

BMJ Open Methodological quality of multivariate prognostic models for intracranial haemorrhages in intensive care units: a systematic review

Jeanne Simon-Pimmel ¹, Johann Foucher,^{1,2} Maxime Léger,^{1,3} Fanny Feuillet,^{1,4} Laetitia Bodet-Contentin,^{1,5} Raphaël Cinotti,⁶ Denis Frasca,^{1,7} Etienne Dantan ¹

To cite: Simon-Pimmel J, Foucher Y, Léger M, *et al.* Methodological quality of multivariate prognostic models for intracranial haemorrhages in intensive care units: a systematic review. *BMJ Open* 2021;**11**:e047279. doi:10.1136/bmjopen-2020-047279

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2020-047279>).

Received 24 November 2020
Accepted 25 July 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Jeanne Simon-Pimmel;
Jeanne.Pimmel@gmail.com

ABSTRACT

Objectives Patients with severe spontaneous intracranial haemorrhages, managed in intensive care units, face ethical issues regarding the difficulty of anticipating their recovery. Prognostic tools help clinicians in counselling patients and relatives and guide therapeutic decisions. We aimed to methodologically assess prognostic tools for functional outcomes in severe spontaneous intracranial haemorrhages.

Data sources Following Preferred Reporting Items for Systematic Reviews and Meta-Analyses recommendations, we conducted a systematic review querying Medline, Embase, Web of Science, and the Cochrane in January 2020.

Study selection We included development or validation of multivariate prognostic models for severe intracerebral or subarachnoid haemorrhage.

Data extraction We evaluated the articles following the CHECKlist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies and Transparent Reporting of multivariable prediction model for Individual Prognosis Or Diagnosis statements to assess the tools' methodological reporting.

Results Of the 6149 references retrieved, we identified 85 articles eligible. We discarded 43 articles due to the absence of prognostic performance or predictor selection. Among the 42 articles included, 22 did not validate models, 6 developed and validated models and 14 only externally validated models. When adding 11 articles comparing developed models to existing ones, 25 articles externally validated models. We identified methodological pitfalls, notably the lack of adequate validations or insufficient performance levels. We finally retained three scores predicting mortality and unfavourable outcomes: the IntraCerebral Haemorrhages (ICH) score and the max-ICH score for intracerebral haemorrhages, the SubArachnoid Haemorrhage International Trialists score for subarachnoid haemorrhages.

Conclusions Although prognostic studies on intracranial haemorrhages abound in the literature, they lack methodological robustness or show incomplete reporting. Rather than developing new scores, future authors should focus on externally validating and updating existing scores with large and recent cohorts.

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ This is the first systematic review of the methodological quality of prognostic tools for severe spontaneous intracranial haemorrhages managed in intensive care units.
- ⇒ A robust search strategy with no language restriction was performed, leading to a high number of eligible articles.
- ⇒ This systematic review follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement, and we evaluated the articles following the Transparent Reporting of multivariable prediction model for Individual Prognosis Or Diagnosis statement to assess the tools' methodological reporting and pitfalls.
- ⇒ This systematic review concerns two types of lesions intracerebral haemorrhages and subarachnoid haemorrhages that present different pathophysiologies and clinical courses but similar long-term consequences, leading us to suspect shared methodological issues.
- ⇒ We were not able to perform a meta-analysis due to the heterogeneity in the included models.

INTRODUCTION

Severe spontaneous intracranial haemorrhages, managed in intensive care units (ICUs), are at high risk of developing complications such as rebleeding or cerebral ischaemia,^{1 2} leading to high morbidity and mortality. Intracerebral haemorrhages (ICH) have a mortality rate of 40% at 1 month,³ while subarachnoid haemorrhages (SAH) have a mortality rate of 25% at 10 years.⁴ Survivors have a high rate of vegetative state or severe disabilities.⁵ This serious statement highlights the initial issues specific to severe strokes and the challenge physicians and surrogates face in deciding to continue invasive care.^{6 7} Indeed, the question arises as to whether advanced resuscitation is justified when the future appears unfavourable.⁸

When considering a limitation of care, the essential issue is to prevent inaccurate self-fulfilling prophecies by predicting outcomes reliably.⁹ In such settings, an individual's patient prognostic may be difficult to assess because of the multiplicity of risk factors involved in the evolution of severe intracranial haemorrhages. Multivariable prognostic scores could assist clinicians in counselling patients and relatives and guide therapeutic decisions.

Previous reviews of prognostic tools,^{10–14} popular in the field of neurocritical care, have not focused on injuries managed in ICUs, for whom the issue of advanced care pursuits is a concern. Indeed, scores are reliable when validated in the population of interest. They also did not address the methodological quality of the selected articles. The PROgnosis REsearch Strategy (PROGRESS) group recently proposed a framework for prognosis concerns^{15 16} that led to the Transparent Reporting of multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) statement.¹⁷ These recommendations efficiently summarised the process for developing and validating a prognostic scoring system.

The objective of our systematic review was to assess the methodology of existing prognostic tools of functional outcomes in patients with severe spontaneous intracranial haemorrhage managed in ICUs. We chose to conduct this systematic review for the two types of lesions (ICH and SAH). While their pathophysiologies and clinical courses are different, the consequences for long-term functional outcomes are similar. The questions that arise at the beginning of the ICU stay about patients' future and the complex ethical decisions are similar. While prognostic models may differ, the way to develop them should follow a similar modelling process. We suspected that studies presenting prognostic tools share the same methodological issues.

MATERIALS AND METHODS

Search strategy

This systematic review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses statement (online supplemental table S1).¹⁸ We searched Medline, Embase, Web of Science and the Cochrane databases on 7 December 2017 and updated on 14 January 2020, without date restriction. We used a query based on Medical Subject Heading terms and keywords. Online supplemental file S2 outlines the detailed search strategy.

Study selection

We included all-language studies focusing on adults with severe spontaneous intracranial haemorrhage (ICH or SAH) managed in ICU, or specified explicitly as 'severe' or 'high grade' injury. We did not include criteria on the location, the cause of the haemorrhage or the type of cases (primary or secondary haemorrhage). We did not include paediatric studies or studies uniquely concerning traumatic injuries. We searched for the development and/or validation of prognostic models, predicting outcomes

using variables collected before or at the beginning of their ICU stay. The targeted outcomes were mortality, functional outcomes or quality-of-life-related outcomes from ICU-discharge or hospital-discharge through to long-term outcomes. Our non-inclusion criteria were reviews or meta-analyses, full texts not found or conference abstracts, models developed without predictor selection, univariate models or the lack of reported prognostic performance. One reviewer (JS-P) screened references by title and abstract. The full eligible texts were assessed independently by four pairs of reviewers (YF-ML, FF-RC, DF-LB-C and JS-P-ED) and discussions resolved any discrepancies.

Data extraction

We predefined a standardised form for data extraction and evaluation of the risk of bias (online supplemental tables S3 and S4). For each eligible article, we collected the author's name, year and journal, data source and study design, inclusion and exclusion criteria, sample size, population characteristics, predicted outcomes (mortality, functional outcomes and quality-of-life), prediction time (ie, the time when one calculates the prediction), horizon time (ie, the end of the prediction time window), predictive tools, development details (such as variables of the scoring systems), internal validation details, external validation details, missing data information and open comments regarding bias and limitations.

