Journal of the Neurological Sciences 379 (2017) 103–108

Contents lists available at ScienceDirect

Journal of the Neurological Sciences

journal homepage: www.elsevier.com/locate/jns

Comparison of all 19 published prognostic scores for intracerebral hemorrhage

Jarno Satopää [a,*], Satu Mustanoja [b], Atte Meretoja [b], Jukka Putaala [b], Markku Kaste [b], Mika Niemelä [a], Turgut Tatlisumak [b], Daniel Strbian [b,c,d]

[a] Department of Neurosurgery, Helsinki University Hospital, Clinical Neurosciences, Neurosurgery, University of Helsinki, Finland
[b] Department of Neurology, Helsinki University Hospital, Clinical Neurosciences, Neurology, University of Helsinki, Finland
[c] Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Sweden
[d] Department of Neurology, Sahlgrenska University Hospital, Sweden

ARTICLE INFO

Article history:
Received 4 January 2017
Received in revised form 17 April 2017
Accepted 16 May 2017
Available online 17 May 2017

Keywords:
Cerebral hemorrhage
Stroke
Prognosis
Patient outcome assessment
Mortality

ABSTRACT

Background and aims: We evaluated the accuracy of 19 published prognostic scores to find the best tool for predicting mortality after intracerebral hemorrhage (ICH).
Methods: A retrospective single-center analysis of consecutive patients with ICH (n = 1013). After excluding patients with missing data (n = 131), we analyzed 882 patients for 3-month (primary outcome), in-hospital, and 12-month mortality. We analyzed the strength of the individual score components and calculated the c-statistics, Youden index, sensitivity, specificity, negative and positive predictive value (NPV and PPV) for the scores. Finally, we included every score component in a multivariable model to analyze the maximum predictive value of the data elements combined.
Results: Observed in-hospital mortality was 23.6%, 3-month mortality was 31.0%, and 12-month mortality was 35.3%. For in-hospital mortality, the National Institutes of Health Stroke Scale (NIHSS) performed equally good as the best score for the other outcomes, the ICH Functional Outcome Score (ICH-FOS). The c-statistics of the scores varied from 0.6293 (95% CI 0.587–0.672) to 0.8802 (0.855–0.906). With all variables from all the scores in a multivariable regression model, the c-statistics did not improve, being 0.89 (0.867–0.913). Using the Youden index cutoff for the ICH-FOS score, the sensitivity (73%), specificity (90%), PPV (76%), and NPV (88%) for the primary outcome were good.
Conclusions: A plethora of scores exists to help clinicians estimate the prognosis of an acute ICH patient. The NIHSS can be used to quantify the risk of in-hospital death while the ICH-FOS performed best for the other outcomes.

1. Introduction

Intracerebral hemorrhage (ICH) has often a dismal prognosis, with average 12-month mortality exceeding 50% [1]. Estimating the patient's risk of death is typically based on the treating physicians' clinical experience. However, over-estimation may lead to unnecessary withdrawal or limitation of care, the so-called "self-fulfilling prophecy", while under-estimation can result in prolonged, unnecessary, costly, and futile treatment [2]. To help the clinician, multiple scores have been published to estimate the prognosis of an ICH patient.

Prognostic scores became hugely popular following the publication of the ICH score in 2001 [3]. Since then, nearly twenty scores have been published. The ICH score is being widely used in clinical practice in many parts of the world.

However, many of the newer scores have been derived from small single-center cohorts and some have neither been compared to each other nor undergone external validation. In addition to age, ICH volume, and ICH location that are included in majority of the scores, some include Glasgow coma scale (GCS) or National Institutes of Health Stroke Scale (NIHSS) as score subcomponents. Some authors have stated that one of these alone might be powerful enough and suffice for mortality prognostication after ICH in clinical practice [4,5]. This could be welcomed, as some of the prognostic models are very complex and this may hinder their real-life usability.

