools; for researchers it may point to common methodological limitations that need to be addressed in future work and for policy developers, if a certain tool has a more compelling evidence base than others, then this scale may be preferred in guidelines.

Previous reviews have reported that many stroke prognosis scales have similar properties such as discrimination and calibration. These reviews also highlight the limited evidence for external validity of many commonly used stroke scales (9, 10). Distinguishing an optimal prognostic tool may not be possible based on psychometric properties alone and factors such as feasibility and acceptability in the real world setting need to be considered.

We sought to collate and appraise a selection of exemplar published stroke scales, designed for use in acute care settings. We used these as a platform to discuss methodological quality of prognostic scale development, while also considering potential barriers or facilitators to implementation of the scales in clinical practice.

METHODS

We performed a focused review of the literature to find scales predicting post-stroke mortality and/or function. Our approach followed that used in a recent comparative efficacy review of stroke scales (9). Rather than assess every tool that has ever been used to make outcome predictions in stroke, we were interested in examples of high profile prognostic scales. Although our intention was not a comprehensive search, we followed, where relevant, Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidance for designing and reporting our study (11). For consistency in use of terminology, we have referred to the prediction models as "scales," and the calculated outputs of models as "scores."

Inclusion/Exclusion Criteria

We defined a scale as any tool that uses more than two determinants to estimate the probability of a certain outcome. We focused on scales with predominant clinical input variables that can be applied without specialist resources or tests and to this end we excluded scales that had more than two neuroimaging input variables. We limited to ischemic or all cause (undifferentiated) stroke scales, recognizing the differing natural progression of ischemic and hemorrhagic strokes.

Search Strategy

Our focus was on scales that are well known in the stroke field and so we adapted our search using an approach that has been used in other focused stroke studies (12, 13). We limited our search to 11 high profile, international journals, chosen based on relevance to stroke and clinical impact, covering fields of stroke, neurology, internal medicine, and geriatric medicine (a full list of journals and the search strategy are included in **Supplementary Materials**).

Searches were from inception to May 2018. Once we had selected chosen scales we used PUBMED and Google Scholar electronic search engines to find the initial development paper and any potential validation papers. A single researcher (SS) performed the search and screened the results We assessed internal validity of the search results by screening title lists twice (October 2015 and May 2018).

Data Extraction and Critical Appraisal

Two researchers (BD, SS) extracted data from selected studies, using a pre-specified proforma. This included information on: data source, study sample characteristics, predictor, and outcome variables, procedures involved in model derivation, methods of validation, measures of performance and presentation of

results. Extracted data were comprehensively reviewed to inform critical appraisal, a process in which all authors (BD, SS, TQ) were involved.

The methodological assessment of prognostic scales is an evolving landscape. Although there is no consensus preferred approach to this, there are certain features common to most tools that purport to assess validity of prognosis research. We based our assessment on recommendations from the Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modeling Studies (CHARMS) checklist (14). Discrepancies in assessment between researchers were discussed and resolved through consensus.

Data Used for Scale Development

We assessed the representativeness of the sample from which information was collected. Generalizability of a scale to a broader patient population may be compromised when recruitment takes place in a highly specific context or is limited to a relatively homogeneous group; when multiple inclusion and exclusion criteria are applied; and finally, when patients with missing data are excluded from the study (complete-case analysis). The latter presents itself as an issue, as it is uncommon for variable values to be missing completely at random. Often this is related to other predictors, the outcome, or even the value of that variable itself (15). Therefore, patients with missing data are likely to form a selective rather than random subsample of the initial baseline cohort, and may substantially differ from those included in the analysis (14, 16).

Scale Variables

For predictor and outcome variables, a particular concern was whether they are precisely defined and measured in a way that can be reproduced across different centers. It is recommended that continuous data (e.g., age) are not categorized when introduced to a model as a predictor (17, 18). Doing so is associated with loss of information and power, and increases the risk of generating inaccurate estimates and residual confounding. Finally, bias may arise from lack of blinding to predictors when assessing an outcome, or blinding to the outcome where predictors are assessed retrospectively.

Scale Development Process

In this context, we assessed study sample size against the number of candidate predictors being tested. For logistic regression procedures, we considered a minimum of ten events (number of patients with the less frequent outcome) per variable to be sufficient (19, 20). Evaluating the selection process of predictors for inclusion in scales presents a challenge, as there is no agreed approach (21). There are however certain practices that are consistently stated to increase risk of bias. One is selecting predictors for inclusion in multivariable analysis based on significance in univariable analyses (22). This approach may lead to exclusion of predictors that could be associated with the outcome after adjusting for the effects of other factors. A data-driven approach to variable selection may lead to model overfitting (23) and forward selection techniques should be avoided in multivariable

modeling (24). Either a full model approach (all candidate variables included in the model) or backwards elimination (beginning with all candidate predictors, removing those that do not satisfy a pre-specified statistical criterion) is preferable (25).

Assessment of Scale Performance

We distinguished three levels of validation: apparent, internal and external (26–28). In apparent validation, predictive ability is assessed in the development set itself and may give overoptimistic performance estimates. With internal validation two approaches are described, split-sample and cross-validation. These involve randomly splitting the baseline sample into development and assessment sets. In the split-sample technique, the population is divided once, in cross-validation, the process of sample division is repeated for consecutive fractions of subjects, thus allowing for each participant to be included in the validation set once. Here, a larger part of the baseline sample can be used for model derivation, avoiding the considerable loss of power associated with split sample approaches (14). The most efficient method of internal validation is considered to be bootstrapping, where samples are drawn with replacement from the original dataset, replicating sampling from an underlying population (26). The generated sample is of the same size as the original dataset.

Importantly, even with use of internal validation techniques, assessing a scale's performance in the development cohort is considered insufficient to confirm its value and general applicability (27, 29), In view of this, we prioritized findings from external validation studies. External data can differ from derivation data in terms of when and where it was collected, as well as by research group. Typically, an external dataset is comparable to the original, however in some studies a model is intentionally tested in a population characterized by different clinical features.

Reviewing study results on predictive performance, we focused on measures of discrimination and calibration, as these properties are necessary (although not sufficient) to ensure clinical usefulness of a prognostic scale (14, 27). Discrimination relates to the ability of a model to accurately distinguish between those who develop a certain outcome and those who do not, and is commonly expressed as the area under the receiver operating characteristic curve (AUROC) (30). To aid interpretation of results, we applied to following AUROC cut-off values: 1.00—perfect discrimination; 0.90 to 0.99—excellent; 0.80 to 0.89—good; 0.70 to 0.79—fair; 0.51 to 0.69—poor; 0.50—of no value, equivalent to chance (31). Calibration refers to the level of agreement between observed and predicted outcome probabilities, with assessment preferably based on inspection of calibration plots/curves (32). Graphical evaluation can be accompanied by reporting results of the Hosmer-Lemeshow test, assessing whether there is a significant difference between observed and predicted outcomes. The test has however limited power for detection of poor calibration, is oversensitive in large samples, and does not allow to determine the direction of miscalibration (33).

RESULTS

Overview of Scale Content and Quality

Our search returned 3817 results. We found 10 primary scales that met our inclusion criteria, six of which also had modified versions published (**Tables 1**, **2**). Scales used from four to thirteen input variables, with a mode of five. There was considerable overlap in the predictors used, including variables relating to demographics, past medical history and the acute stroke event (**Table 3**). The most commonly incorporated predictors were: age (all ten scales), a measure of stroke severity (eight); pre-stroke function, comorbidities and stroke subtype (each present in five scales). Individual scale content, with scoring, is presented in **Table 4**.

For seven scales the outcome of interest was a specified range of scores on the modified Rankin Scale (mRS) (34–36). The scale is a measure of functional outcome following stroke, ranging from no symptoms (a score of zero) through increasing levels of disability, to death (a score of six). Five scales focused on mortality, while one scale aimed to predict recurrent stroke and another length of hospital stay. Four scales predicted more than one outcome.

Our critical appraisal of scales' development process identified potential sources of bias in each study, as well as issues related to incompleteness of reporting for methods and results. Most common limitations were around handling missing data and model development. In relation to the former, in two studies it was not clearly stated how missing data was handled. Six studies used complete-case analysis, and the remaining two excluded participants from analyses involving the particular variables they had no data for. For model development, six studies selected variables for multivariable modeling based on the univariable significance (**Table 5**).

We present an overview of each scale, focusing on scale content, development, validation, and where applicable any modification to the scale. We summarize our critical appraisal of derivation studies and discuss potential issues around implementation of the scales in routine clinical practice.

Acute Stroke Registry and Analysis of Lausanne (ASTRAL)

Scale Content and Development

The ASTRAL scale uses six input variables to predict unfavorable functional outcome at 3 months (mRS>2): age, stroke severity according to the National Institutes of Health Stroke Scale (NIHSS) (37), time from symptom onset to admission, range of visual fields, acute glucose, and level of consciousness (38). Based on these variables, an integer score is assigned, from zero with no upper limit. Higher scores are associated with a greater probability of an unfavorable outcome. Through a logistic regression procedure, the scale was developed in a sample of 1,633 ischemic stroke patients from the Acute Stroke Registry and Analysis of Lausanne (39).

Scale Validation and Updating

Using a 2-fold cross-validation technique for internal validation, the scale was found to have good discriminatory power,

AUROC:0.85 for prediction of mRS>2 at 3 months. The derivation paper further described external validation of the scale in two independent cohorts from Athens and Vienna (40, 41), reporting AUROC values of 0.94 and 0.77, respectively. Calibration was assessed in all three cohorts based on Hosmer-Lemeshow test and inspection of calibration plots, indicating a good fit with the data.

The ASTRAL scale has been subsequently externally validated by seven studies, with six assessing predictive value based on AUROC estimates (42–48). Within these, ASTRAL was found to have fair to good discriminatory power, with the exception of one study, involving a Brazilian cohort (AUROC 0.67) (44). These external validation studies used differing time points for outcome assessment (up to 5 years post-stroke) and differing outcomes, including mortality and symptomatic intracerebral hemorrhage (sICH).

Critical Appraisal and Clinical Application

In the ASTRAL derivation study we found potential sources of bias relating to participant selection, namely excluding all patients with pre-stroke dependency and any missing data. In addition, treatment effects were not accounted for. We also noted that some issues relevant to scale development were unclear: whether any method of blinding was used, the number of candidate predictors (which allows to estimate whether the sample size was sufficient), and finally whether there were any significant baseline differences between the derivation and validation cohorts.

Despite these concerns, evidence from validation studies suggests that the predictive performance of ASTRAL is sufficient for the clinical setting. The scale was designed with the acute context in mind, and does not require sophisticated diagnostic tests. Nonetheless, in some cases, estimating onset to admission time may not be possible. Where all necessary information is accessible, the ASTRAL offers an easily-calculable score, with use aided by color-coded graphs to assign a percentage probability of unfavorable outcome based on clinical features. There is also a score calculator available online (49).

Dense Artery, mRS, Age, Glucose, Onset-to-Treatment, and NIHSS (DRAGON)

Scale Content and Development

The DRAGON scale incorporates the six variables in its acronym, as well as early infarct signs on computed tomography (CT). It was developed to predict functional outcome at 3 months in stroke patients treated with intravenous tissue plasminogen activator (IV-tPA) (50). The outcome was trichotomized according to mRS scores, where mRS 0–2 was defined as "good outcome," mRS 3–4 as "poor outcome" and mRS 5–6 as "miserable outcome." Scale scores range from one to ten, with higher values associated with poorer outcomes. The scale was derived in a single-center Finish cohort of 1,319 ischemic stroke patients, using a logistic regression procedure.