Articles and prognostic tools selection based on quality assessment

To include the articles, we followed the CHecklist for critical Appraisal and data extraction for systematic Reviews of prediction Modelling Studies (CHARMS)¹⁹ and the TRIPOD statements.²⁰ Specifically, they recommend developing a score from a learning sample set and validating the prognostic performance from an independent sample (internal and/or external validation). This step avoids reporting the prognostic capacities on the training sample only because no internal nor external validation led to overestimating their performance.²¹ The articles reporting model development without any validation were thus not retained. They also recommend having a sufficient sample size and a sufficient number of events (known as the effective sample size). We considered at least 250 patients and 50/50 events and non-events as sufficient. The modelling strategy must also consider enough events per predictor, usually at least 10, to avoid overfitting.^{22 23} We did not include articles that did not follow these recommendations.

Assessment of the performance of the prognostic tools should use discrimination (ability to differentiate between patients who do or do not experience the event, eg, area under the curve (AUC) receiver operating characteristic (ROC)), calibration (agreement between predictions from the model and observed outcomes) and global measures (simultaneous evaluation of calibration and discrimination, eg, Brier Score). Among the

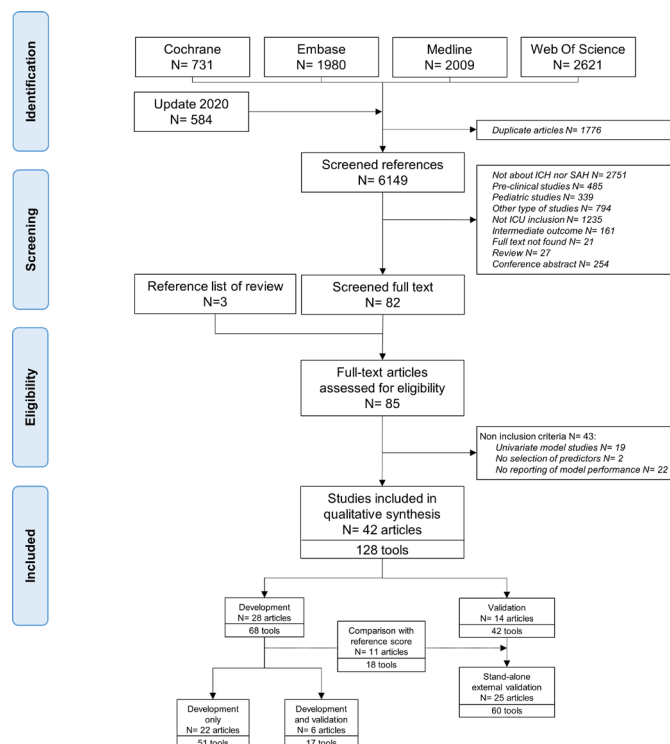


Figure 1 PRISMA flow diagram, selection of included studies. PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

included articles, the retained prognostic tools were those presenting good prognostic performances reported on internal and/or external validation.

Patient and public involvement

This study has no patient or public involvement.

RESULTS

Description of studies

The electronic database search identified 6149 unique references. Screening of titles/abstracts and references checking of included articles and reviews identified 85 eligible papers for full-text review. We did not include 43 articles for the following reasons: 19 univariate models, 2 models without predictor selection and 22 multivariate models without performance reporting. Finally, we included 42 articles (figure 1).

All articles were in English. There were 11 articles published before 2010, 12 between 2010 and 2015 and 19 after 2015. The published teams were mainly from Europe (n=17, 40%) and North America (n=14, 33%). Patients were mostly recruited into an ICU (n=33, 79%). Inclusion criteria were heterogeneous in terms of location or aetiology of the haemorrhage. For ICH, most studies included only spontaneous ICH, some excluding malformations and/or coagulation disorder. For SAH, most included aneurysmal SAH online supplemental tables S3 and S4 present the information regarding inclusion and non-inclusion criteria of each study. The pooled mean age was 59.3 years (SD 13.7) (data not available for

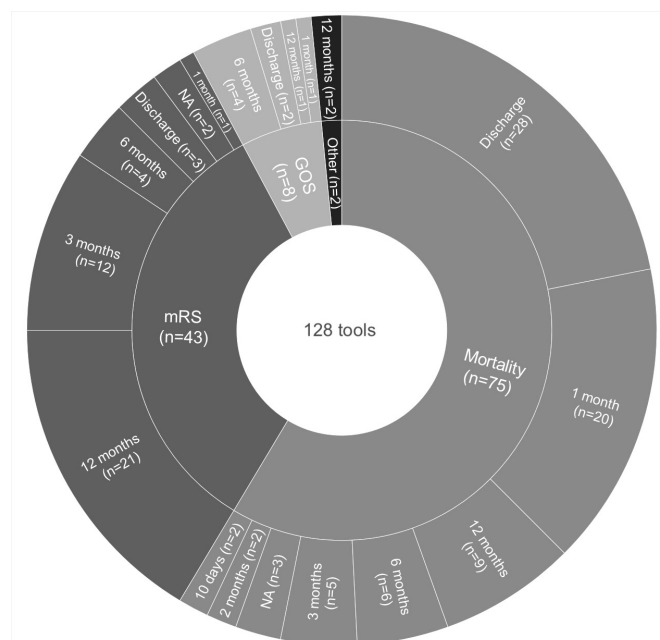


Figure 2 Predicted outcomes and corresponding horizon times of the 128 prognostic tools. GOS: Glasgow Outcome Scale; mRS: modified Rankin Scale; NA, not available.

six studies). Fifty-three per cent (range 21%–73%) were female (missing data for five studies). The 42 eligible articles reported 128 prognostic tools (figure 1): five articles reported one tool, 16 reported two tools, 7 reported three tools and 14 articles more than three tools, differing by their predictors, their types of outcome or their horizon times. Regardless of the types of predicted outcomes, the sample sizes ranged from 68 to 1629 patients (median 290, IQR 128–413), and the number of events ranged from 21 to 786 (median 64.5, IQR 34–164). Regardless of the time of prediction, most of the prognostic tools predicted mortality (n=75, 59%) (figure 2). Fifty-one (40%) tools studied functional outcomes using the modified Rankin Scale (mRS) or the Glasgow Outcome Scale (GOS). The horizon time for mortality data was mostly short-term (67% at discharge or 1 month), unlike functional outcomes (14% at discharge or 1 month) (figure 2). One study predicted the cognitive status and physical quality-of-life at 12 months. The 452 predictors of these 128 tools mainly involved baseline characteristics (n=95, 21%), admission clinical variables (n=104, 23%), biological measures (n=86, 19%), CT variables (n=95, 21%), ICU-evolution variables (n=29, 6%), existing scores (n=40, 9%) and others (n=3, 1%). Most variables were available on admission, others within 72 hours after ICU admission, and few were available throughout the ICU stay. The prediction time was sometimes unknown.