Instead of creating another new prognostic score for ICH-related mortality, our objective was to find the best tool among the existing prognostic models and to validate the lesser-known scores in a large, external cohort. We evaluated and compared the prognostic accuracy of all published scores and their subcomponents in predicting mortality

* Corresponding author at: Department of Neurosurgery, Helsinki University Hospital, Topeliuksenkatu 5, 00029 Helsinki, Finland.
E-mail address: jarno.satopaa@hus.fi (J. Satopää).

http://dx.doi.org/10.1016/j.jns.2017.05.034
0022-510X/

after ICH in a large, consecutive, single-center cohort of typical ICH patients treated in a tertiary specialist acute stroke center.

2. Methods

2.1. Patient selection

A retrospective analysis of 1013 consecutive ICH patients from the Helsinki ICH Study (HICHS) presenting at the Helsinki University Hospital neurological emergency department between January 2005 and March 2010. All data were collected retrospectively from charts, electronic patient records, and imaging archives. The patient population and data retrieval methods have been previously described [6]. We excluded cases with missing data to calculate all of the scores. The catchment population is 1.8 million and the hospital has the only neurological emergency room in the province of Uusimaa. Patients were treated according to the European guidelines [7]. Mortality data was collected from the National Death Registry. The Helsinki University Hospital scientific board approved the study as a registry study with no patient contact or consent. According to Finnish law, no patient consent is needed for retrospective registry studies.

2.2. Prognostic scores

A thorough search was performed in PubMed, OvidSP, Web of Science, and Google to include all the available prognostic models for ICH published after 1990. We used the terms 'ICH', 'prognosis', 'score', 'scale', and 'validation'. None of the scores was used for clinical prognostication at our institute.

We calculated all the scores for each included patient. The FUNC score [8] was derived to estimate good outcome instead of mortality and was hence reciprocated for the receiver operator curve (ROC) analyses. For determining the ICH volume, we preferred the same method as the original articles. The Graeb [9] and Hallevi [10] intraventricular hemorrhage (IVH) scores were determined by JS and SM. The Tuhrim equation [11] included the IVH volume (cm3) estimated by computer-based image analysis. As this was not available at our center, we used the mathematical method by Hallevi [10]. Due to unavailability of some computerized volumetric methods and structured questionnaires at our institute, we used the best possible alternatives (Online supplement, Methods).

2.3. Statistical methods

The accuracy of the scores was evaluated for three different measures: 3-month all-cause mortality was chosen as the primary outcome, whereas in-hospital and 12-month mortality were measured as secondary outcomes.

The univariate association of all the different score components and outcome was analyzed. The ordinal and continuous score components were tested for normality. The Kruskal-Wallis and Mann-Whitney *U* tests were used for skewed, and Analysis of Variance (ANOVA) for normally distributed data. Non-parametric tests were used for skewed data. Because of the time series nature of the analysis, we used Kaplan-Meier survival analysis with the time from ICH to death up to 12 months as a continuous variable for the dichotomous score components. The Breslow-Wilcoxon test was used to test for univariate differences.

To estimate the best possible prognostic performance of all the score variables in a tailored logistic regression model, we divided the study population to a derivation and a validation cohort stratified for age and ICH volume. The cohorts were tested for statistically significant differences. With all the score components forced to enter the model, a logistic regression model was constructed from the derivation cohort. No dichotomization was performed neither on the continuous nor the ordinal variables. The resulting model was tested on the validation cohort.

The c-statistics of a) the best performing individual score component, b) the best performing score, c) the original ICH score, d) GCS as advised by Parry-Jones and colleagues, [4] and e) our optimized logistic regression model were tested against each other. The Youden index (J) was calculated to determine the optimal cut off points and, further, the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of the selected scores. In addition, we created a bivariate correlation matrix and used Spearman's ρ and tolerance as analyses of collinearity to measure the relative independence of the different score components.

The analyses were conducted using SPSS 20 (IBM corp., NY). For the statistical differences in the c-statistics, we used Z derived from ROC-kit v.1.0.3 software [12] (http://metz-roc.uchicago.edu/) with non-

Table 1
Included scores and their derivation cohorts, locations and primary outcomes.