TABLE 1 | Derivation study characteristics.

Scale	Derivation dataset	Location	Recruitment type	Enrolment dates	Stroke type	Exclusion criteria of note
ASTRAL	Acute Stroke Registry and Analysis of Lausanne	Switzerland	Single-center	Jan 2003 to Jul 2010	Ischemic	Pre-stroke mRS > 2
DRAGON	Bespoke cohort	Finland	Single-center	1995 to Sep 2010	Ischemic	Basilar artery occlusions
FSV	Stroke Outcome Study	Canada	Single-center	2001 to 2002	Ischemic, hemorrhagic	
iSCORE	Registry of the Canadian Stroke Network	Canada	Multi-center	Jul 2003 to Jun 2008	Ischemic	
PLAN	Registry of the Canadian Stroke Network	Canada	Multi-center	Jul 2003 to Mar 2008	Ischemic	Patients receiving IV-tPA
SNARL	Endovascular registry	Unclear	Multi-center	Sep 2009 to Jul 2011	Ischemic	
SOAR	Anglia Stroke and Heart Clinical Network Database	United Kingdom	Multi-center	1997 to 2010	Ischemic, hemorrhagic	
SPI	Carotid ultrasound register	United States	Single-center	Jan 1984 to Feb 1987	Ischemic	Artificial heart valves, previous cerebrovascular event
S-TPI	NINDS 1 & 2, ATLANTIS A & B, ECASS 2 cohorts	International	Multi-center	Jan 1991 to Jul 1998*	Ischemic	Multiple, related to primary randomized control trials; minor strokes
THRIVE	MERCI, Multi MERCI cohorts	International	Multi-center	May 2001 to Dec 2003; Jan 2004 to Jul 2006	Ischemic	Multiple, related to primary single-arm trials.

IV-tPA, intravenous tissue plasminogen activator; mRS, modified Rankin Scale.
*Combined recruitment period for all trials.

Scale Validation and Updating

On internal validation, using 1,000 bootstrap replications, DRAGON was found to have an AUROC of 0.84 (95%CI: 0.80–0.87) for prediction of miserable outcome. A comparable AUROC value was obtained through external validation, performed by the authors in a cohort of 330 Swiss patients: 0.80 (95%CI: 0.74–0.86). Calibration was not assessed. DRAGON has undergone subsequent external validation in ten studies (42, 43, 46, 51–57), all of which concluded the scale performs well, and (where assessed) had fair to good discriminatory power. In majority of cases, the scale was used in a similar context and for the same purpose as in the derivation study. However, one study assessed prediction of sICH (42).

Recognizing the increasing use of magnetic resonance imaging (MRI), the original DRAGON scale was adapted to include MRI based variables (58). Namely, with all clinical variables remaining unchanged, proximal middle cerebral artery occlusion on MR angiography replaced hyperdense artery sign, and the diffusion-weighted imaging Alberta Stroke Program Early Computed Tomography Score (DWI ASPECTS) replaced CT early infarct signs (59, 60). The scale was derived in a French cohort of 228 patients treated with IV-tPA. Internal validation was performed using a bootstrapping method. For prediction of 3-month mRS>2, MRI-DRAGON was found to have an AUROC of 0.83 (95%CI: 0.78–0.88). The scale was externally validated in one subsequent study, where reported AUROC values for prediction of poor and miserable outcome were 0.81 (95%CI: 0.75–0.87) and 0.89 (95%CI: 0.84–0.95), respectively (61).

Critical Appraisal and Clinical Application

We identified issues in the DRAGON derivation study. All continuous candidate predictors were categorized. A complete-case analysis approach was employed and discriminatory power was only estimated for prediction of miserable outcome, while calibration was not assessed at all. Moreover, it seemed unclear whether any blinding method was applied for assigning mRS scores, and there were no description of the multivariable method for selection of final predictors.

In the context of clinical practice, DRAGON score should be easy to calculate [online score calculator available (62)]. Again, estimating symptom onset-to-treatment time may not be possible in some cases. There is potential for misinterpretation of early infarct and hyperdense cerebral artery signs (63–65). From this point of view, MRI-DRAGON appears a valuable alternative. MRI has been found to be a more sensitive method for ischemia detection than CT, and use of a semi-quantitative assessment of lesions is likely to ensure higher reproducibility (66). Importantly, based on results of validation studies, both versions of the scale seem to have satisfactory predictive ability, although evidence on performance of MRI-DRAGON is still limited.

Five Simple Variables (FSV)
Scale Content and Development

The FSV scale incorporates two models for predicting functional outcome at 6 months post-stroke (67–69). One is used for good (mRS<3) or excellent (mRS<2) outcomes, and one for prediction of a devastating outcome (mRS>4; FSV_{DEV}). The two models share four input variables: age, pre-stroke functional status (Oxford Handicap Score) (70), ability to lift both arms off the bed, and normal verbal response on the Glasgow Coma Scale (71). The first model additionally includes ability to walk unaided, while the FSV_{DEV} incorporates stroke subtype. Prediction scores created based on the models range from−5 to 5 for the positive outcomes, and 0 to 15 for the devastating outcome. In both cases, a higher score is associated with a

TABLE 2 | Derivation study participants, outcomes and scale discriminatory power.

Scale	Baseline sample size	Total sample used for model derivation (% of baseline sample)	Outcome	Timepoint of outcome assessment	Number of patients with outcome (%)	AUROC (95% CI)* for apparent/internal validation
ASTRAL	1,967	1,645 (84%)	mRS 3–6	3 months	559 (34%)	0.85
DRAGON	1,529	1,319 (86%)	mRS 0–2	3 months	798 (60%)	Not reported
			mRS 3–4	3 months	339 (26%)	Not reported
			mRS 5–6	3 months	182 (14%)	0.84 (0.80 – 0.87)
FSV	598	538 (90%)	mRS 0–1	6 months	Not reported	0.86
			mRS 0–2	6 months		0.89
			mRS 5–6	6 months		0.87
iSCORE	Not available	8,223 (n/a)	Death	1 month	1,004 (12.2%)	0.85**
			Death	1 year	1,853 (22.5%)	0.82**
PLAN	12,576	4,943 (39%)	Death	1 month	569 (11.5%)	0.85 (0.84–0.87)
			Death	1 year	1,088 (22.0%)	0.82 (0.81–0.84)
			mRS 5–6	Discharge	735 (14.9%)	0.89 (0.87–0.90)
SNARL	556	511 (92%)	mRS 0–2	3 months	186 (36.4%)	0.79 (0.75–0.83)
SOAR	Not available	12,355 (n/a)	Death	7 days	Not reported	0.79 (0.78–0.80)
			Death	Inpatient	Not reported	0.79 (0.78–0.80)
			LOS < 8 days	n/a	n/a	0.61 (0.60–0.62)
SPI	352	142 (40%)	Stroke or death	2 years	38 (27%)	Not reported
S-TPI	2,184	2,131 (98%)	mRS 0–1	3 months	773 (36%)	0.79**
			mRS 5–6	3 months	464 (22%)	0.78**
THRIVE	305	Not reported	mRS 0–2	3 months	94 (n/a)	0.71**
			Death	3 months	111 (n/a)	Not reported

mRS, modified Rankin Scale; LOS, length of hospital stay.
*Where reported in derivation study.
**AUROC for model, not derived risk score.

TABLE 3 | Categories of input variables included in post-stroke prognostic scales.

Scale	Age	Sex	Pre-stroke functional status	Comorbidity	Time from symptom onset	Acute physiology	Stroke severity	Stroke classification	Imaging findings	Treatment
Astral	■				■	■	■			
Dragon	■		■		■	■	■		■	
FSV	■		■			■	■			
iScore	■	■	■	■		■	■	■		
PLAN	■		■	■		■	■			
SNARL	■		■	■		■	■		■	
SOAR	■		■				■	■		
SPI	■			■		■	■			
S-TPI	■	■	■				■		■	■
THRIVE	■			■			■			

Green fill denotes inclusion of a variable from the given category in a prognostic scale.

greater likelihood of having the outcome of interest. Both FSV models were derived in a single-center Canadian cohort of 538 stroke patients.

Scale Validation and Updating
Internal validation of the prediction scores, using 500 bootstrap replications, indicated good discriminatory power, with AUROC values of 0.88, 0.87, and 0.86 for good, excellent and devastating outcomes, respectively. Similar results were reported for initial external validation, conducted in a sample of patients from the Oxfordshire Community Stroke Project (OCSP), with AUROC values ranging from 0.86 to 0.89. Calibration was assessed only in the derivation sample for prediction of good outcome, and, based plotted calibration curves, concluded to be good (67).

FSV scores have been externally validated in one study (72), reporting good discriminatory power for prediction of good and devastating outcomes at 6 months in a Scottish stroke cohort. The use of five variables for predicting post-stroke functional outcome was also assessed in a cohort combining six European populations. However, here a similar scale was being

TABLE 4 | Scale variables and scoring systems.

Scale	Variable	Level/category	Score
ASTRAL	Acute glucose	<3.7 or >7.3 mmol/L	1
	Age	Per every 5 years	1
	Any stroke-related visual field defect	Yes	2
	Level of consciousness	Decreased	3
	Symptom onset to treatment time	>3 hours	2
	Stroke severity (NIHSS)	Per every point	1
DRAGON	Hyperdense cerebral artery or early infarct signs on CT scan	None	0
		Either	1
		Both	2
	Pre-stroke functioning (mRS)	≤1	0
		>1	1
	Age	<65 years	0
		65–79 years	1
		>79 years	2
	Acute glucose	≤8 mmol/L	0
		>8 mmol/L	1
	Symptom onset to treatment time	≤90 min	0
		>90 min	1
	Stroke severity (NIHSS)	0–4	0
		5–9	1
		10–15	2
		>15	3
FSV	Able to lift both arms	Yes	2
	Able to walk unaided	Yes	1
	Age	<80 years	1
	Verbal GCS	Normal	1
	Pre-stroke functioning (mRS)	Per every level	−1
FSV$_{DEV}$	Able to lift both arms	No	2
	Age	≥80 years	3
	Verbal GCS	Abnormal	3
	Pre-stroke functioning (mRS)	Per every level	1
	Stroke subtype classification (OCSP)	TACS	2
iSCORE; 1-month outcome	Age	Per every year	1
	Sex	Male	10
	Stroke severity (CNS)	0	105
		≤4	65
		5-7	40
		>8	0
	Stroke subtype (TOAST)	Lacunar	0
		Non-lacunar	30
		Undetermined	35
	History of atrial fibrillation	Yes	10
	History of congestive heart failure	Yes	10
	Cancer	Yes	10
	Renal dialysis	Yes	35
	Pre-stroke functioning	Dependent	15

(Continued)