Model development studies

Twenty-eight studies developed prediction models. Online supplemental table S3 provides complete standardised form and references. Twelve articles focused on patients with ICH only, 15 on patients with SAH only and one on patients with both ICH and SAH. Of the 16 articles on

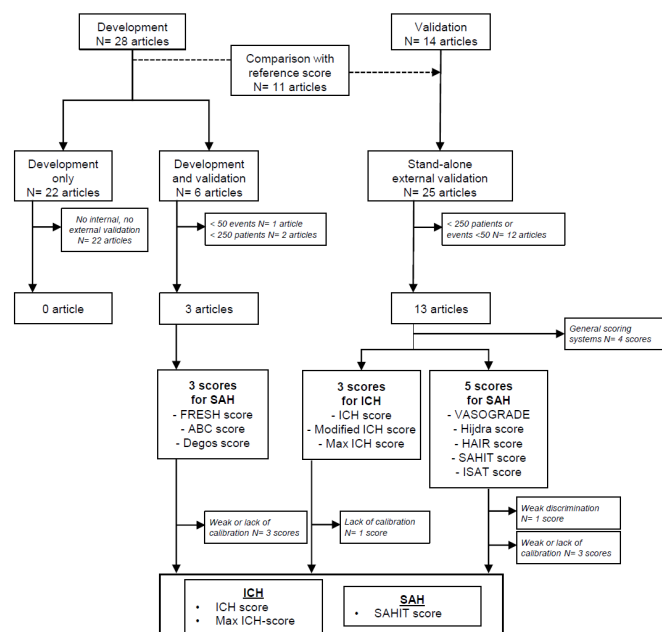


Figure 3 Retained prognostic scores. ICH, intracerebral haemorrhage; SAH, subarachnoid haemorrhages; SAHIT, SubArachnoid Haemorrhage International Trialists.

SAH, 14 (87%) reported a functional outcome, while they represented 6 (46%) of the 13 ICH articles. The primary statistical analysis used to develop the scoring system was logistic regression. Other analyses were linear models, Cox models or less well-known statistical methods such as decision tree analysis, Bayesian networks and artificial neural networks. One article did not specify the type of modelling used (see online supplemental table S3) for corresponding references). Predictor selection strategy, which describes the initial pool of variables and the analysed variables, was rarely mentioned.

Among the 28 included articles, 22 articles developed their tool without validation, that is, they reported the apparent prognostic capacities on the training sample only. They were thus not retained. However, few of these studies were well conducted, with a large cohort and long-term outcome and would deserve validations.^{24–27} Among the 28 included articles, six articles presented a development with internal validation (two using bootstrapping, three cross-validations, one temporal validation). One also reported additional external validation. Online supplemental table S5 lists the methods used to quantify prognostic performances. The authors seldom presented global performances. All reported the discrimination with the AUC of the ROC curve, while calibration measures were not systematic. Of the six studies that developed and validated models, two included fewer than 250 patients and one had less than 50 events. We did not retain them due to this insufficient sample size (figure 3).

Finally, three articles proposed a prognostic tool developed and validated based on recommendations: the FRESH score for SAH (excluding rupture of arteriovenous malformation),¹¹ the ABC score for patients with aneurysmal SAH²⁸ and the score by Degos *et al* for elderly

patients with aneurysmal SAH.²⁹ Table 1 summarises the collected information regarding source population, development approach, validation details and prognostic performances of these three retained scores.

External validation studies

Fourteen articles aimed to externally validate one or more existing models, most of which were not initially developed with severe injuries managed in ICUs. Eleven out of the 28 articles that developed a tool also compared their score to one or more existing models. Finally, 25 articles presented a stand-alone external validation. Online supplemental table S4 provides complete standardised form and references. Online supplemental table S5 lists the methods used to report prognostic performances. Most reported the AUC of the ROC curve; 15 articles had at least one calibration measurement. The authors rarely compared external validation cohorts to the population of the original article. One study proposed recalibration to predict another outcome than the development study.³⁰

Of the 25 studies that externally validated models, 12 included fewer than 250 patients or less than 50 events (figure 3). There were four externally validated general scoring systems. The APACHE II, the SIRS summary score, the SOFA score and the SAPS II showed encouraging performance values when predicting short-term mortality. Because they did not include specific predictors of brain injuries, their use in clinical practice to predict functional or long-term outcomes is not appropriate (figure 3). Injury-specific predictors could extend these scoring systems to improve their predictive capacities and clinical utilities.

There were eight injury-specific externally validated scores. In the ICH population, we retained three externally validated scores: the ICH score,^{25 31–33} the modified ICH score (MICH)^{25 34} and the max ICH score.^{25 33} For the SAH, we retained five tools. Two tools were bivariate, including Glasgow Coma Scale or World Federation of NeuroSurgeons (WFNS) scale associated with CT features: a three-coloured grading system termed the VASOGRADE^{35 36} and the Hijdra score for aneurysmal SAH.^{37 38} Three tools were multivariate models: the HAIR score^{11 36 39 40} and the SubArachnoid Haemorrhage International Trialists (SAHIT) score for SAH,^{41 42} and the international subarachnoid aneurysm trial score for aneurysmal SAH.^{43 44} Tables 2 and 3 summarise, for ICH and SAH, respectively, the collected information regarding the source population, development approach, validation details and prognostic performances of these eight scores.

Retained prognostic scores from included studies

Finally, for each included study (Development and validation or Stand-alone external validation), we reviewed the levels of prognostic performances for the final selection of multivariate prognostic scores that can be easily applicable for practical use. Among the prognostic tools for

Table 1 Extract of the complete standardised form for the retained developed and validated tools (SAH only)

Author, date, score name	Population	Population characteristics	Sample size, nb of events	Outcomes	Time of prediction, predictors	Discrimination	Calibration	Global performance	Strengths	Limitations
Witsch <i>et al</i> 2016 FRESH score ¹¹	Prospective US monocentric cohort (SHOP).	<p>► Age: mean 55y. (SD 14)</p> <p>► Female: 38%</p>	n=1526 Unfav. outcome: 79%	Unfav. outcome (mRS 4–6) at 12 m	48 hours after admission. FRESH: Hunt & Hess, APACHE w/o GCS, age, aneurysmal rebleed	AUC: 89.8 (88.1–91.6)	No	<p>► N-R²: 0.50</p> <p>► C/S-R²: 0.35</p>	<p>► Considering QOL outcome</p> <p>► Good discrimination</p>	<p>► Lack of calibration</p> <p>► Use of linear regression to model ordinal mRS</p>
Degos <i>et al</i> 2012 ABC-score ²⁸	Prospective French monocentric cohort.	<p>► Age: mean 50y. (SD 13)</p> <p>► Female: 64%</p>	Training n=368 Validation n=158 Mortality: 17% Training Validation 18%	<p>► Mortality at 12 m</p> <p>► Independent function (mRS 0–3) at 12 m</p> <p>► Full recovery (mRS 0–1) at 12 m</p>	On admission. Troponin-I, S100B, GCS	AUC: 0.76 (0.67–0.85) ► mRS 0–3 0.76 (0.67–0.86) ► mRS 0–1 0.83 (0.79–0.88)	HL-GOF: NA	No	<p>► Two different cut-offs for mRS</p> <p>► High discrimination</p>	<p>► Use of HL-GOF test</p>
Degos <i>et al</i> 2012 ²⁸	Prospective French monocentric cohort.	<p>► Age: <60 y.: 708 60–70 y.: 138 ≥70 y.: 87</p> <p>► Female: 62%</p>	n=933 Unfav. outcome: 19%	Unfavourable outcome (mRS 4–6) at 12 m	Unknown time of prediction. ► Without interaction: IH on admission, severe IH, isch. vasospasm, rebleeding, endovascular complication, surgery complication, Fisher III–V, hydrocephalus, age >60y With interaction: above + hydrocephalus * age >60y	AUC: 0.84 (0.82–0.88) ► Without interaction: 0.85 (0.82–0.88) ► With interaction: 0.85 (0.82–0.88)	<p>► Without interaction: p=0.18</p> <p>► With interaction: p=0.22</p>	No	<p>► High discrimination</p>	<p>► Use of HL-GOF test</p> <p>► Unknown time of collection of predictors (endovascular or surgery complications, hydrocephalus)</p>