Year	Score name	n	Prospective	Location (number of centers)	Mortality outcomes	Functional outcomes
1993	Cincinnati model [24]	162	No	Ohio, USA [20]	30-day	30-day OHS
1995	Masé equation [25]	138	No	Trieste, Italy [1]	30-day	–
1999	Tuhrim equation [11]	129	Yes	New York, USA [1]	30-day	–
2001	ICH score (oICH) [3]	152	No	San Francisco, USA [2]	30-day	–
2003	new ICH score (nICH) and modified ICH score (mICH) [16]	142	No	Hong Kong, China [1]	30-day	30-day mRS 0–2
2006	modified ICH-A and -B (mICH-A and –B) [26]	153	Yes	Junin and Bahia Blanca, Argentina [2]	30-day	180-day GOS 4–5, 180-day GOS 2–3
2006	Essen ICH score [17]	340	No	Germany [30]	100-day	100-day BI 95–100, 100-day BI 0–90
2006	GP on stage score (GPoS) [38]	995	Yes	Asia [14]	–	Discharge mRS 5–6
2007	ICH grading scale (ICH-GS) [19]	378	Yes	Guadalaraya, Mexico [1]	In-hospital, 30-day	30-day GOS 4–5
2008	FUNC score [8]	418	Yes	Boston, USA [1]	–	90-day GOS 4–5
2008	Cho's MICH score [22]	226	Yes	Taichung, Taiwan [1]	180-day	180-day GOS 4–5, 180-day BI 55–100
2008	ICH outcome score (ICHOS) [29]	107	No	Taoyan, Taiwan [1]	30-day	–
2009	Simplified ICH score (sICH) [27]	293	No	Taichung, Taiwan [1]	30-day	–
2011	Landseed ICH score (LSICH) [28]	285	No	Taoyan, Taiwan [1]	In-hospital	Discharge BI <40
2012	ICH index (ICHI) [34]	227	No	Chongqing, China [1]	In-hospital	
2013	ICH functional outcome score (ICH-FOS) [14]	1953	No	China (132)	–	1-year mRS 3–6
2013	GWTG-stroke score [18]	~6000	Yes	USA (1046)	In-hospital	–

OHS, Oxford Handicap Scale; GOS, Glasgow Outcome Scale; mRS, modified Rankin Scale; BI, Barthel Index.

parametric assumptions and U-statistic based method [13]. A two-sided $p < 0.05$ was considered significant.

2.4. Role of the funding sources

The funding sources had no impact on the study design, collection, analysis nor interpretation of the data and neither in writing nor submitting the report.

3. Results

We found 19 prognostic scores (Table 1). Many of the scores shared similar point assignments (Online supplement, Table I). Additionally, five scores provided a formula for the probability of death or survival (Online supplement, Table II). Our retrospective database included 1013 consecutive patients with a spontaneous ICH. Of these, 131 were excluded due to missing data, leaving 882 patients for the analyses. The demographics are presented in Tables 2 and 3. The observed in-

Table 2
Demographics of the study population and association of the dichotomous score components with mortality. Kaplan-Meier analysis up to 12 months and Breslow statistic were used to test for group differences.