TABLE 4 | Continued

Scale	Variable	Level/category	Score
	Acute glucose	≥7.5 mmol/L	15
iSCORE; 1-year outcome	Age	Per every year	1
	Sex	Male	5
	Stroke severity (CNS)	0	70
		≤4	40
		5-7	25
		>8	0
	Stroke subtype (TOAST)	Lacunar	0
		Non-lacunar	15
		Undetermined	20
	History of atrial fibrillation	Yes	5
	History of congestive heart failure	Yes	10
	Previous myocardial infarction	Yes	5
	Current smoker	Yes	5
	Cancer	Yes	15
	Renal dialysis	Yes	40
	Pre-stroke functioning	Dependent	20
	Acute glucose	≥7.5 mmol/L	10
PLAN	Pre-stroke functioning	Dependent	1.5
	Cancer	Yes	1.5
	History of congestive heart failure	Yes	1.0
	History of atrial fibrillation	Yes	1.0
	Level of consciousness	Reduced	5.0
	Age	Per decade	1.0
	Arm weakness	Significant/total	2.0
	Leg weakness	Significant/total	2.0
	Neglect or aphasia	Either/both	1.0
SNARL	Symptomatic hemorrhage	No	2
	Stroke severity (NIHSS)	>20	0
		10–20	1
		<10	3
	Age	>80 years	0
		60–79 years	1
		<60 years	2
	Reperfusion (TICI)	≥2b	3
	Location of occlusion	M2 or distal	1
SOAR	Age	≤65	0
		66–85	1
SNARL		>85	2
	Stroke type	Hemorrhagic	1
	Stroke subtype classification (OCSP)	LACS	0
		PACS	0
		POCS	1
		TACS	2
	Pre-stroke functioning (mRS)	≤2	0
		3–4	1
		5	2
SPI	Age	>65	3
	Diabetes mellitus	Yes	3

(Continued)

TABLE 4 | Continued

Scale	Variable	Level/category	Score
	Acute severe hypertension	Yes	2
	Type of cerebrovascular event	TIA	0
		Stroke	2
	History of coronary heart disease	Yes	1
S-TPI; Good outcome	IV-tPA treatment	Yes	N/a*
	Age	Per every year	N/a*
	Acute systolic blood pressure	Per every 1 mmHg	N/a*
	Diabetes	Yes	N/a*
	Sex	Male	N/a*
	Stroke severity (NIHSS)	Per every point	N/a*
	Previous stroke	Yes	N/a*
	Symptom onset to treatment time	Per every minute	N/a*
	Treatment × systolic blood pressure		N/a*
	Treatment × sex		N/a*
	Treatment × previous stroke		N/a*
	Treatment × symptom onset to treatment time		N/a*
	Age × stroke severity		N/a*
S-TPI; Poor outcome	Age	Per every year	N/a*
	Stroke severity (NIHSS)	Per every point	N/a*
	Acute glucose	Per every mmol/L	N/a*
	ASPECTS score**	Per every point	N/a*
THRIVE	Age	≤59	0
		60–79	1
		≥80	2
	Stroke severity (NIHSS)	≤10	0
		11–20	2
		≥80	4
	Diabetes mellitus	Yes	1
	History of hypertension	Yes	1
	History of atrial fibrillation	Yes	1

ASPECTS, Alberta Stroke Program Early Computed Tomography Score; CNS, Canadian Neurological Scale; CT, computed tomography; GCS, Glasgow Coma Scale; IV-tPA, intravenous tissue plasminogen activator; LACS, lacunar stroke syndrome; mRS, modified Rankin Scale; NIHSS, National Institutes of Health Stroke Scale; OCSP, Oxfordshire Community Stroke Project; PACS, partial anterior circulation stroke syndrome; POCS, posterior circulation stroke syndrome; TACS, total anterior circulation stroke syndrome; TIA, transient ischemic attack; TICI, Thrombolysis In Cerebral Ischemia; TOAST, Trial of ORG 10172 in Acute Stroke Treatment.
*Risk score was not developed.
**Optional variable.

independently developed rather than the FSV being externally validated (73). The described model included the same variables, although a different measure was used for estimating pre-stroke functional status [Barthel index (74)]. The authors reported good discriminatory power on both internal and external validation.

Critical Appraisal and Clinical Application
To assess FSV derivation, we reviewed three publications and identified potential sources of bias. The sample size was insufficient for the number of tested candidate predictors.

Although a complete-case analysis method was not applied, with no data imputation, participants with missing data were excluded from particular analyses. Blinding was unclear. Input variables for multivariable analyses were selected based on univariate significance. In the paper where models for excellent and devastating outcomes were developed, calibration was not assessed, while in the remaining two, the procedure was mentioned but no calibration plots were presented. In relation to study results, differences in baseline characteristics between derivation and validation datasets were not assessed, and a data-driven approach was applied when selecting cut-off scores for outcome prediction (75).

In clinical practice, a significant advantage of FSV is the use of easily accessible and often routinely collected information. Moreover, for patients and their families, the differentiation between recovering to a level of functional independence with and without disability can be of particular value. It is unlikely however for this useful concept to be transferred into practice, as the same FSV cut-off score was chosen for both outcomes, the difference lying in prognostic accuracy for prediction of each. Finally, although reports on FSV performance are encouraging, further external validation studies are necessary before it can be considered for use in a clinical setting.

iScore
Scale Content and Development
The iScore was developed using a logistic regression procedure to predict death at two timepoints. The derivation study included 12,262 ischemic stroke patients from the Registry of the Canadian Stroke Network (76). For outcome prediction at 3 months, an integer score (from zero, with no defined upper limit) is calculated based on: age, sex, stroke severity assessed with the Canadian Neurological Scale (77), stroke subtype according to the Trial of ORG 10172 in Acute Stroke Treatment (TOAST) (78), acute glucose, history of atrial fibrillation, congestive heart failure, cancer, kidney disease, and preadmission dependency. For predicting one-year mortality, previous myocardial infarction and smoking status are added. Higher scores associated with greater mortality.

Scale Validation and Updating
In the derivation study, a split-sample validation method was chosen, with 8223 patients assigned to the development set (AUROC: 0.85 and 0.82 for 30-day and 1-year mortality, respectively) and 4039 in the internal validation (AUROC: 0.85 and 0.84). External validation used data from 3270 patients from the Ontario Stroke Audit. AUROC: 0.79 and 0.78, for 30-day and 1-year mortality, respectively.

The scale has been further externally validated in 15 studies (48, 54, 79–91). The iScore has been applied not only to predict mortality, but also poor functional outcome, institutionalization, clinical response, hemorrhagic transformations following thrombolytic therapy, and healthcare costs. All studies concluded that iScore is useful, predicting outcomes of interest with sufficient accuracy. Where AUROC values were estimated, they were fair to good, apart from one study where AUROC was 0.68 for 30-day mortality or disability at discharge (79).

TABLE 5 | Assessment of risk of bias in scale derivation studies.

Scale	Data source	Participants	Outcome	Predictors	Sample size	Missing data	Model development	Model performance	Model evaluation	Results
Astral	green	red	yellow	green	green	red	green	green	green	yellow
Dragon	green	green	green	red	green	red	yellow	red	green	green
FSV	green	green	green	red	yellow	green	red	green	green	red
iScore	green	yellow	green	yellow	green	red	red	green	yellow	red
PLAN	green	green	green	red	green	yellow	red	yellow	yellow	green
SNARL	green	yellow	yellow	green	green	red	red	green	green	green
SOAR	green	yellow	green	green	green	red	red	green	green	red
SPI	green	yellow	yellow	red	red	yellow	red	red	yellow	green
S-TPI	yellow	red	green	green	green	yellow	green	green	yellow	green
THRIVE	green	red	green	green	green	yellow	red	green	green	green

Color code: green, low risk of bias; yellow, unclear or medium risk of bias; red, high risk of bias.

Recognizing the difficulty of etiological classification (92), a revised iScore (iScore-r) was developed, replacing TOAST with OCSP (93). The revised scale was validated in a Taiwanese cohort of 3,504 ischemic stroke patients, for prediction of poor functional outcome (mRS>2) at discharge and at 3-months. Assessment of discriminatory power in an external cohort of iScore and iScore-r indicated comparable performance of the scales. AUROC of 0.78 and 0.77 for discharge outcome, and AUROC of 0.81 and 0.80 for 3-month outcome, with lower values reported for iScore-r.

Critical Appraisal and Clinical Application

We identified limitations in the iScore derivation. A complete-case analysis approach was applied. Variables were selected based on univariable significance. Administration of treatments was not accounted for. A split-sample method was used for internal validation, while the external validation cohort was partially recruited from the same centers as the derivation cohort, which gives overoptimistic estimates of performance in independent populations. It was unclear whether blinding was applied; which inputs were included in the model as continuous and which were categorized; and how pre-stroke dementia (a candidate predictor) and dependency were operationalized.

The iScore scale has many external validation studies, which indicate sufficient prognostic ability for outcomes other than just mortality. Use of the scale can be aided by an online score calculator (94). Nonetheless, compared to most scales included in this review iScore require substantial baseline information. The revised scale may offer a solution to the issues of acute classification, yet the iScore-r derivation study reported high attrition rates, and with no further external validation studies, the generalizability of the scale remains uncertain.

Preadmission Comorbidities, Level of Consciousness, Age, and Neurological Deficit (PLAN)

Scale Content and Development

The PLAN scale was developed to estimate probability of death and severe disability following ischemic stroke, specifically 30-day and 1-year mortality, and mRS>4 at discharge (95). A risk score ranging from 0-25 is calculated based on: pre-admission dependency, history of cancer, congestive heart failure, atrial fibrillation, consciousness, age, proximal weakness of the leg, weakness of the arm, aphasia and neglect. Higher scores are associated with greater likelihood of death or severe disability. The scale was derived through logistic regression using the same multicenter data source as in the case of iScore. The baseline sample comprised 9,847 patients. However, as a split-sample validation method was applied, only 4,943 of subjects were included in the development set.

Scale Validation and Updating

The derivation study reported results of both apparent and internal validation, with AUROC values ranging from 0.82 to 0.89 for all three outcomes. The scale's performance was not assessed in an independent dataset. External validation was however conducted in two subsequent studies (48, 73). The scale was applied for prediction of good functional outcome, poor outcome, and mortality. In all analyses, PLAN was found to have AUROC values above 0.80.

Critical Appraisal and Clinical Application

Our assessment of PLAN revealed issues predominantly related to three aspects of scale development: predictors, the model derivation procedure, and assessment of performance. In relation to predictors, all originally continuous variables were categorized. There was also a lack of reporting on how pre-stroke dementia and dependency were operationalized, as well as on blinding to outcome for assessment of input variables. In terms of creating the model, variables for multivariable analysis were chosen based on estimated associations in univariable analysis, while the method for selecting final predictors in multivariable analysis seemed unclear. Finally, the scale was only internally validated, using a split-sample method. Calibration was assessed alongside discrimination, however this was limited to performing the Hosmer-Lemeshow test and correlations between observed and expected outcomes. An additional concern is the lack of statement on the method of handling missing data.

Given the increasing use of IV-tPA as a treatment option in ischemic stroke, it is noteworthy that patients receiving this intervention were excluded from the PLAN derivation study. This does not necessarily entail limited applicability of the scale,

particularly as the external validation studies, reporting good performance for PLAN, both included IV-tPA-treated patients. However, as the scale was only applied in two independent dataset, it seems that more evidence is necessary before reaching conclusions on PLAN's generalizability. If an acceptable level of performance is consistently indicated, another issue worth investigating will be whether the relative complexity in scoring impedes implementation of the scale in clinical practice.

Symptomatic Hemorrhage, Baseline NIHSS, Age, Reperfusion, and Location of Clot (SNARL)

Scale Content and Development

The SNARL scale uses the three clinical and two imaging variables in its acronym to predict a good outcome (mRS<3) at 3 months following ischemic stroke treated with endovascular therapy (96). Scores can range from zero to eleven, with higher scores associated with a greater probability of a good outcome. The scale was derived through a logistic regression procedure, using data of 511 patients from a multicenter registry.

Scale Validation and Updating

Based on results of apparent validation, reported AUROC was 0.79 (95%CI: 0.75–0.83). The study also assessed the scale's performance in an independent cohort, comprising 223 patients from the North American Solitaire Acute Stroke registry. For this dataset, AUROC was 0.74 (95%CI: 0.68–0.81). In addition, the authors reported that compared to the THRIVE scale (described below), SNARL presented a 35% improvement in terms of accurately classifying patients' probability of a good outcome. We did not identify any further external validation studies assessing this scale.