AP, arterial pressure; AUC, area under the receiving operative curve; AVM, arterio-vascular malformation; compl., complication; GCS, Glasgow Coma Scale; HL-GOF, Hosmer-Lemeshow-Goodness-Of-Fit test; HR, heart rate; ICH, intracerebral haemorrhage; ICU, intensive care unit; IH, intracranial hypertension; isch, ischaemic; m, months; mRS, modified Rankin-Scale; NA, not available; nb, number; NICU, NeuroICU; phys, physical; QOL, quality-of-life; SAH, subarachnoid haemorrhage; TBI, traumatic brain injury; Unfav., unfavourable; y, years.

Table 2 Extract of the complete standardised form for the retained externally validated tools for ICH

Initial paper (author, date), score name	Initial study design	Initial outcomes	Initial predictors	Paper (external validations), date	Design, sample size	Outcomes, nb of events	Discrimination	Calibration	Global performance	Strengths	Limitations
Hemphill <i>et al</i> 2001 ³¹ ICH score	American bicentric cohort	Mortality at 1 m	Age, GCS, ICH volume, IVH, Infratentorial origin	Rodriguez-Fernandez <i>et al</i> 2018 ³²	Prospective multicentre cohort, Spain, 2009–2012 n=336	Mortality at 1 m (52%)	AUC 0.74 (0.69–0.79)	▲ HL-GOF p<0.001 ▲ GiVTI belt p<0.001	No	▲ Good discrimination	▲ Bad calibration ▲ Short-term outcome
				Schmidt <i>et al</i> 2018 ³³	Prospective US monocentric cohort, 2010–2017 n=372	▲ Mortality at 3 m (41%) ▲ Unfav. outcome (mRS 4–6) at 3 m (63%)	AUC ▲ 3 m mortality 0.83 (0.79–0.88) ▲ 3 m unfav. outcome 0.85 (0.81–0.89)	▲ Histogram pred. vs obs. ▲ HL-GOF p>0.3	▲ Good discrimination ▲ Functional outcome	▲ Short-term outcome	
				Sembill <i>et al</i> 2017 ²⁵	Prospective monocentric German cohort, 2007–2011 n=471	▲ Mortality at (3 and) 12 m (30.1%) ▲ Unfav. outcome (mRS 4–6) at (3 and) 12 m (45.4%)	AUC ▲ 12 m mortality 0.69 (0.64–0.74) ▲ 12 m unfav. outcome 0.72 (0.67–0.76)	Histogram pred. vs obs.	No	▲ Long-term outcome ▲ Functional outcome	▲ Weaker discrimination
Cho <i>et al</i> 2008 ³⁴ MICH	Taiwanese monocentric cohort	▲ Mortality at 6 m ▲ Unfav. outcome (GOS 4–5) at 12 m ▲ Barthel index (≥55) at 12 m	GCS, ICH volume, IVH or hydrocephalus	Sembill <i>et al</i> 2017 ²⁵	Prospective monocentric German cohort, 2007–2011 n=471	▲ Mortality at (3 and) 12 m (30.1%) ▲ Unfav. outcome (mRS 4–6) at (3 and) 12 m (45.4%)	AUC ▲ 12 m mortality: 0.6 ▲ 12 m unfav. outcome: 0.69	No	No	Long-term functional outcome	▲ Weak discrimination ▲ No calibration
Sembill <i>et al</i> 2017 ²⁵ Max ICH score	Germanic monocentric cohort	▲ Mortality at 3 and 12 m ▲ Unfav. outcome (mRS 4–6) at 3 and 12 m	Lobar ICH volume, non-lobar ICH volume, age, NIHSS, IVH, oral anticoag	Schmidt <i>et al</i> 2018 ³³	Prospective US monocentric cohort, 2010–2017 n=372	▲ Mortality at 3 m (41%) ▲ Unfav. outcome (mRS 4–6) at 3 m (63%)	AUC ▲ 3 m mortality: 0.82 (0.78–0.86) ▲ 3 m unfav. outcome: 0.88 (0.85–0.92)	▲ Histogram pred. vs obs. ▲ HL-GOF p>0.3	▲ Likelihood ratio χ^2 p<0.001	▲ Functional outcome ▲ Good discrimination and calibration	Short-term outcome for external validation
anticoag, anticoagulant therapy; AUC, Area Under the Receiving Operative Curve; GCS, Glasgow Coma Scale; HL-GOF, Hosmer-Lemeshow-Goodness-Of-Fit test; ICH, Intracranial Haemorrhage; IVH, Intraventricular haemorrhage; m, months; MICH, modified ICH; mRS, modified Rankin-Scale; NIHSS, National Institutes of Health Stroke Scale; obs, observed; pred, predicted; unfav, unfavourable.											

anticoag, anticoagulant therapy; AUC, Area Under the Receiving Operating Curve; GCS, Glasgow Coma Scale; HL-GOF, Hosmer-Lemeshow-Goodness-Of-Fit test; ICH, Intracranial Haemorrhage; IVH, intraventricular haemorrhage; m, months; MICH, modified Rankin-Scale; mRS, modified Rankin-Scale; NIHSS, National Institutes of Health Stroke Scale; obs, observed; pred, predicted; unfav, unfavourable.

Table 3 Extract of the complete standardised form for the retained externally validated tools for SAH