	Total (n = 882)	In-hospital mortality, n (row %)	3-month mortality, n (row %)	12-month mortality, n (row %)	p-Value
Sex					0.0185
Male	508 (57.6%)	136 (26.8%)	171 (33.7%)	192 (37.8%)	
Female	374 (42.4%)	72 (19.3%)	102 (27.3%)	119 (31.8%)	
Cortical location					0.5101
Yes	338 (38.3%)	76 (22.5%)	96 (28.4%)	115 (34.0%)	
No	554 (61.7%)	132 (24.3%)	177 (32.5%)	196 (36.0%)	
Deep (including thalamic) location					0.1291
Yes	483 (54.8%)	122 (25.3%)	164 (34.0%)	182 (37.7%)	
No	399 (45.2%)	86 (21.6%)	109 (27.3%)	129 (32.3%)	
Infratentorial location					0.0002
Yes	123 (13.9%)	49 (39.8%)	55 (44.7%)	56 (45.5%)	
No	759 (86.1%)	159 (20.9%)	218 (28.7%)	255 (33.6%)	
Intraventricular hemorrhage					<0.0001
Yes	354 (40.1%)	143 (40.4%)	186 (52.5%)	198 (55.9%)	
No	528 (59.9%)	65 (12.3%)	87 (16.5%)	113 (21.4%)	
Subarachnoid hemorrhage					0.0764
Yes	113 (12.8%)	46 (40.7%)	61 (54.5%)	69 (61.1%)	
No	769 (87.2%)	162 (21.1%)	212 (27.6%)	242 (31.5%)	
Hydrocephalus					<0.0001
Yes	123 (13.9%)	83 (67.5%)	95 (77.2%)	96 (78.0%)	
No	759 (86.1%)	125 (16.5%)	178 (23.5%)	215 (28.3%)	
Dialysis					0.6309
Yes	14 (1.6%)	4 (28.6%)	5 (35.7%)	5 (35.7%)	
No	868 (98.4%)	204 (23.5%)	268 (30.9%)	306 (35.3%)	
Comorbidities					0.0002
Yes	384 (43.5%)	102 (26.6%)	139 (36.2%)	162 (42.2%)	
No	498 (56.5%)	106 (21.3%)	134 (26.9%)	149 (29.9%)	
Pre-ICH cognitive deficit					0.0003
Yes	40 (4.5%)	14 (35.0%)	21 (52.5%)	26 (65.0%)	
No	842 (95.5%)	194 (23.0%)	252 (29.9%)	285 (33.8%)	
Atrial fibrillation					<0.0001
Yes	90 (10.2%)	34 (37.8%)	40 (44.4%)	45 (50.0%)	
No	792 (89.8%)	174 (22.0%)	233 (29.4%)	266 (33.6%)	
Previous stroke or transient ischemic attack					0.1032
Yes	144 (16.3%)	38 (26.4%)	49 (34.0%)	59 (41.0%)	
No	738 (83.7%)	170 (23.0%)	224 (30.4%)	252 (34.1%)	
Peripheral artery disease					0.0321
Yes	15 (1.7%)	7 (46.7%)	7 (46.7%)	8 (53.5%)	
No	867 (98.3%)	201 (23.2%)	266 (30.7%)	303 (34.9%)	
Coronary heart disease					<0.0001
Yes	122 (13.8%)	37 (30.3%)	54 (44.3%)	63 (51.6%)	
No	760 (86.2%)	171 (22.5%)	219 (28.8%)	248 (32.6%)	
Dyslipidemia					0.2921
Yes	177 (20.1%)	33 (18.6%)	51 (28.8%)	57 (32.2%)	
No	705 (79.9%)	175 (24.8%)	222 (31.5%)	254 (36.0%)	
After-hours arrival					0.6375
Yes	524 (59.4%)	121 (23.1%)	160 (30.5%)	182 (34.7%)	
No	358 (40.6%)	87 (24.3%)	113 (31.6%)	129 (36.0%)	
Arrival by ambulance					<0.0001
Yes	774 (87.8%)	200 (25.8%)	264 (34.1%)	301 (38.9%)	
No	108 (12.2%)	8 (7.4%)	9 (8.3%)	10 (9.3%)	
Arrival by other than emergency room					0.9110
Yes	13 (1.5%)	4 (30.8%)	5 (38.5%)	5 (38.5%)	
No	869 (98.5%)	204 (23.5%)	268 (30.8%)	306 (35.2%)	
Arrival by private transport					<0.0001
Yes	95 (10.8%)	4 (4.2%)	4 (4.2%)	5 (5.3%)	
No	787 (89.2%)	204 (25.9%)	269 (34.2%)	306 (38.9%)	
Diabetes					0.3211
Yes	125 (14.2%)	32 (25.6%)	41 (32.8%)	50 (40.0%)	
No	757 (85.8%)	176 (23.2%)	232 (30.6%)	261 (34.5%)	

hospital mortality was 23.6%, 3-month mortality was 31.0%, and 12-month mortality was 35.3%.