Critical Appraisal and Clinical Application

Through our critical appraisal of the SNARL derivation study we identified two sources of bias, both common across the reviewed scales, use of a complete-case analysis approach and selection of predictors based on associations in a univariable statistical procedure. The applied input selection process in multivariable analysis, on the other hand, seemed unclear, as did the use of any blinding methods. Finally, although predictors were well-operationalized, interpretation of imaging findings may be subject to relatively high interobserver variability.

Stroke Subtype, OCSP, Age, and Pre-stroke mRS (SOAR)

Scale Content and Development

SOAR was developed to predict early mortality (inpatient and 7-day) and length of hospital stay, based on the four clinical variables of the scale's acronym (97). Using a logistic regression model, a scoring system ranging from 0 to 8 was derived, with higher scores associated with a greater likelihood of death and extended length of stay. The derivation cohort included 12,355 acute stroke patients (91% ischemic) from a multicenter register, based in the United Kingdom.

Scale Validation and Updating

SOAR was internally validated using a bootstrapping resampling method, with reported AUROC values being the same for both 7-day and inpatient mortality: 0.79 (95%CI: 0.78–0.80). For predicting length of hospital stay, dichotomized at seven days, AUROC was 0.61 (95%CI: 0.60–0.62). Although external validation was not included as part of the derivation paper, SOAR has been subsequently assessed in independent datasets in five studies (98–103). Four studies assessed the scale's performance for predicting early mortality (inpatient, 7-day, discharge, and 90-day). Three found SOAR to have fair discriminatory power, and one, good. One study applied the scale for prediction of length of hospital stay. Discrimination was not formally assessed, however the authors reported that SOAR scores were significantly associated with the outcome (100).

Three external validation studies additionally aimed to improve the scales predictive performance by adding new variables. The modified SOAR (mSOAR) added stroke severity (NIHSS) (98). When compared to SOAR, mSOAR was found to have superior discriminatory power: AUROC of 0.83 (95%CI: 0.79–0.86) vs. 0.79 (95%CI: 0.75–0.84). Noteworthy, performance was assessed in the mSOAR derivation set. Nonetheless, the finding was confirmed in an independent Chinese patient sample: AUROC of 0.78 (95%CI: 0.76–0.81) and 0.79 (95%CI: 0.77–0.80) for discharge and 90-day mortality, respectively, compared to 0.72 (95%CI: 0.70–0.75) and 0.70 (95%CI: 0.69–0.72), respectively, with higher estimates reported for mSOAR. The remaining two updates of SOAR included adding admission blood glucose levels (SOAR-G) and admission sodium (SOAR-Na) (101, 103). Both concluded that the original and revised scales performed well, however without evidence of the latter offering a significant improvement in discriminatory power.

Critical Appraisal and Clinical Application

Reviewing the SOAR derivation study, we noted that the authors intended to select predictors for multivariable analysis based on univariable associations. However, as all candidate predictors were found to be significantly associated with the outcome, using this approach would not have influenced the results. In this case, what seems to be a greater issue, is that sex was not included in the final model, despite the significance of its association in both univariable and multivariable analyses. Risk of bias was increased by excluding all patient with missing data from the study, as well as by not accounting for effects of administered treatments.

For implementation in clinical practice, the simplicity of SOAR appears a major advantage, including easily accessible information on only four variables. Adding NIHSS is the only attempted modification that has significantly improved scale performance. In many centers, where the measure is not routinely used, this will introduce an additional challenge, yet it is worth considering that stroke severity has been consistently found to be associated with post-stroke outcomes. Calculation of mSOAR can be aided by use of an online tool (104).

Stroke Prognosis Instrument (SPI)
Scale Content and Development
SPI was developed to predict risk of stroke or death within 2 years of TIA or minor stroke (105). A score ranging from 0 to 11 is calculated based on five variables: age, history of diabetes and coronary heart disease, acute hypertension, and presentation (TIA or minor stroke). This score assigns patients to one of three risk groups: low (0–2 points), medium (3–6 points), and high (7–11 points). The scale was developed based on survival analysis, specifically using a Cox proportional hazards model. The derivation cohort included 142 patients, who had undergone carotid ultrasonography in a United States tertiary care hospital. Based on data from this sample, an initial SPI score was developed, including only three variables: age, diabetes, and hypertension.

Scale Validation and Updating
In the derivation study, the SPI score was assessed based on its ability to accurately stratify patients according to risk of stroke or death, using data from the development sample, as well as in an independent Canadian cohort, including 330 patients. In the derivation set, the results showed that 3% of patients estimated as being at low risk had a subsequent stroke or died within 2 years of the initial neurovascular event, while the incidence for patients assigned to the medium risk group was 27%, and for those in the high-risk group 48%. For the validation cohort, the incidence of stroke and death were 10, 21, and 59%, for the 3 risk groups respectively. To ameliorate decreased performance estimates in the external set, two more variables were added to the scale, differentiation between a TIA and a minor stroke, and a history of coronary heart disease.

The authors of SPI subsequently externally validated the final scale in four independent cohorts, and used one of these cohorts to develop a modified version of the scale (106). SPI-II was derived based on data from 525 female patients, who participated in the Women's Estrogen for Stroke Trial (107). In addition to the original variables, SPI-II incorporates history of congestive heart failure and prior stroke, with total scores ranging from 0 to 15. Data from three cohorts, with a total of 9,220 patients, were used in a pooled analysis to estimate the AUROC values for both scales, concluding that SPI-II (0.63; 95%CI: 0.62–0.65) had superior discriminatory power to SPI-I (0.59; 95%CI: 0.57–0.60).

SPI-II has been subsequently externally validated in two studies (108, 109). The first found that for prediction of both stroke and death at 1 year, SPI-II had poor discriminatory power (0.62; 95%CI: 0.61–0.64), which further decreased when limiting the outcome measure to recurrent stroke (0.55; 95%CI: 0.51–0.59). In the second study, groups identified as medium and high risk were combined, and the scale applied to predict 3-month recurrence of ischemic events. Here, the scale was found to have an AUROC of 0.55 (95%CI: 0.41–0.69).

Critical Appraisal and Clinical Application
The SPI derivation study had a high risk of bias. Exclusion criteria for study participation included previous stroke and any missing data on variables of interest. As a result, close to 60% of the baseline sample were not included in the analyses, leaving an insufficient number of participants relative to the number of predictors that were investigated. All of these predictors were categorized. Distinguishing between TIAs, minor strokes, and stroke has potential for interobserver variability. All candidate predictors were included in analysis, however forward selection method was used.

In relation to assessing scale performance, we noted that neither discrimination nor calibration were assessed in the SPI derivation study. The chosen validation cohort also differed from the derivation set in that some predictors were measured in alternative ways, and patients with previous strokes were included. The latter introduced an additional problem, as a history of cerebrovascular events was found to be significantly associated with the outcome. However, as this could not be investigated in the derivation set, the variable was not incorporated as a predictor. The he final SPI score seemed to be derived on the basis of a partially erroneous process of rounding up variable coefficient values.

SPI-II is also at high risk of bias, using data from a female-only patient sample. Although the revised scale was found to have significantly increased discriminatory power compared to the original, it was nonetheless poor, as confirmed in subsequent validation studies. The scale's predicted outcome is also problematic, creating a highly heterogenous risk group. On one hand, with a highly diverse range of possible scenarios, identifying a set of predictors both necessary and sufficient for accurate outcome prognosis seems extremely difficult. On the other hand, for clinicians, and particularly for patients, identifying that one belongs to a high-risk group seems of limited value, when this can indicate increased likelihood of anything from a minor stroke with no residual disability to death.

Stroke Thrombolytic Predictive Instrument (S-TPI)
Scale Content and Development
S-TPI was developed to assist clinicians in predicting the outcome of ischemic stroke patients following intravenous IV-tPA (110). Two logistic regression models were created: one for prediction of good outcome (mRS<2) and one for prediction of catastrophic outcome (mRS>4), at 3 months. In addition to IV-tPA treatment, the former model included the following variables: age, initial systolic blood pressure, diabetes, sex, baseline NIHSS score, prior stroke, and symptom onset to treatment time; as well as interaction terms: treatment with blood pressure, sex, prior stroke, and onset to treatment time, and age with NIHSS. For prediction of catastrophic outcome, the model consisted of considerably fewer inputs: age, NIHSS, serum glucose and ASPECTS score, the latter treated as an optional variable, with inclusion subject to availability. The models were derived using a combined dataset from five randomized clinical trials of IV-tPA, involving 1983, 1967, and 1883 patients (depending on the model), out of an initial cohort of 2184.

Scale Validation and Updating
The models were internally validated using a bootstrapping method, creating development and independent test datasets. In the latter, AUROC values were 0.77 [interquartile range

(IQR): 0.76–0.78] and 0.76 (IQR: 0.75–0.78), for prediction of good outcome and catastrophic outcome without ASPECTS, respectively. Calibration was graphically assessed through plotting mean predicted vs. observed rates of patient outcomes across quintiles, and was concluded to be excellent.

S-TPI was subsequently externally validated in three studies (111–113). Two studies assessed discriminatory power based on AUROC, finding the scale to have fair to good performance for both good and catastrophic outcomes. Calibration curves were investigated in all three studies. In each case, S-TPI was found to overestimate the likelihood of a good outcome, particularly at higher levels of observed probabilities. In relation to a catastrophic outcome, findings were mixed, two studies reported the scale to underestimate the likelihood of this outcome, while the third concluded the opposite. One of the studies undertook recalibration of the scale and further added two variables for prediction of good outcome, signs of infarction on brain scan and serum glucose level. The authors reported this improved the scales discriminatory power (AUROC of 0.77 vs. 0.75) (112).

In contrast, the group that developed S-TPI sought to simplify the scale by reducing the number of predictors, with an aim to makes its implementation in routine clinical practice more feasible (114). The process involved removing interaction terms with limited external supporting evidence, removing the ASPECT score, and exploring the use of simpler stroke severity measures. A total of nine models were generated through logistic regression, for prediction of three outcome levels: mRS<2, mRS<3, and mRS>4. Results from apparent validation showed that AUROC values for all models ranged from 0.75 to 0.80. External validation was performed for models predicting mRS<2 and mRS<3, with findings indicating comparable discriminatory power as in the derivation set. The authors concluded that reducing model components did not lead to a substantial deterioration in performance. We have not identified any further publications externally validating the simplified S-TPI models.

Critical Appraisal and Clinical Application
Risk of bias in the derivation paper was increased by use of data from randomized control trials. Inclusion and exclusion criteria for such trials typically lead to recruiting a highly selective group of participants, thus decreasing the generalizability of scales developed based on their data. In addition, it seemed that trial investigators were not blinded to predictors (with the exception of use of IV-tPA vs. placebo) when assessing the outcome. Finally, although patients with missing data were not excluded from the derivation study outright, lack of data imputation would have led to participants being excluded from particular analyses, when they had no data for one or more of the variables used.

In view of scale implementation, it is important to note that individual patient outcome predictions were to be estimated automatically using a computer system, with an open-access version of the instrument also published online. The latter is however no longer available. With no presentation of an easily calculable risk score, estimating probabilities of patient outcomes would be a challenging task for clinicians, particularly taking into account the complexity of the S-TPI models. Despite the effort to simplify the scale, its use would nonetheless require applying the regression model formulae itself. There also seems to be no clear indication of which version of the multiple S-TPI models is the best candidate for implementation. Overall, it appears more evidence of predictive performance is needed before the simplified models can be considered for clinical use, as well as an easily-applicable scoring system.