Initial paper (author, date), score name	Initial study design	Initial outcomes	Initial predictors	Paper (external validations), date	Design, sample size	Outcomes, nb of events	Discrimination	Calibration	Global performance	Strengths	Limitations
De Oliveira et al 2015 ³⁵ VASOGRADE	3 SAHIT trials (CONSCIOUS-1, EPO trial, statin trial)+1 Canadian centre	DCI	WFNS, modified Fisher Scale	Dengler et al 2017 ³⁶	German hosp. registry, monocentric, 2009–2015, n=423	<ul style="list-style-type: none"> Unfav. outcome (mRS 3–6) at 12 m (53.1%) Unfav. outcome (mRS 4–6) at 12 m 	<ul style="list-style-type: none"> AUC mRS 3–6: 0.711 mRS 4–6: 0.709 	<ul style="list-style-type: none"> Histogram pred. vs obs. 	<ul style="list-style-type: none"> No 	<ul style="list-style-type: none"> Good discrimination Choice of a functional outcome for EV 	<ul style="list-style-type: none"> Weak calibration for 2/3 of patients due to a three-level ordinal scale with two predictors
Hijdra et al 1988 ³⁷ Hijdra score	NA	<ul style="list-style-type: none"> Unfav. outcome (GOS 1 to 2) at 1 m DCI Rebleeding 	SAH volume, GCS	Claassen et al 2004 ³⁸	Prospective US monocentric cohort, 1996–2002, n=413	Unfav. Outcome (mRS 4–6) at 3 m (40.4%)	<ul style="list-style-type: none"> AUC 0.67 (0.61–0.73) 	No	No	<ul style="list-style-type: none"> Weaker discrimination 	<ul style="list-style-type: none"> No calibration
Risselada et al 2010 ⁴³ ISAT	ISAT European cohorts	Mortality at 2 m	Age, lumen size, Fisher grade, WFNS	Dijkland et al 2016 ⁴⁴	Dutch hosp. registry, Monocentric, 2007–2011, n=307	Mortality at 2 m (30.6%)	<ul style="list-style-type: none"> AUC WFNS at ttt: 0.89 WFNS on adm: 0.82 	<ul style="list-style-type: none"> Calibration curve Intercept at ttt: 2.248 on adm: 1.502 Slope at ttt: 1.417 on adm: 1.959 	No	<ul style="list-style-type: none"> Good discrimination Calibration curve 	<ul style="list-style-type: none"> Low calibration for high-risk SAH Choice of the outcome (short-term mortality)
Jaja et al 2018 ⁴¹ SAHIT	SAHIT (nine international trials and registries)	<ul style="list-style-type: none"> Mortality at 3 m Unfav. outcome (GOS 1 to 3) at 3 m 	<ul style="list-style-type: none"> Core: age, premorbid history of HT, WFNS NI: core +CT vol of SAH, aneurysm size, aneurysm location Full: NI +ttt modality 	Mascitelli et al 2018 ⁴²	US trial cohort, Monocentric, 2003–2007, n=338	<ul style="list-style-type: none"> Unfav. outcome (mRS 3–6) at 6 m (29.6%) Mortality at 6 m (10.1%) 	<ul style="list-style-type: none"> AUC Unfav. outcome: core: 72.8 (66.8–78.9) NI: 73.2 (67.1–79.2) Full: 73.4 (67.5–79.4) Mortality: core: 72.1 (62.1–82.2) NI: 73.9 (64.4–83.5) Full: 74.4 (65.1–83.8) 	<ul style="list-style-type: none"> Calibration curve Intercept/slope 	<ul style="list-style-type: none"> R² Brier score Brier scale 	<ul style="list-style-type: none"> Extended choice of predictors Good discrimination Good calibration curves Reporting of global performances 	
Lee et al 2014 ³⁹ HAIR	American monocentric cohort	In-hosp. mortality	Hunt & Hess, age, IVH, rebleeding within 24 hours	Witsch et al 2016 ⁴¹	Cohort SHOP: Prospective US monocentric cohort, 1996–2014, n=1526 CONSCIOUS-1: multicentric Israel, Europe, North America, n=413	<ul style="list-style-type: none"> Unfav. outcome (mRS 4–6) at 12 m (79% SHOP) CONSCIOUS-1 (NA CONSCIOUS-1) 	<ul style="list-style-type: none"> AUC SHOP: 88.3 (86.4–90.2) CONSCIOUS-1: 71.8 (66.0–77.5) 	No	<ul style="list-style-type: none"> SHOP: N-R²: 0.45 C/S-R²: 0.32 CONSCIOUS-1: N-R²: 0.17 C/S-R²: 0.11 	<ul style="list-style-type: none"> Long-term functional outcome Good discrimination 	<ul style="list-style-type: none"> Weaker perf. on CONSCIOUS-1 No calibration
				Dengler et al. 2017 ³⁶	Hosp. registry Germany, monocentric 2009–2015, n=423	<ul style="list-style-type: none"> Unfav. outcome (mRS 3–6) at 12 m (53.1%) Unfav. outcome (mRS 4–6) at 12 months 	<ul style="list-style-type: none"> AUC mRS 3–6: 0.739 mRS 4–6: 0.737 	<ul style="list-style-type: none"> Histogram pred. vs obs. 	No	<ul style="list-style-type: none"> Long-term functional outcome Good discrimination 	<ul style="list-style-type: none"> Weak calibration for high-risk SAH
				Abulhasan et al 2017 ⁴⁰	Canadian retrospective monocentric cohort, 2010–2016, n=434	Mortality at discharge (14.1%)	AUC: 0.89	<ul style="list-style-type: none"> Calibration curve Intercept: -0.05 Slope: 0.77 	No	<ul style="list-style-type: none"> Excellent discrimination Calibration curve 	<ul style="list-style-type: none"> Choice of outcome (short-term mortality)

adm, admission; AUC, Area Under the Receiving Operating Curve; C/S-R², Cox/Snell R²; DCI, Delayed Cerebral Ischaemia; GCS, Glasgow Coma Scale; GOS, Glasgow Outcome Scale; Hosp, hospital; HT, Hypertension; ISAT, international subarachnoid aneurysm trial; IVH, intraventricular haemorrhage; m, months; mRS, modified Rankin Scale; NA, Not available; NI, Neuroimaging; N-R², Nagelkerke R²; obs, observed; pred, predicted; SAH, Subarachnoid Haemorrhage; SAHIT, Subarachnoid Haemorrhage International Trials; ttt, treatment; unfav, unfavourable; vol, volume; WFNS, World Federation of Neurosurgeons.

ICH, we did not retain the MICH score because of the lack of reporting calibration that did not guarantee agreement between predictions and observed outcomes. We thus highlighted two scores (figure 3). The ICH score³¹ was externally validated in three large ICU cohorts, predicting 1-month, 3-month and 12-month mortality or functional outcome (mRS 4–6).^{25 32 33} The max ICH score²⁵ predicted 3-month and 12-month mortality and functional outcome (mRS 4–6), based on CT predictors (lobar and non-lobar ICH volume, age, National Institutes of Health Stroke Scale, presence of intraventricular haemorrhage and anticoagulant therapy). This showed good performances in a large external ICU cohort.³³ Table 2 presents the original publication, the external validation studies and corresponding performances (discrimination and calibration).

Among the retained SAH tools, the level of clinical utility and prognostic capacities was debatable. Tables 1 and 3 detail the strengths and limitations of each of these scores. The vast majority of tools presented high discrimination. We did not retain the Hjdra score³⁷ because of weak discrimination or absence of calibration. Additionally, the VASOGRADE,³⁵ the FRESH score,¹¹ the ABC score²⁸ and the Degos score for the elderly²⁹ lacked reporting calibration or used the Hosmer-Lemeshow goodness-of-fit test. The ISAT score and HAIR score, which had a low calibration for high-risk SAH, would probably benefit from recalibration or updating.^{39 43} We thus only retained the SAHIT score⁴¹ (figure 3). In a single external validation,⁴² it predicted either an unfavourable outcome (mRS 3–6) or mortality at 6 months, based on clinical predictors (age, history of hypertension and WFNS preoperative neurological grade) and CT (Fisher grade, aneurysm size and location). It revealed good discrimination and calibration.