The majority of the score components were associated with mortality in univariate analyses (Tables 2 and 3). Of the continuous variables, only "systolic blood pressure", "pulse pressure", and "temperature" were not associated with mortality. The included and excluded patients did not differ in regard to the main score components (Online supplement, Table III).

In the receiver operating curve (ROC) analyses, the National Institutes of Health Stroke Scale (NIHSS) had the highest c-statistic of the individual score components at every time point. In the ROC analyses of the scores, the ICH Functional Outcome Score (ICH-FOS) score [14] performed best. For 3-month mortality, the c-statistic was 0.8802 and, for in-hospital and 12-month mortality, 0.8661 and 0.8642, respectively (Table 4).

The c-statistics of the ICH-FOS, NIHSS, GCS, the original ICH score, and our experimental logistic model were tested against each other (Table 5). For the 3- and 12-month mortality, the ICH-FOS was the best performer with a statistically significant difference to the other tested scores, with a minor edge over NIHSS. For the in-hospital mortality, our analyses did not show a statistically significant difference between the NIHSS, GCS, and ICH-FOS scores. The original ICH score was equal to NIHSS and CGS but inferior to the ICH-FOS. We calculated the Youden index to determine the optimal cut-off limit for mortality in the best performing scores (Table 4). For the ICH-FOS, the optimal cut off was different for the different outcomes: >8 points for in-hospital and 3-month mortality ($J_{max} = 0.599$ and $J_{max} = 0.626$, respectively) and >7 points for 12-month mortality ($J_{max} = 0.591$).

The differences of the derivation and validation cohorts for our experimental regression model were minor (Online supplement, Table IV). It produced good c-statistic values for all time points but did not perform significantly better than the best available score (Table 5). In correlation analyses the "NIHSS score", "GCS score", and "NIHSS alertness" score components, as well as systolic and pulse pressure were highly correlated ($|\rho| \geq 0.7$). In tests of collinearity, adding the aforementioned variables in the same regression model produced remarkably high tolerances (0.256 for GCS, 0.229 for NIHSS, and 0.240 for NIHSS alertness).

4. Discussion

Our results showed that the ICH-FOS score performed best with a small edge over the NIHSS for predicting 3-month and 12-month mortality, and for in-hospital mortality they performed equally well. The NIHSS is a purely clinical score, whereas using the ICH-FOS requires both GCS and NIHSS scores, a CT scan and blood glucose. In modern neurological emergency units, all these data are readily available, but in some environments, a simpler alternative could be welcomed. The NIHSS was created to assess the impairment severity in ischemic stroke patients, but it has since then been validated for use in patients with ICH [15]. It seems logical that the presentation neurological status should be prognostic of mortality, at least short-term. The same result has been suggested previously [5]. The best performing score, ICH-FOS, includes GCS and NIHSS total scores, both measures of clinical neurological status and level of consciousness thus reinforcing their importance. In addition to ICH-FOS, NIHSS was included in the Modified and New ICH scores, [16] the Essen ICH score, [17] and the Get With The Guidelines (GWTG) score [18]. They do not, however, report the c-statistics for NIHSS or the other individual score components.

Multiple authors have underlined the need for a score better suited to prognosticate poor functional outcome. In a recent study the FUNC [8] and ICH grading scale (ICH-GS) scores [19] outperformed the original ICH score in predicting poor neurological outcome [20]. Majority of the scores, however, try to estimate mortality. Parry-Jones and colleagues compared the original ICH score, [21] GCS, ICH-GS, and modified ICH score [22] in estimating 30-day mortality. They found that the three former scores performed equally well and suggested that the simplest alternative, GCS, should suffice for predicting prognosis after an ICH [4].