Totaled Health Risks in Vascular Events Stroke (THRIVE)
Scale Content and Development
The THRIVE scale was originally developed with an aim to support identification of patients who may benefit from endovascular stroke treatments, in terms of 3-month functional outcome and risk of death (115). The scale includes five clinical variables: age, stroke severity, and history of hypertension, diabetes mellitus and atrial fibrillation. On their basis, an integer score is calculated ranging from 0 to 9, with higher scores associated with a greater probability of a poor outcome. The scale was developed using logistic and ordinal regression models. The derivation cohort included participants of the MERCI and Multi MERCI trials of mechanical thrombectomy, with a total of 305 ischemic stroke patients (116, 117).

Scale Validation and Updating
The derivation paper reported results of apparent validation for prediction of good outcome (mRS<3), finding that the final prognostic model had an AUROC of 0.71. The THRIVE score, developed based on estimated odds ratios for each predictor, was assessed in terms of its association with the percentage of patients with a particular outcome. A good outcome was observed in 64.7% of patients with a score of 0–2 and in 10.6% of cases with a score of 6–9. Reported mortality rates were 5.9 and 56.4% for patients with low and high THRIVE scores, respectively.

There have been 16 subsequent studies externally validating THRIVE. These involved patient groups receiving intra-arterial therapy, intravenous thrombolysis, and no acute treatments, and focused on a number of different outcomes – good functional outcome, poor outcome, risk of hemorrhagic transformations, infarct size, and even pulmonary infection (44, 118–132). The majority of studies aimed to predict multiple outcomes, and 15 assessed the scales predictive performance in terms of AUROC values, typically alongside other estimates. Seven studies found the discriminatory power to be either poor or fair, depending on the outcome, four reported it to be poor for all used outcomes, three to be fair, and only one study found the performance to be good, specifically for prediction of mortality rates.

With an aim to improve the scale's predictive performance, a revised version was developed, the THRIVE-c Calculation (133). The modified tool includes the same variables as the original scale, with age and NIHSS score entered as continuous rather than categorized variables. The derivation study reported results of apparent, internal (split-sample) and external validation, with AUROC values ranging from 0.77 to 0.80 for prediction of poor outcome. In the overall study cohort, THRIVE-c was found to have significantly superior predictive performance compared to the original THRIVE score (0.79, 95%CI: 0.78–0.79 vs. 0.75, 95%CI: 0.74–0.76). THRIVE-c has been subsequently externally

validated in a Chinese population of patients receiving IV-tPA (134). The scale was used to predict symptomatic hemorrhage, poor functional outcome and mortality, with reported AUROC values of 0.70, 0.75, and 0.81, respectively.

Critical Appraisal and Clinical Application

Our critical appraisal of the THRIVE derivation study indicated issues with each of the assessed aspects, either due to methodological quality or incomplete reporting. The derivation set consisted of participants recruited to a clinical trial, thus leading to participation of a selective group of subjects. Moreover, the final sample size and method of handling missing data seemed unclear. In relation to input variables, the initial set of candidate predictors appeared limited, omitting a number of factors found to be associated with functional outcome in previous research. Inclusion of specific chronic diseases in multivariable analysis was based on significance of associations in univariable analysis. Three factors, age, stroke severity and success of vessel recanalization, were included in multivariable analyses outright, and all were found to be independently associated with the outcome. It is however unclear why vessel recanalization was not incorporated into the final THRIVE scale. There was also no report of how cut-offs were determined for the derived THRIVE score.

Assessment of scale performance in the derivation study was limited to apparent validation. Moreover, discriminatory power was tested only for the model predicting good outcome; it was not assessed for the model predicting mortality or for the derived THRIVE score. Although the scale has undergone extensive external validation since its development, findings from these studies do not seem to support a favorable judgement on the scale's prognostic performance. THRIVE-c appears to be a superior alternative, yet up-to-date we have found only one independent validation study assessing the scale's predictive ability. In view of use in routine clinical practice, inclusion of relatively few variables, based on information typically available in an acute setting, is a relevant advantage of THRIVE. Score calculation can additionally be aided be use of an online tool (135). However, existing evidence on predictive ability does not seem to merit implementation.

DISCUSSION

There are many prognostic tools available for use in acute stroke settings. We have reviewed a selection of these and common themes emerge. Our primary interests were methodological quality of derivation, subsequent external validation and scale usability in routine clinical practice. Across 10 primary derivation studies of better-known scales, we identified potential sources of bias in each. However, it is the results of external validation studies that allow us to conclude on the scales prognostic value and applicability. We found that all scales, but one, were externally validated.

While there was a range of prognostic accuracies reported, most scales had properties that would be considered "acceptable." This is perhaps not surprising as the scales tended to measure the same concepts of demographics, comorbidity, initial stroke severity and pre-stroke functional status. Where scale developers

have tried to add additional elements to these core predictors, the gain in predictive power has been limited. However, most scales have been developed with a biomedical focus and it is plausible that other less traditional factors could improve utility of the scales, for example incorporating measures of frailty, resilience, provision of rehabilitation services and social support, or the clinician's clinical gestalt.

Based on our literature search, we identified the highest number of external validation studies for THRIVE. Yet results indicated a level of predictive performance insufficient to merit the scale's use in a clinical setting. Our critical appraisal may partly explain this, identifying concerns relating to all aspects of THRIVE's derivation process. Four other scales, ASTRAL, DRAGON, iScore, and SOAR, have also been validated in multiple independent datasets. For all, findings suggested a level of predictive ability that would merit implementing the scales in clinical practice. Moreover, one study reported ASTRAL and DRAGON to predict patient outcomes more accurately than clinicians (46). Although evidence regarding the performance of other scales included in this review seemed insufficient to reach firm conclusions, a number of these tools were derived with relatively low risk of bias, and future research is likely to confirm their prognostic value. These include PLAN, SNARL, S-TPI, as well as updated scale versions: MRI-DRAGON, iScore-r and mSOAR.

For a number of reasons, it is challenging to directly compare the reviewed prognostic tools, and we have deliberately chosen to avoid naming a single preferred tool. Firstly, studies assessing more than one scale as part of an external validation process are relatively uncommon. When conducted, findings are often difficult to interpret, small differences in predictive ability between scales may arise from the superiority of one over another, but they could also be attributed to an incidentally greater similarity between the validation set and the derivation set of the scale found to perform best. Secondly, although satisfactory predictive ability is essential for a scale to be clinically useful, it is not sufficient. A number of other factors need to be considered, including feasibility in routine clinical practice, the relevance of the predicted outcome to the specific context, and whether applying the tool improves clinical decision-making, patient outcomes or cost-effectiveness of services (136). To help answer these questions, it is necessary to conduct impact studies, a stage in prognostic research that to our knowledge none of the described scales have yet reached.

Our focused literature review has strengths and limitations. We recognize that there have been many high quality systematic and narrative reviews of stroke prognosis scales (10, 137). We hope that our review offers a novel focus. We have appraised relevant stroke scales against each other; very few derivation papers have done this, despite its importance when choosing which scale to use. Additionally, we have followed the PRISMA systematic review guidelines (9) when designing and completing our study and based our appraisal on the CHARMS checklist (16). Our intention was not to offer a comprehensive review, rather we choose exemplar scales that featured in high impact journals and so by implication would be amongst the best known in the clinical community. In our assessment of feasibility we identified clinical and radiological features that

may be challenging to assess in the acute setting (63, 138). Our focus was routine stroke practice (139) and our comments on feasibility may not apply to specialist stroke centers. It takes time for scales to become established and our review did not include recently published scales, for example those designed to inform thrombectomy decisions (140). However, the literature describing these scales is increasing rapidly and soon there may be sufficient validation studies.

We used data from our focused literature review to compare long-term stroke prognosis scales. We found many scales with similar content and properties. Although development of the scales did not always follow methodological best practice, most of these scales have been subsequently validated. Rather than developing new scales, prognostic research in stroke should now focus on implementation and comparative analyses.

AUTHOR CONTRIBUTIONS

SS and TQ contributed to the conception and design of the study. TQ designed the literature search strategy, SS performed the search and screened the results. SS and BD extracted study data, all authors were involved in critical appraisal of included studies. BD and SS drafted the final manuscript. All authors critically revised the manuscript, approved its final version, and agreed to be accountable for its content.

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fneur.2019.00274/full#supplementary-material

REFERENCES

1. Johnston KC, Wagner DP, Wang XQ, Newman GC, Thijs V, Sen S, et al. Validation of an acute ischemic stroke model: does diffusion-weighted imaging lesion volume offer a clinically significant improvement in prediction of outcome? *Stroke*. (2007) 38:1820–5. doi: 10.1161/STROKEAHA.106.479154

2. National Institute for Health and Care Excellence. *Stroke and Transient Ischaemic Attack in Over 16s: Diagnosis and Initial Management*. NICE Guidance and Guidelines. (2008). Available online at: https://www.nice.org.uk/guidance/cg68 (Cited August 17, 2018).

3. Counsell C, Dennis M. Systematic review of prognostic models in patients with acute stroke. *Cerebrovasc Dis*. (2001) 12:159–70. doi: 10.1159/000047699

4. Mattishent K, Kwok CS, Mahtani A, Pelpola K, Myint PK, Loke YK. Prognostic indices for early mortality in ischaemic stroke - meta-analysis. *Acta Neurol Scand*. (2016) 133:41–8. doi: 10.1111/ane.12421

5. Teale EA, Forster A, Munyombwe T, Young JB. A systematic review of case-mix adjustment models for stroke. *Clin Rehabil*. (2012) 26:771–86. doi: 10.1177/0269215511433068

6. Moons KGM, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? *BMJ*. (2009) 338:b375. doi: 10.1136/bmj.b375

7. Reilly BM, Evans AT. Translating clinical research into clinical practice: Impact of using prediction rules to make decisions. *Ann Int Med*. (2006) 144:201–9. doi: 10.7326/0003-4819-144-3-200602070-00009

8. Steyerberg EW, Moons KGM, 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. doi: 10.1371/journal.pmed.1001381

9. Quinn TJ, Singh S, Lees KR, Bath PM, Myint PK, Collaborators V. Validating and comparing stroke prognosis scales. *Neurology*. (2017) 89:997–1002. doi: 10.1212/WNL.0000000000004332

10. Fahey M, Crayton E, Wolfe C, Douiri A. Clinical prediction models for mortality and functional outcome following ischemic stroke: a systematic review and meta-analysis. *PLoS ONE*. (2018) 13:e0185402. doi: 10.1371/journal.pone.0185402

11. Moher D, Liberati A, Tetzlaff J, Altman DG, Grp P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Int Med*. (2009) 151:264–W64. doi: 10.7326/0003-4819-151-4-200908180-00135

12. Quinn TJ, Dawson J, Walters MR, Lees KR. Functional outcome measures in contemporary stroke trials. *Int J Stroke*. (2009) 4:200–5. doi: 10.1111/j.1747-4949.2009.00271.x

13. Lees R, Fearon P, Harrison JK, Broomfield NM, Quinn TJ. Cognitive and mood assessment in stroke research focused review of contemporary studies. *Stroke*. (2012) 43:1678. doi: 10.1161/STROKEAHA.112.653303

14. 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. doi: 10.1371/journal.pmed.1001744

15. Gorelick MH. Bias arising from missing data in predictive models. *J Clin Epidemiol*. (2006) 59:1115–23. doi: 10.1016/j.jclinepi.2004.11.029