DISCUSSION

While studies labelled as ‘prognostic’ abound in the literature on intracranial haemorrhage, our systematic review dedicated explicitly to critical patients revealed a lack of methodological robustness. Of the 85 read articles, we identified six articles that developed a prognostic tool supported by a validation study and 25 external validation studies. After critical appraisal of the articles, we retained, for the ICH population, the ICH score,³¹ which has better performances for the shorter outcome, and the max ICH score.²⁵ For the SAH population, we retained the SAHIT score for its high methodological quality.⁴²

The ICH score,³¹ developed in 2001, has benefited from multiple external validations in many different populations. The American Heart Association guidelines⁴⁵ recommend its reporting. In external validations with severe ICH, its performances could be better, particularly for longer term and functional outcomes.^{25 32 33} It would be interesting to consider updating or recalibrating this tool. The max ICH score,²⁵ developed in 2017, showed good calibration and discrimination on only one external

cohort, with satisfying calibration and better performances than the ICH score on the same sample.³³ It would benefit from further validations in other large and recent cohorts. The SAHIT score, developed in 2018, predicted unfavourable outcome or mortality at 3 months in a low to severe SAH population.⁴¹ The single external validation in an ICU cohort revealed good prognostic performances that further studies have yet to be confirmed.⁴²

In our systematic review, the authors rarely highlighted the clinical objective, which leads us to believe that clinical purposes did not drive most score elaborations. Functional outcomes in the modern setting of critical care make more sense than mortality outcomes for patients who are more likely to survive but face disabilities.⁴⁶ The ordinal functional outcomes scales are almost systematically dichotomised (GOS 1–3 vs 4–5, mRS 4–6 or 3–6 vs 0–3 or 0–4). These thresholds, though never justified, should depend on the clinical objective. If the score’s purpose is to support clinicians in making ethically challenging decisions, such as withdrawal of care, it is not reasonable to place severe disabilities, vegetative state and death on the same unfavourable side. Besides, a prognostic tool on its own, as rigorous as it may be, is hardly capable of integrating the strong human dimension of such a complex decision. Multidisciplinary clinical teams should rely on a combination of considerations, which include multi-variable scoring systems. If the clinical objective is instead to inform patients and their relatives of the evolution prospects, the condition they consider to be favourable should be determined by themselves and ideally over the very long-term.^{47 48} In our systematic review, the longest prediction horizon was 12 months, that is, before stabilisation of functional recovery and the ability to adapt to such a consolidated statement.⁴⁹ Moreover, patient perception could weigh the different levels of functional disabilities.^{50 51} Indeed, survivors have a wide range of life-long consequences such as neuropsychological difficulties, memory problems, fatigue and physical complaints, that is, dimensions not explored with functional outcome scales.^{51 52} As these symptoms are not always apparent, only validated patient (or caregivers) reported questionnaires can reflect the subjective perception of their quality of life.^{51 53} In our systematic review, the only article mentioning quality of life concerns the FRESH score.¹¹ Even though some methodological choices are questionable in this study, we think that it deserves attention because it surpasses the functional outcomes by integrating the quality of life as an objective of prediction.

In our systematic review, we identified several methodological pitfalls. A large proportion of eligible studies are wrongly labelled ‘prognostic models’. Some authors did not report prognostic performances, sometimes because they wrongly interpreted the odds ratio as a prognostic ability. These mistakes revealed considerable confusion in the literature between the notions of correlation and prediction.⁵⁴ Some development studies only reported apparent prognostic performances. This lack of internal or external validation led to overestimating the

performances of the prognostic tools.²¹ Several studies based on small sample size or a small number of events resulted in the risk of overfitting or low credibility in terms of prognostic performances.⁵⁵ These studies would benefit from external validations with recent and large cohorts. There was heterogeneity in the prognostic performances' reports: discrimination was systematic, only about half of the retained studies assessed calibration and 10% global performance. Calibration curves, rarely reported, allow future external validation to assess the eventual need for recalibration or updating, to adapt it to the population of interest. The popular Hosmer-Lemeshow Goodness-of-fit test is known to perform poorly, making its use regrettable.⁵⁶ We discarded several studies due to the absence of variables selection. The included studies rarely specified the predictor selection strategy, which describes the initial pool of variables and the analysed variables. This precision allows the reader to assess the risk of overfitting. The prediction time was sometimes unknown, making the score challenging to apply. Authors should clearly state this information to inform the user of when to calculate the prediction. Authors who studied long-term outcomes always chose to use logistic regression by excluding patients lost to follow-up when time-to-event methods would have been more appropriate in the presence of such censoring. Finally, this resulted in a very low number of prognostic tools that seemed methodologically correct and presenting a reasonable prognostic performance level. However, weaker validation results do not mean that the model is incorrect. If scores' development approaches were optimal, relevant predictors could be recalibrated and combined with new data to validate a strong tool.^{57 58}

A consortium of experts published the TRIPOD statement in 2015, clearly setting out how to report prognostic information.¹⁷ Of the 85 full texts screened in our review, 35 (41%) were printed after the TRIPOD publication in 2015. Of these, we finally retained 19 (54%) articles published after the TRIPOD publication, whereas we retained only 11 (22%) from the 50 articles published before the TRIPOD publication. Similar to Zamanipour Najafabadi *et al*, we noticed a trend towards quality improvement, reinforced with the necessary ongoing validation of existing scores.⁵⁹ Our systematic review revealed that some robust published scores, outlined in reviews focusing on non-severe to severe intracranial haemorrhages¹⁰⁻¹³ (eg, FUNC score,⁶⁰ Essen ICH score⁶¹ and ICHOP score⁶²), have not yet been validated in the ICU population. To use them reliably in such settings, they should be externally validated with critical patients. We also did not find tools dedicated to severe specific populations (such as haemorrhages secondary to malformation or patients with coagulation disorders). External validations would be interesting for these populations (eg, patients under anticoagulant⁶³ or arteriovenous malformation⁶⁴). Another option would be to extend existing scores with these risk predictors, such as the Max ICH-score, which includes the variable 'presence of

oral anticoagulant'.²⁵ With the rapid evolution of therapeutic advances in neurocritical care, the ongoing prognostic studies should focus on temporal validation and updating/recalibrating existing good scores to ensure their performance validity.⁵⁵ It is also possible to extend this by incorporating additional modern variables.

This review has several limitations. First, we aimed to include tools dedicated to ICH or SAH managed in the ICU. Because of the lack of severity classification for these pathologies, and heterogeneity of patients admitted to ICUs, we defined our proper severity criteria, which is debatable. Second, only one assessor conducted the study screening on title/abstract. This may have resulted in some missing eligible studies. Third, we did not use a formal tool to study the risk of bias such as the recent Prediction model Risk Of Bias ASsessment Tool (PROBAST) based on the TRIPOD.^{65 66} Following the TRIPOD recommendations, we built our own standardised form collecting similar information than the PROBAST items. Fourth, due to the heterogeneity in the included models, we could not to perform a meta-analysis. Finally, as with any systematic review, our work underwent publication bias issue. Similar to randomised clinical trials, we cannot exclude that unpublished studies may have negative results or size effects different from published studies.⁶⁷ One consequence could be, for instance, the underrepresentation of external validation studies with non-confirmatory prognostic performances.

CONCLUSIONS

Our review identified several methodological pitfalls and incomplete reporting in prognostic articles on intracranial haemorrhages managed in ICU. Among the many published scores for ICH and SAH, some deserve further attention. Rather than developing new scores, future authors should focus on externally validating and updating well-developed existing scores with large and recent cohorts, relying on methodological syntheses such as the TRIPOD statement.^{17 57 68} We have chosen to emphasise the ICH score, the max ICH score and the SAHIT scores for their superior prognostic performances. Nevertheless, they need ongoing validations, recalibrations and impact studies to improve them. The use of 'patient-centred' outcomes that have yet to be defined could also enhance the tools in the delicate, medical and ethical setting of critical care. Beyond all methodological issues, patient-centred clinical finality should guide prognostic tools to be convincing.

Author affiliations

¹UMR 1246 Methods in Patients-Centered Outcomes and Health Research, INSERM, Nantes, France

²Biostatistician, University Hospital Centre Nantes, Nantes, Pays de la Loire, France

³Medical Intensive Care, Angers University Hospital, Nantes, France

⁴Biostatistics and Methodology Unit, University Hospital Centre Nantes, Nantes, Pays de la Loire, France

⁵Intensive Care Unit, Regional University Hospital Centre Tours, Tours, Centre, France

⁶Anaesthesia and Intensive Care Unit, Hôpital Laennec, Saint-Herblain, University Hospital of Nantes, France, Université de Nantes, CHU Nantes, Saint-Herblain, France

⁷Anesthesia and Critical Care Department, University Hospital Centre Poitiers, Poitiers, France

Twitter Jeanne Simon-Pimmel @PimmelSimon

Contributors JS-P: conceptualisation and design of the review, data acquisition, statistical analysis, writing: original draft, review and editing. YF: conceptualisation, data acquisition, providing statistical expertise, writing: original draft, review and editing. FF: data acquisition, writing: review and editing. ML, LB-C, RC and DF: data acquisition, providing clinical expertise, writing: review and editing. ED: project management, conceptualisation, methodology, data acquisition, writing: original draft, review and editing.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests YF reports personal fees for statistical training from Sanofi and Biogen, outside the submitted work.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. Raw review data are available by reasonable request to the corresponding author.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Jeanne Simon-Pimmel <http://orcid.org/0000-0003-0699-5793>

Etienne Dantan <http://orcid.org/0000-0001-7137-5051>

REFERENCES

- van Gijn J, Kerr RS, Rinkel GJE. Subarachnoid haemorrhage. *Lancet* 2007;369:306–18.
- Balami JS, Buchan AM. Complications of intracerebral haemorrhage. *Lancet Neurol* 2012;11:101–18.
- van Asch CJ, Luitse MJ, Rinkel GJ, et al. Incidence, case fatality, and functional outcome of intracerebral haemorrhage over time, according to age, sex, and ethnic origin: a systematic review and meta-analysis. *Lancet Neurol* 2010;9:167–76.
- Rinkel GJE, Algra A. Long-Term outcomes of patients with aneurysmal subarachnoid haemorrhage. *Lancet Neurol* 2011;10:349–56.
- Katrak PH, Black D, Peeva V. Do stroke patients with intracerebral hemorrhage have a better functional outcome than patients with cerebral infarction? *Pm R* 2009;1:427–33.
- Hamel MB, Goldman L, Teno J. Identification of comatose patients at high risk for death or severe disability. *JAMA* 1995;273:1842–8.
- Pratt AK, Chang JJ, Sederstrom NO. A fate worse than death: prognostication of devastating brain injury. *Crit Care Med* 2019;47:591–8.
- Gillett GR, Honeybul S, Ho KM, et al. Neurotrauma and the rub: where tragedy meets ethics and science. *J Med Ethics* 2010;36:727–30.
- Jain A, Jain M, Bellolio MF, et al. Is early DNR a self-fulfilling prophecy for patients with spontaneous intracerebral hemorrhage? *Neurocrit Care* 2013;19:342–6.
- Jaja BNR, Cusimano MD, Etminan N, et al. Clinical prediction models for aneurysmal subarachnoid hemorrhage: a systematic review. *Neurocrit Care* 2013;18:143–53.
- Witsch J, Frey H-P, Patel S, et al. Prognostication of long-term outcomes after subarachnoid hemorrhage: the fresh score. *Ann Neurol* 2016;80:46–58.
- Sreekrishnan A, Dearborn JL, Greer DM, et al. Intracerebral hemorrhage location and functional outcomes of patients: a systematic literature review and meta-analysis. *Neurocrit Care* 2016;25:384–91.
- Satopää J, Mustanoja S, Meretoja A, et al. Comparison of all 19 published prognostic scores for intracerebral hemorrhage. *J Neurol Sci* 2017;379:103–8.
- Gregório T, Pipa S, Cavaleiro P, et al. Prognostic models for intracerebral hemorrhage: systematic review and meta-analysis. *BMC Med Res Methodol* 2018;18:145.
- Hemingway H, Croft P, Perel P, et al. Prognosis research strategy (progress) 1: a framework for researching clinical outcomes. *BMJ* 2013;346:e5595.
- Steyerberg EW, Moons KGM, van der Windt DA, et al. Prognosis research strategy (progress) 3: prognostic model research. *PLoS Med* 2013;10:e1001381.
- Collins GS, Reitsma JB, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 2015;162:55–63.
- Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Ann Intern Med* 2009;151:264–9.
- Moons KGM, de Groot JAH, Bouwmeester W, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the charms checklist. *PLoS Med* 2014;11:e1001744.
- Moons KGM, Altman DG, Reitsma JB, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): explanation and elaboration. *Ann Intern Med* 2015;162:W1–73.
- Steyerberg EW, Bleeker SE, Moll HA, et al. Internal and external validation of predictive models: a simulation study of bias and precision in small samples. *J Clin Epidemiol* 2003;56:441–7.
- Collins GS, Ogundimu EO, Altman DG. Sample size considerations for the external validation of a multivariable prognostic model: a resampling study. *Stat Med* 2016;35:214–26.
- 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.
- Fallenius M, Skrifvars MB, Reinikainen M, et al. Spontaneous intracerebral hemorrhage. *Stroke* 2019;50:2336–43.
- Sembill JA, Gerner ST, Volbers B, et al. Severity assessment in maximally treated ICH patients: the max-ICH score. *Neurology* 2017;89:423–31.
- Weimer JM, Nowacki AS, Frontera JA. Withdrawal of life-sustaining therapy in patients with intracranial hemorrhage: self-fulfilling prophecy or accurate prediction of outcome? *Crit Care Med* 2016;44:1161–72.
- Zheng K, Zhong M, Zhao B, et al. Poor-Grade aneurysmal subarachnoid hemorrhage: risk factors affecting clinical outcomes in intracranial aneurysm patients in a multi-center study. *Front Neurol* 2019;10:123.
- Degos V, Apfel CC, Sanchez P, et al. An admission bioclinical score to predict 1-year outcomes in patients undergoing aneurysm coiling. *Stroke* 2012;43:1253–9.
- Degos V, Gourraud P-A, Tursis VT, et al. Elderly age as a prognostic marker of 1-year poor outcome for subarachnoid hemorrhage patients through its interaction with admission hydrocephalus. *Anesthesiology* 2012;117:1289–99.
- Fallenius M, Skrifvars MB, Reinikainen M, et al. Common intensive care scoring systems do not outperform age and Glasgow coma scale score in predicting mid-term mortality in patients with spontaneous intracerebral hemorrhage treated in the intensive care unit. *Scand J Trauma Resusc Emerg Med* 2017;25:102.
- Hemphill JC, Bonovich DC, Besmertis L, et al. The ICH score: a simple, reliable grading scale for intracerebral hemorrhage. *Stroke* 2001;32:891–7.
- Rodríguez-Fernández S, Castillo-Lorente E, Guerrero-Lopez F, et al. Validation of the ICH score in patients with spontaneous intracerebral hemorrhage admitted to the intensive care unit in southern Spain. *BMJ Open* 2018;8:e021719.
- Schmidt FA, Liotta EM, Prabhakaran S, et al. Assessment and comparison of the max-ICH score and ICH score by external validation. *Neurology* 2018;91:e939–46.