16. Van der Heijden GJMG, Donders ART, Stijnen T, Moons KGM. Imputation of missing values is superior to complete case analysis and the missing-indicator method in multivariable diagnostic research: a clinical example. *J Clin Epidemiol*. (2006) 59:1102–9. doi: 10.1016/j.jclinepi.2006.01.015

17. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med*. (2006) 25:127–41. doi: 10.1002/sim.2331

18. Bennette C, Vickers A. Against quantiles: categorization of continuous variables in epidemiologic research, and its discontents. *BMC Med Res Methodol*. (2012) 12:21. doi: 10.1186/1471-2288-12-21

19. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol*. (1996) 49:1373–9. doi: 10.1016/S0895-4356(96)00236-3

20. Austin PC, Steyerberg EW. Events per variable (EPV) and the relative performance of different strategies for estimating the out-of-sample validity of logistic regression models. *Stat Methods Med Res*. (2017) 26:796–808. doi: 10.1177/0962280214558972

21. Sauerbrei W, Royston P, Binder H. Selection of important variables and determination of functional form for continuous predictors in multivariable model building. *Stat Med*. (2007) 26:5512–28. doi: 10.1002/sim.3148

22. Sun GW, Shook TL, Kay GL. Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. *J Clin Epidemiol*. (1996) 49:907–16. doi: 10.1016/0895-4356(96)00025-X

23. Babyak MA. What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic Med*. (2004) 66:411–21. doi: 10.1097/01.psy.0000127692.23278.a9

24. Mantel N. Why stepdown procedures in variable selection. *Technometrics*. (1970) 12:621–5. doi: 10.1080/00401706.1970.10488701

25. Royston P, Moons KGM, Altman DG, Vergouwe Y. Prognosis and prognostic research: developing a prognostic model. *Brit Med J*. (2009) 338:b604. doi: 10.1136/bmj.b604

26. Steyerberg EW, Harrell FE, Borsboom GJJM, Eijkemans MJC, Vergouwe Y, Habbema JDF. Internal validation of predictive models: Efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol*. (2001) 54:774–81. doi: 10.1016/S0895-4356(01)00341-9

27. Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. New York, NY: Springer Science & Business Media. (2008).

28. Minne L, Ludikhuize J, de Rooij SEJA, Abu-Hanna A. Characterizing predictive models of mortality for older adults and their validation for use in clinical practice. *J Am Geriatr Soc*. (2011) 59:1110–5. doi: 10.1111/j.1532-5415.2011.03411.x

29. Moons KGM, 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. doi: 10.1136/heartjnl-2011-301247

30. Pencina MJ, D'Agostino RB. Evaluating discrimination of risk prediction models the c statistic. *Jama-J Am Med Assoc*. (2015) 314:1063–4. doi: 10.1001/jama.2015.11082

31. Carter JV, Pan JM, Rai SN, Galandiuk S. ROC-ing along: Evaluation and interpretation of receiver operating characteristic curves. *Surgery*. (2016) 159:1638–45. doi: 10.1016/j.surg.2015.12.029

32. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models a framework for traditional and novel measures. *Epidemiology*. (2010) 21:128–38. doi: 10.1097/EDE.0b013e3181c30fb2

33. Altman DG, Vergouwe Y, Royston P, Moons KGM. Prognosis and prognostic research: validating a prognostic model. *Brit Med J*. (2009) 338:b605. doi: 10.1136/bmj.b605

34. Newcommon NJ, Green TL, Haley E, Cooke T, Hill MD. Improving the assessment of outcomes in stroke: use of a structured interview to assign grades on the modified Rankin Scale. *Stroke*. (2003) 34:377–8. doi: 10.1161/01.STR.0000055766.99908.58

35. Saver JL, Filip B, Hamilton S, Yanes A, Craig S, Cho M, et al. Improving the Reliability of Stroke Disability Grading in Clinical Trials and Clinical Practice The Rankin Focused Assessment (RFA). *Stroke*. (2010) 41:992–5. doi: 10.1161/STROKEAHA.109.571364

36. Vanswieten JC, Koudstaal PJ, Visser MC, Schouten HJA, Vangijn J. Interobserver agreement for the assessment of handicap in stroke patients. *Stroke*. (1988) 19:604–7. doi: 10.1161/01.STR.19.5.604

37. *Stroke Scales and Related Information: National Institute of Health, National Institute of Neurological Disorders and Stroke*. (2016). Available online at: https://www.ninds.nih.gov/Disorders/Patient-Caregiver-Education/Preventing-Stroke/Stroke-Scales-and-Related-Information (Cited December 16, 2018).

38. Ntaios G, Faouzi M, Ferrari J, Lang W, Vemmos K, Michel P. An integer-based score to predict functional outcome in acute ischemic stroke: the ASTRAL score. *Neurology*. (2012) 78:1916–22. doi: 10.1212/WNL.0b013e318259e221

39. Michel P, Odier C, Rutgers M, Reichhart M, Maeder P, Meuli R, et al. The Acute STroke Registry and Analysis of Lausanne (ASTRAL). Design and Baseline Analysis of an Ischemic Stroke Registry Including Acute Multimodal Imaging. *Stroke*. (2010) 41:2491–8. doi: 10.1161/STROKEAHA.110.596189

40. Vemmos KN, Takis CE, Georgilis K, Zakopoulos NA, Lekakis JP, Papamichael CM, et al. The Athens stroke registry: results of a five-year hospital-based study. *Cerebrovasc Dis*. (2000) 10:133–41. doi: 10.1159/000016042

41. Vienna Stroke Registry, Vienna Stroke Study Group. The Vienna Stroke Registry: objectives and methodology. The Vienna Stroke Study Group. *Wien Klin Wochensch*. (2001) 113:141–7.

42. Asuzu D, Nystrom K, Amin H, Schindler J, Wira C, Greer D, et al. Comparison of 8 scores for predicting symptomatic intracerebral hemorrhage after IV thrombolysis. *Neurocrit Care*. (2015) 22:229–33. doi: 10.1007/s12028-014-0060-2

43. Cooray C, Mazya M, Bottai M, Dorado L, Skoda O, Toni D, et al. External Validation of the ASTRAL and DRAGON Scores for Prediction of Functional Outcome in Stroke. *Stroke*. (2016) 47:1493–9. doi: 10.1161/STROKEAHA.116.012802

44. Kuster GW, Dutra LA, Brasil IP, Pacheco EP, Arruda MJC, Volcov C, et al. Performance of four ischemic stroke prognostic scores in a Brazilian population. *Arq Neuro-Psiquiat*. (2016) 74:133–7. doi: 10.1590/0004-282X20160002

45. Liu G, Ntaios G, Zheng H, Wang Y, Michel P, Wang DZ, et al. External Validation of the ASTRAL Score to Predict 3- and 12-Month Functional Outcome in the China National Stroke Registry. *Stroke*. (2013) 44:1443–5. doi: 10.1161/STROKEAHA.113.000993

46. Ntaios G, Gioulekas F, Papavasileiou V, Strbian D, Michel P. ASTRAL, DRAGON and SEDAN scores predict stroke outcome more accurately than physicians. *Eur J Neurol*. (2016) 23:1651–7. doi: 10.1111/ene.13100

47. Papavasileiou V, Milionis H, Michel P, Makaritsis K, Vemmou A, Koroboki E, et al. ASTRAL Score Predicts 5-Year Dependence and Mortality in Acute Ischemic Stroke. *Stroke*. (2013) 44:1616–20. doi: 10.1161/STROKEAHA.113.001047

48. Wang WY, Sang WW, Jin D, Yan SM, Hong Y, Zhang H, et al. The prognostic value of the iScore, the PLAN Score, and the ASTRAL Score in Acute Ischemic Stroke. *J Stroke Cerebrovasc*. (2017) 26:1233–8. doi: 10.1016/j.jstrokecerebrovasdis.2017.01.013

49. *ASTRAL Score for Ischemic Stroke: MDCalc*. Available online at: https://www.mdcalc.com/astral-score-stroke (Cited October 22, 2018).

50. Strbian D, Meretoja A, Ahlhelm FJ, Pitkäniemi J, Lyrer P, Kaste M, et al. Predicting outcome of IV thrombolysis–treated ischemic stroke patients. *The DRAGON Score*. (2012) 78:427–32. doi: 10.1212/WNL.0b013e318245d2a9

51. Giralt-Steinhauer E, Rodríguez-Campello A, Cuadrado-Godia E, Ois Á, Jiménez-Conde J, Soriano-Tárraga C, et al. External Validation of the DRAGON Score in an Elderly Spanish Population: Prediction of Stroke Prognosis after IV Thrombolysis. *Cerebrovasc Dis*. (2013) 36:110–4. doi: 10.1159/000352061

52. Ovesen C, Christensen A, Nielsen JK, Christensen H. External validation of the ability of the DRAGON score to predict outcome after thrombolysis treatment. *J Clin Neurosci*. (2013) 20:1635–6.

53. Strbian D, Seiffge DJ, Breuer L, Numminen H, Michel P, Meretoja A, et al. Validation of the DRAGON Score in 12 Stroke Centers in Anterior and Posterior Circulation. *Stroke*. (2013) 44:2718–21. doi: 10.1161/STROKEAHA.113.002033

54. Van Hooff RJ, Nieboer K, De Smedt A, Moens M, De Deyn PP, De Keyser J, et al. Validation assessment of risk tools to predict outcome after thrombolytic therapy for acute ischemic stroke. *Clin Neurol Neurosur*. (2014) 125:189–93. doi: 10.1016/j.clineuro.2014.08.011

55. Baek JH, Kim K, Lee YB, Park KH, Park HM, Shin DJ, et al. Predicting Stroke Outcome Using Clinical-versus Imaging-based Scoring System. *J Stroke Cerebrovasc*. (2015) 24:642–8. doi: 10.1016/j.jstrokecerebrovasdis.2014.10.009

56. Zhang XM, Liao XL, Wang CJ, Liu LP, Wang CX, Zhao XQ, et al. Validation of the DRAGON score in a chinese population to predict functional outcome of intravenous thrombolysis-treated stroke patients. *J Stroke Cerebrovasc*. (2015) 24:1755–60. doi: 10.1016/j.jstrokecerebrovasdis.2015.03.046

57. Wang A, Pednekar N, Lehrer R, Todo A, Sahni R, Marks S, et al. DRAGON score predicts functional outcomes in acute ischemic stroke patients receiving both intravenous tissue plasminogen activator and endovascular therapy. *Surg Neurol Int*. (2017) 8:149. doi: 10.4103/2152-7806.210993

58. Turc G, Apoil M, Naggara O, Calvet D, Lamy C, Tataru AM, et al. Magnetic resonance imaging-DRAGON Score: 3-month outcome prediction after intravenous thrombolysis for anterior circulation stroke. *Stroke*. (2013) 44:1323–8. doi: 10.1161/STROKEAHA.111.000127

59. Barber PA, Demchuk AM, Zhang JJ, Buchan AM, Grp AS. Validity and reliability of a quantitative computed tomography score in predicting outcome of hyperacute stroke before thrombolytic therapy. *Lancet*. (2000) 355:1670–4. doi: 10.1016/S0140-6736(00)02237-6

60. Barber PA, Hill MD, Eliasziw M, Demchuk AM, Pexman JHW, Hudon ME, et al. Imaging of the brain in acute ischaemic stroke:

comparison of computed tomography and magnetic resonance diffusion-weighted imaging. *J Neurol Neurosur Psychiatry.* (2005) 76:1528–33. doi: 10.1136/jnnp.2004.059261

61. Turc G, Aguettaz P, Ponchelle-Dequatre N, Hénon H, Naggara O, Leclerc X, et al. External Validation of the MRI-DRAGON score: early prediction of stroke outcome after intravenous thrombolysis. *PLoS ONE.* (2014) 9:e99164. doi: 10.1371/journal.pone.0099164

62. *DRAGON Score for Post-TPA Stroke Outcome: MDCalc.* Available online at: https://www.mdcalc.com/dragon-score-post-tpa-stroke-outcome (Cited October 22, 2018).