- 34 Cho D-Y, Chen C-C, Lee W-Y, *et al.* A new Modified Intracerebral Hemorrhage score for treatment decisions in basal ganglia hemorrhage--a randomized trial. *Crit Care Med* 2008;36:2151-6.
- 35 de Oliveira Manoel AL, Jaja BN, Germans MR, *et al.* The VASOGRADE: a simple grading scale for prediction of delayed cerebral ischemia after subarachnoid hemorrhage. *Stroke* 2015;46:1826-31.
- 36 Dengler NF, Sommerfeld J, Diesing D, *et al.* Prediction of cerebral infarction and patient outcome in aneurysmal subarachnoid hemorrhage: comparison of new and established radiographic, clinical and combined scores. *Eur J Neurol* 2018;25:111-9.
- 37 Hijdra A, van Gijn J, Nagelkerke NJ, *et al.* Prediction of delayed cerebral ischemia, rebleeding, and outcome after aneurysmal subarachnoid hemorrhage. *Stroke* 1988;19:1250-6.
- 38 Claassen J, Vu A, Kreiter KT, *et al.* Effect of acute physiologic derangements on outcome after subarachnoid hemorrhage. *Crit Care Med* 2004;32:832-8.
- 39 Lee VH, Ouyang B, John S, *et al.* Risk stratification for the in-hospital mortality in subarachnoid hemorrhage: the hair score. *Neurocrit Care* 2014;21:14-19.
- 40 Abulhasan YB, Alabdulraheem N, Simoneau G, *et al.* Mortality after spontaneous subarachnoid hemorrhage: causality and validation of a prediction model. *World Neurosurg* 2018;112:e799-811.
- 41 Jaja BNR, Saposnik G, Lingsma HF, *et al.* Development and validation of outcome prediction models for aneurysmal subarachnoid haemorrhage: the SAHIT multinational cohort study. *BMJ* 2018;360:j5745.
- 42 Mascitelli JR, Cole T, Yoon S, *et al.* External validation of the subarachnoid hemorrhage international Trialists (SAHIT) predictive model using the Barrow ruptured aneurysm trial (Brat) cohort. *Neurosurgery* 2020;86:101-6.
- 43 Risselada R, Lingsma HF, Bauer-Mehren A, *et al.* Prediction of 60 day case-fatality after aneurysmal subarachnoid haemorrhage: results from the International subarachnoid aneurysm trial (ISAT). *Eur J Epidemiol* 2010;25:261-6.
- 44 Dijkland SA, Roozenbeek B, Brouwer PA, *et al.* Prediction of 60-Day case fatality after aneurysmal subarachnoid hemorrhage: external validation of a prediction model. *Crit Care Med* 2016;44:1523-9.
- 45 Hemphill JC, Greenberg SM, Anderson CS, *et al.* Guidelines for the management of spontaneous intracerebral hemorrhage: a guideline for healthcare professionals from the American heart Association/ American stroke association. *Stroke* 2015;46:2032-60.
- 46 Peigne V, Chaize M, Falissard B, *et al.* Important questions asked by family members of intensive care unit patients. *Crit Care Med* 2011;39:1365-71.
- 47 Saigle V, Asad S, Presseau J, *et al.* Do patient-reported outcome measures for SAH include patient, family, and caregiver priorities? A scoping review. *Neurology* 2019;92:281-95.
- 48 Gaudry S, Messika J, Ricard J-D, *et al.* Patient-important outcomes in randomized controlled trials in critically ill patients: a systematic review. *Ann Intensive Care* 2017;7:28.
- 49 Ashley JG, Ashley MJ, Masel BE, *et al.* The influence of post-acute rehabilitation length of stay on traumatic brain injury outcome: a retrospective exploratory study. *Brain Inj* 2018;32:600-7.
- 50 Chaisinanunkul N, Adeoye O, Lewis RJ, *et al.* Adopting a patient-centered approach to primary outcome analysis of acute stroke trials using a Utility-Weighted modified Rankin scale. *Stroke* 2015;46:2238-43.
- 51 Nobels-Janssen E, van der Wees PJ, Verhagen WIM, *et al.* Patient-Reported outcome measures in subarachnoid hemorrhage: a systematic review. *Neurology* 2019;92:1096-112.
- 52 Tagliaferri F, Compagnone C, Korsic M, *et al.* A systematic review of brain injury epidemiology in Europe. *Acta Neurochir* 2006;148:255-68.
- 53 Reeves M, Lisabeth L, Williams L, *et al.* Patient-Reported outcome measures (PROMs) for acute stroke: rationale, methods and future directions. *Stroke* 2018;49:1549-56.
- 54 Pepe MS, Janes H, Longton G, *et al.* Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. *Am J Epidemiol* 2004;159:882-90.
- 55 Steyerberg EW, Uno H, Ioannidis JPA, *et al.* Poor performance of clinical prediction models: the harm of commonly applied methods. *J Clin Epidemiol* 2018;98:133-43.
- 56 Hosmer DW, Hosmer T, Le Cessie S, *et al.* A comparison of goodness-of-fit tests for the logistic regression model. *Stat Med* 1997;16:965-80.
- 57 Van Calster B, Van Hoorde K, Vergouwe Y, *et al.* Validation and updating of risk models based on multinomial logistic regression. *Diagn Progn Res* 2017;1:2.
- 58 Steyerberg EW. Validation in prediction research: the waste by data splitting. *J Clin Epidemiol* 2018;103:131-3.
- 59 Zamanipoor Najafabadi AH, Ramspek CL, Dekker FW, *et al.* Tripod statement: a preliminary pre-post analysis of reporting and methods of prediction models. *BMJ Open* 2020;10:e041537.
- 60 Rost NS, Smith EE, Chang Y, *et al.* Prediction of functional outcome in patients with primary intracerebral hemorrhage: the FUNC score. *Stroke* 2008;39:2304-9.
- 61 Weimar C, Roth M, Willig V, *et al.* Development and validation of a prognostic model to predict recovery following intracerebral hemorrhage. *J Neurol* 2006;253:788-93.
- 62 Gupta VP, Garton ALA, Sisti JA, *et al.* Prognosticating functional outcome after intracerebral hemorrhage: the ICHOP score. *World Neurosurg* 2017;101:577-83.
- 63 Katsanos AH, Krogias C, Lioutas V-A, *et al.* The prognostic utility of ICH-score in anticoagulant related intracerebral hemorrhage. *J Neurol Sci* 2020;409:116628.
- 64 Appelboom G, Hwang BY, Bruce SS, *et al.* Predicting outcome after arteriovenous malformation-associated intracerebral hemorrhage with the original ICH score. *World Neurosurg* 2012;78:646-50.
- 65 Wolff RF, Moons KGM, Riley RD, *et al.* PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med* 2019;170:51-8.
- 66 Moons KGM, Wolff RF, Riley RD, *et al.* PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. *Ann Intern Med* 2019;170:W1-33.
- 67 Rezaeian M. Reducing publication bias in biomedical research: reviewing and registering protocol with a suitable Journal. *J Clin Epidemiol* 2016;69:248-9.
- 68 Collins GS, de Groot JA, Dutton S, *et al.* External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC Med Res Methodol* 2014;14:40.