63. Wardlaw JM, Dorman PJ, Lewis SC, Sandercock PAG. Can stroke physicians and neuroradiologists identify signs of early cerebral infarction on CT? *J Neurol Neurosur Psychiatry.* (1999) 67:651–3. doi: 10.1136/jnnp.67.5.651

64. Wardlaw JM, Mielke O. Early signs of brain infarction at CT: observer reliability and outcome after thrombolytic treatment - Systematic review. *Radiology.* (2005) 235:444–53. doi: 10.1148/radiol.2352040262

65. Mair G, Boyd EV, Chappell FM, von Kummer R, Lindley RI, Sandercock P, et al. Sensitivity and specificity of the hyperdense artery sign for arterial obstruction in acute ischemic stroke. *Stroke.* (2015) 46:102–7. doi: 10.1161/STROKEAHA.114.007036

66. Chalela JA, Kidwell CS, Nentwich LM, Luby M, Butman JA, Demchuk AM, et al. Magnetic resonance imaging and computed tomography in emergency assessment of patients with suspected acute stroke: a prospective comparison. *Lancet.* (2007) 369:293–8. doi: 10.1016/S0140-6736(07)60151-2

67. Reid JM, Gubitz GJ, Dai DW, Kydd D, Eskes G, Reidy Y, et al. Predicting functional outcome after stroke by modelling baseline clinical and CT variables. *Age Ageing.* (2010) 39:360–6. doi: 10.1093/ageing/afq027

68. Reid JM, Dai DW, Christian C, Reidy Y, Counsell C, Gubitz GJ, et al. Developing predictive models of excellent and devastating outcome after stroke. *Age Ageing.* (2012) 41:560–4. doi: 10.1093/ageing/afs034

69. Reid JM, Dai D, Thompson K, Christian C, Reidy Y, Counsell C, et al. Five-simple-variables risk score predicts good and devastating outcome after stroke. *Int J Phys Med Rehabil.* (2014) 2:186. doi: 10.4172/2329-9096.1000186

70. Bamford JM, Sandercock PAG, Warlow CP, Slattery J. Interobserver agreement for the assessment of handicap in stroke patients. *Stroke.* (1989) 20:828. doi: 10.1161/01.STR.20.6.828

71. Teasdale G, Jennett B. Assessment of coma and impaired consciousness. A practical scale. *Lancet.* (1974) 2:81–4. doi: 10.1016/S0140-6736(74)91639-0

72. Reid JM, Dai DW, Delmonte S, Counsell C, Phillips SJ, MacLeod MJ. Simple prediction scores predict good and devastating outcomes after stroke more accurately than physicians. *Age Ageing.* (2017) 46:421–6. doi: 10.1093/ageing/afw197

73. Ayis SA, Coker B, Rudd AG, Dennis MS, Wolfe CDA. Predicting independent survival after stroke: a European study for the development and validation of standardised stroke scales and prediction models of outcome. *J Neurol Neurosur Psychiatry.* (2013) 84:288–96. doi: 10.1136/jnnp-2012-303657

74. Quinn TJ, Langhorne P, Stott dJ. Barthel Index for stroke trials. *Stroke.* (2011) 42:1146–51. doi: 10.1161/STROKEAHA.110.598540

75. Leeflang MMG, Moons KGM, Reitsma JB, Zwinderman AH. Bias in sensitivity and specificity caused by data-driven selection of optimal cutoff values: Mechanisms, magnitude, and solutions. *Clin Chem.* (2008) 54:729–37. doi: 10.1373/clinchem.2007.096032

76. Saposnik G, Kapral MK, Liu Y, Hall R, O'Donnell M, Raptis S, et al. IScore: a risk score to predict death early after hospitalization for an acute ischemic stroke. *Circulation.* (2011) 123:739-U172. doi: 10.1161/CIRCULATIONAHA.110.983353

77. Cote R, Hachinski VC, Shurvell BL, Norris JW, Wolfson C. The Canadian Neurological scale: a preliminary study in acute stroke. *Stroke.* (1986) 17:731–7. doi: 10.1161/01.STR.17.4.731

78. Adams HP, Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, et al. Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in acute stroke treatment. *Stroke.* (1993) 24:35–41. doi: 10.1161/01.STR.24.1.35

79. Saposnik G, Raptis S, Kapral MK, Liu Y, Tu JV, Mamdani M, et al. The iScore predicts poor functional outcomes early after hospitalization for an acute ischemic stroke. *Stroke.* (2011) 42:3421-U161. doi: 10.1161/STROKEAHA.111.623116

80. Saposnik G, Fang JM, Kapral MK, Tu JV, Mamdani M, Austin P, et al. The iScore predicts effectiveness of thrombolytic therapy for acute ischemic stroke. *Stroke.* (2012) 43:1315–22. doi: 10.1161/STROKEAHA.111.646265

81. Saposnik G, Demchuk A, Tu JV, Johnston SC, SORCan SORC. The iScore predicts efficacy and risk of bleeding in the national institute of neurological disorders and stroke tissue plasminogen activator stroke trial. *J Stroke Cerebrovasc.* (2013) 22:876–82. doi: 10.1016/j.jstrokecerebrovasdis.2012.09.001

82. Saposnik G, Reeves MJ, Johnston SC, Bath PMW, Ovbiagele B, Collaboration V. Predicting clinical outcomes after thrombolysis using the iscore results from the virtual international stroke trials archive. *Stroke.* (2013) 44:2755–9. doi: 10.1161/STROKEAHA.113.003258

83. Bejot Y, Jacquin A, Daubail B, Durier J, Giroud M. Population-based validation of the iscore for predicting mortality and early functional outcome in ischemic stroke patients. *Neuroepidemiology.* (2013) 41:169–73. doi: 10.1159/000354634

84. Bejot Y, Daubail B, Sensenbrenner B, Legris N, Durier J, Giroud M. iScore for predicting institutional care after ischemic stroke: a population-based study. *J Stroke Cerebrovasc.* (2015) 24:694–8. doi: 10.1016/j.jstrokecerebrovasdis.2014.11.010

85. Park TH, Saposnik G, Bae HJ, Lee SJ, Lee KB, Lee J, et al. The iScore predicts functional outcome in korean patients with ischemic stroke. *Stroke.* (2013) 44:1440–2. doi: 10.1161/STROKEAHA.111.000748

86. Park TH, Park SS, Ko Y, Lee SJ, Lee KB, Lee J, et al. The iScore predicts clinical response to tissue plasminogen activator in korean stroke patients. *J Stroke Cerebrovasc.* (2014) 23:367–73. doi: 10.1016/j.jstrokecerebrovasdis.2013.05.025

87. Dragoumanos V, Tzirogiannis KN, Panoutsopoulos GI, Krikonis K, Fousteris E, Vourvou M, et al. Evaluation of IScore validity in a Greek cohort of patients with type 2 diabetes. *BMC Neurol.* (2013) 13:121. doi: 10.1186/1471-2377-13-121

88. Ewara EM, Isaranuwatchai W, Bravata DM, Williams LS, Fang J, Hoch JS, et al. The iScore predicts total healthcare costs early after hospitalization for an acute ischemic stroke. *Int J Stroke.* (2015) 10:1179–86. doi: 10.1111/ijs.12641

89. Nikneshan D, Raptis R, Pongmoragot J, Zhou LM, Johnston SC, Saposnik G, et al. Predicting clinical outcomes and response to thrombolysis in acute stroke patients with diabetes. *Diabetes Care.* (2013) 36:2041–7. doi: 10.2337/dc12-2095

90. Xu J, Tao YL, Xie XW, Liu GF, Wang AX, Wang YJ, et al. A Comparison of mortality prognostic scores in ischemic stroke patients. *J Stroke Cerebrovasc.* (2016) 25:241–7. doi: 10.1016/j.jstrokecerebrovasdis.2015.09.021

91. Zhang N, Liu GF, Zhang GH, Fang JM, Wang YL, Zhao XQ, et al. External validation of the iScore for predicting ischemic stroke mortality in patients in China. *Stroke.* (2013) 44:1924–9. doi: 10.1161/STROKEAHA.111.000172

92. Abdul-Rahim AH, Dickie DA, Selvarajah JR, Lees KR, Quinn TJ. Stroke aetiological classification relaibility and effect on trial sample size. *Trials.* (2019) 20:107. doi: 10.1186/s13063-019-3222-x

93. Sung SF, Chen YW, Hung LC, Lin HJ. Revised iScore to predict outcomes after acute ischemic stroke. *J Stroke Cerebrovasc.* (2014) 23:1634–9. doi: 10.1016/j.jstrokecerebrovasdis.2014.01.010

94. *ISCORE Ischemic Stroke Predictive Risk Score: SORCan.* Available online at: http://www.sorcan.ca/iscore/index.html (Cited October 22, 2018).

95. O'Donnell MJ, Fang JM, D'Uva C, Saposnik G, Gould L, McGrath E, et al. The PLAN score a bedside prediction rule for death and severe disability following acute ischemic stroke. *Arch Intern Med.* (2012) 172:1548–56. doi: 10.1001/2013.jamainternmed.30

96. Prabhakaran S, Jovin TG, Tayal AH, Hussain MS, Nguyen TN, Sheth KN, et al. Posttreatment variables improve outcome prediction after intra-arterial therapy for acute ischemic stroke. *Cerebrovasc Dis.* (2014) 37:356–63. doi: 10.1159/000362591

97. Myint PK, Clark AB, Kwok CS, Davis J, Durairaj R, Dixit AK, et al. The SOAR (Stroke subtype, Oxford Community Stroke Project classification, Age, prestroke modified Rankin) score strongly predicts early outcomes in acute stroke. *Int J Stroke.* (2014) 9:278–83. doi: 10.1111/ijs.12088

98. Abdul-Rahim AH, Quinn TJ, Alder S, Clark AB, Musgrave SD, Langhorne P, et al. Derivation and validation of a novel prognostic scale (Modified-Stroke Subtype, Oxfordshire Community Stroke Project Classification,

Age, and Prestroke Modified Rankin) to Predict Early Mortality in Acute Stroke. *Stroke.* (2016) 47:74–9. doi: 10.1161/STROKEAHA.115.009898

99. Kwok CS, Potter JF, Dalton G, George A, Metcalf AK, Ngeh J, et al. The SOAR stroke score predicts inpatient and 7-day mortality in acute stroke. *Stroke.* (2013) 44:2010–2. doi: 10.1161/STROKEAHA.113.001148

100. Kwok CS, Clark AB, Musgrave SD, Potter JF, Dalton G, Day DJ, et al. The SOAR stroke score predicts hospital length of stay in acute stroke: an external validation study. *Int J Clin Pract.* (2015) 69:659–65. doi: 10.1111/ijcp.12577

101. McCall SJ, Alanazi TA, Clark AB, Musgrave SD, Bettencourt-Silva JH, Bachmann MO, et al. Hyperglycaemia and the SOAR stroke score in predicting mortality. *Diabetes Vasc Dis Res.* (2018) 15:114–21. doi: 10.1177/1479164117743034

102. Wan H, Pan YS, Meng X, Wang CJ, Liao XL, Wang D, et al. Validation of the mSOAR and SOAR scores to predict early mortality in Chinese acute stroke patients. *Plos ONE.* (2017) 12:e0180444. doi: 10.1371/journal.pone.0180444

103. Adekunle-Olarinde IR, McCall SJ, Barlas RS, Wood AD, Clark AB, Bettencourt-Silva JH, et al. Addition of sodium criterion to SOAR stroke score. *Acta Neurol Scand.* (2017) 135:553–9. doi: 10.1111/ane.12634

104. *Modified SOAR Score for Stroke: MDCalc.* Available online at: https://www.mdcalc.com/modified-soar-score-stroke (Cited October 22, 2018).

105. Kernan WN, Horwitz RI, Brass LM, Viscoli CM, Taylor KW. A prognostic system for transient ischemia or minor stroke. *Ann Int Med.* (1991) 114:552–7. doi: 10.7326/0003-4819-114-7-552

106. Kernan WN, Viscoli CM, Brass LM, Makuch RW, Sarrel PM, Roberts RS, et al. The Stroke Prognosis Instrument II (SPI-II) - A clinical prediction instrument for patients with transient ischemia and nondisabling ischemic stroke. *Stroke.* (2000) 31:456–62. doi: 10.1161/01.STR.31.2.456

107. Kernan WN, Brass LM, Viscoli CM, Sarrel PM, Makuch R, Horwitz RI. Estrogen after ischemic stroke: clinical basis and design of the Women's Estrogen for Stroke Trial. *J Stroke Cerebrovasc Dis.* (1998) 7:85–95. doi: 10.1016/S1052-3057(98)80026-8

108. Navi BB, Kamel H, Sidney S, Klingman JG, Nguyen-Huynh MN, Johnston SC. Validation of the stroke prognostic instrument-ii in a large, modern, community-based cohort of ischemic stroke survivors. *Stroke.* (2011) 42:3392–6. doi: 10.1161/STROKEAHA.111.620336

109. Liu J, Li MN, Liu JY. Evaluation of the ESRS and SPI-II scales for short-term prognosis of minor stroke and transient ischemic attack. *Neurol Res.* (2013) 35:568–72. doi: 10.1179/1743132813Y.0000000172

110. Kent DM, Selker HP, Ruthazer R, Bluhmki E, Hacke W. The stroke-thrombolytic predictive instrument - A predictive instrument for intravenous thrombolysis in acute ischemic stroke. *Stroke.* (2006) 37:2957–62. doi: 10.1161/01.STR.0000249054.96644.c6

111. Uyttenboogaart M, Stewart RE, Vroomen PC, Luijckx GJ, De Keyser J. Utility of the stroke-thrombolytic predictive instrument. *J Neurol Neurosur Psychiatry.* (2008) 79:1079–81. doi: 10.1136/jnnp.2008.145698

112. McMeekin P, Flynn D, Ford GA, Rodgers H, Thomson RG. Validating the stroke-thrombolytic predictive instrument in a population in the United Kingdom. *Stroke.* (2012) 43:3378–81. doi: 10.1161/STROKEAHA.112.671073

113. Decourcelle A, Moulin S, Dequatre-Ponchelle N, Bodenant M, Rossi C, Girot M, et al. Are the results of intravenous thrombolysis trials reproduced in clinical practice? Comparison of observed and expected outcomes with the stroke-thrombolytic predictive instrument (STPI). *Rev Neurol.* (2017) 173:381–7. doi: 10.1016/j.neurol.2017.03.023

114. Kent DM, Ruthazer R, Decker C, Jones PG, Saver JL, Bluhmki E, et al. Development and validation of a simplified stroke-thrombolytic predictive instrument. *Neurology.* (2015) 85:942–9. doi: 10.1212/WNL.0000000000001925

115. Flint AC, Cullen SP, Faigeles BS, Rao VA. Predicting long-term outcome after endovascular stroke treatment: the totaled health risks in vascular events score. *Am J Neuroradiol.* (2010) 31:1192–6. doi: 10.3174/ajnr.A2050

116. Smith WS, Sung G, Starkman S, Saver JL, Kidwell CS, Gobin YP, et al. Safety and efficacy of mechanical embolectomy in acute ischemic stroke - Results of the MERCI trial. *Stroke.* (2005) 36:1432–8. doi: 10.1161/01.STR.0000171066.25248.1d

117. Smith WS, Sung G, Saver J, Budzik R, Duckwiler G, Liebeskind DS, et al. Mechanical thrombectomy for acute ischemic stroke -

Final results of the multi MERCI trial. *Stroke.* (2008) 39:1205–12. doi: 10.1161/STROKEAHA.107.497115

118. Flint AC, Faigeles BS, Cullen SP, Kamel H, Rao VA, Gupta R, et al. THRIVE Score predicts ischemic stroke outcomes and thrombolytic hemorrhage risk in VISTA. *Stroke.* (2013) 44:3365–9. doi: 10.1161/STROKEAHA.113.002794

119. Flint AC, Kamel H, Rao VA, Cullen SP, Faigeles BS, Smith WS. Validation of the Totaled Health Risks In Vascular Events (THRIVE) score for outcome prediction in endovascular stroke treatment. *Int J Stroke.* (2014) 9:32–9. doi: 10.1111/j.1747-4949.2012.00872.x

120. Flint AC, Gupta R, Smith WS, Kamel H, Faigeles BS, Cullen SP, et al. The THRIVE score predicts symptomatic intracerebral hemorrhage after intravenous tPA administration in SITS-MOST. *Int J Stroke.* (2014) 9:705–10. doi: 10.1111/ijs.12335

121. Flint AC, Xiang B, Gupta R, Nogueira RG, Lutsep HL, Jovin TG, et al. THRIVE score predicts outcomes with a third-generation endovascular stroke treatment device in the TREVO-2 trial. *Stroke.* (2013) 44:3370–5. doi: 10.1161/STROKEAHA.113.002796

122. Flint AC, Cullen SP, Rao VA, Faigeles BS, Pereira VM, Levy EI, et al. The THRIVE score strongly predicts outcomes in patients treated with the Solitaire device in the SWIFT and STAR trials. *Int J Stroke.* (2014) 9:698–704. doi: 10.1111/ijs.12292

123. Chen WQ, Pan YS, Zhao XQ, Liao XL, Liu LP, Wang CJ, et al. Totaled health risks in vascular events score predicts clinical outcome and symptomatic intracranial hemorrhage in chinese patients after thrombolysis. *Stroke.* (2015) 46:864–6. doi: 10.1161/STROKEAHA.114.007979

124. Chen WQ, Liu GF, Fang JM, Wang YL, Song YL, Pan YS, et al. External validation of the totaled health risks in vascular events score to predict functional outcome and mortality in patients entered into the china national stroke registry. *J Stroke Cerebrovasc.* (2016) 25:2331–7. doi: 10.1016/j.jstrokecerebrovasdis.2016.03.021

125. Boehme AK, Rawal PV, Lyerly MJ, Albright KC, Bavarsad Shahripour R, Palazzo P, et al. Investigating the utility of previously developed prediction scores in acute ischemic stroke patients in the stroke belt. *J Stroke Cerebrovasc Dis.* (2014) 23:2001–6. doi: 10.1016/j.jstrokecerebrovasdis.2014.02.003

126. Fjetland L, Roy S, Kurz KD, Solbakken T, Larsen JP, Kurz MW. Neurointerventional treatment in acute stroke. Whom to treat? (Endovascular treatment for acute stroke: utility of THRIVE score and HIAT score for patient selection). *Cardiovasc Intervent Radiol.* (2013) 36:1241–6. doi: 10.1007/s00270-013-0636-9

127. Ishkanian AA, McCullough-Hicks ME, Appelboom G, Piazza MA, Hwang BY, Bruce SS, et al. Improving patient selection for endovascular treatment of acute cerebral ischemia: a review of the literature and an external validation of the Houston IAT and THRIVE predictive scoring systems. *Neurosurg Focus.* (2011) 30:E7. doi: 10.3171/2011.3.FOCUS1144

128. Kamel H, Patel N, Rao VA, Cullen SP, Faigeles BS, Smith WS, et al. The Totaled Health Risks in Vascular Events (THRIVE) Score Predicts Ischemic Stroke Outcomes Independent of Thrombolytic Therapy in the NINDS tPA Trial. *J Stroke Cerebrovasc.* (2013) 22:1111–6. doi: 10.1016/j.jstrokecerebrovasdis.2012.08.017

129. Kastrup A, Brunner F, Hildebrandt H, Roth C, Winterhalter M, Giessing C, et al. THRIVE score predicts clinical and radiological outcome after endovascular therapy or thrombolysis in patients with anterior circulation stroke in everyday clinical practice. *European Journal of Neurology.* (2017) 24:1032–9. doi: 10.1111/ene.13328

130. Kurre W, Aguilar-Perez M, Niehaus L, Fischer S, Schmid E, Bazner H, et al. Predictors of outcome after mechanical thrombectomy for anterior circulation large vessel occlusion in patients aged >/=80 years. *Cerebrovasc Dis.* (2013) 36:430–6. doi: 10.1159/000356186

131. Lei CY, Wu B, Liu M, Chen YC, Yang HL, Wang DR, et al. Totaled health risks in vascular events score predicts clinical outcomes in patients with cardioembolic and other subtypes of ischemic stroke. *Stroke.* (2014) 45:1689–94. doi: 10.1161/STROKEAHA.113.004352

132. You S, Han Q, Xiao G, Zhang C, Zhang X, Liu H, et al. The role of THRIVE score in prediction of outcomes of acute ischemic stroke patients with atrial fibrillation. *Zhonghua Nei Ke Za Zhi.* (2014) 53:532–6.

133. Flint AC, Rao VA, Chan SL, Cullen SP, Faigeles BS, Smith WS, et al. Improved ischemic stroke outcome prediction using model estimation of outcome

probability: the THRIVE-c Calculation. *Int J Stroke*. (2015) 10:815–21. doi: 10.1111/ijs.12529

134. Pan YS, Peng YJ, Chen WQ, Wang YJ, Lin Y, He Y, et al. THRIVE-c score predicts clinical outcomes in Chinese patients after thrombolysis. *Brain Behav*. (2018) 8:e00927. doi: 10.1002/brb3.927

135. *THRIVE Score for Stroke Outcome: MDCalc.* Available online at: https://www.mdcalc.com/thrive-score-stroke-outcome (Cited October 22, 2018).

136. Moons KGM, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *Brit Med J*. (2009) 338:b606. doi: 10.1136/bmj.b606

137. Ntaios G, Papavasileiou V, Michel P, Tatlisumak T, Stribian D. Predicting functional outcome and symptomatic intracranial haemorrhage in patients with acute ischemic stroke: a glimpse into the crystal ball. *Stroke*. (2015) 46:899–908. doi: 10.1161/STROKEAHA.114.003665

138. Fearon P, McArthur KS, Garrity K, Graham LJ, McGroarty G, Vincent S, et al. Prestroke modified rankin stroke scale has moderate interobserver reliability and validity in an acute stroke setting. *Stroke*. (2012) 43:3184–8. doi: 10.1161/STROKEAHA.112.670422

139. McArthur KS, Quinn TJ, Higgins P, Langhorne P. Post-acute care and secondary prevention after ischaemic stroke. *BMJ*. (2011) 342:d2083. doi: 10.1136/bmj.d2083

140. Ryu CW, Kim BM, Kim HG, Heo JH, Nam HS, Kim DJ, et al. Optimizing outcome prediction scores in patients undergoing endovascular thrombectomy for large vessel occlusions using collateral grade on computed tomography angiography. *Neurosurgery*. (2018) doi: 10.1093/neuros/nyy316. [Epub ahead of print].

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.