In ischemic stroke, experienced clinicians' ability to correctly predict 30-day mortality may be as low as 33.1%, [23] and in comparison with these figures, all the scores performed extremely well. However, many of the scores seem to include outcomes defined *a posteriori* and some have been developed on small single-center cohorts. External validation of the lesser-known scores has been scarce (Supplemental data, Table V). Our study is the first to include all available prognostic scores for mortality after ICH, making it possible to compare their performance in an external population. The three mathematical models – Cincinnati model, [24] Masé model, [25] and the Tuhrim equation [11] – seem hard to implement in day-to-day clinical usage, and their validation in other studies has been very limited.

In majority of the models, the selection of score components may have been based on choosing the "most significant" variables in univariate analysis for the logistic regression. Some could argue this can be used to find new explanatory variables, although this clearly has not been the case with ICH prognostication as majority of the scores share the same variables.

In addition to GCS as a measure of level of consciousness, the original ICH score includes age, infratentorial location, ICH volume, and intraventricular extension. The best performing score, ICH-FOS, adds NIHSS as a measure of clinical neurological status and hyperglycemia on admission as a measure of uncontrolled diabetes [14]. Many of the scores include comorbidities, [18,26–28] some include hyper- or hypotension, [16,27,29] and one prognostic score even includes out-of-hours arrival [18]. These factors probably affect the mortality after ICH, but they may not be well generalizable to all ICH patients. In our study, body temperature, blood sugar level, and blood pressure did not have a strong influence on mortality. However, comorbidities, hydrocephalus and arrival to hospital by ambulance (and not by private transport) were highly associated with mortality in our study. Hydrocephalus has been found a significant prognostic factor in patients with ICH, [30] but it was not included in any of the scores. One explanation may be that ICH-related hydrocephalus is usually caused by intraventricular or cerebellar hemorrhage, which are, in turn, included in majority of the scores. Although age was included in majority of the scores, it had a rather

Table 3
Demographics of the study population and the univariate association of the continuous score components with mortality. C-statistics with 95% confidence intervals.

Variable	Median (IQR)	In-hospital mortality, c-statistic	3-month mortality, c-statistic	12-month mortality, c-statistic
Age (years)	68 (58–78)	0.559 (0.515–0.604)	0.633 (0.593–0.674)	0.658 (0.620–0.696)
ICH volume (cm3)	9.70 (3.76–26.1)	0.759 (0.720–0.799)	0.761 (0.726–0.797)	0.757 (0.723–0.791)
GCS	14 (11–15)	0.825 (0.790–0.860)	0.790 (0.755–0.824)	0.767 (0.732–0.801)
NIHSS	11 (4–19)	0.852 (0.821–0.882)	0.848 (0.820–0.876)	0.826 (0.797–0.855)
NIHSS alertness	0 (0–1)	0.813 (0.776–0.849)	0.778 (0.742–0.814)	0.753 (0.717–0.789)
Graeb score	0 (0–4)	0.733 (0.690–0.776)	0.743 (0.705–0.781)	0.717 (0.680–0.755)
Temperature (°C)	36.7 (36.3–37.2)	0.584 (0.534–0.633)	0.582 (0.539–0.626)	0.559 (0.516–0.601)
Blood glucose (mmol/L)	7.2 (6.1–9.1)	0.686 (0.646–0.726)	0.663 (0.625–0.701)	0.645 (0.608–0.683)
Systolic blood pressure (mm Hg)	172 (149–193)	0.541 (0.493–0.590)	0.533 (0.490–0.576)	0.516 (0.476–0.557)
Pulse pressure (mm Hg)	79 (63–96)	0.545 (0.496–0.593)	0.546 (0.502–0.589)	0.539 (0.497–0.580)

IQR, interquartile range; GCS, Glasgow coma scale; NIHSS, National Institutes of Health Stroke Scale.

Table 4
Performance of the scores. C-statistics with 95% confidence intervals. Youden index, cut-off, sensitivity, specificity, PPV and NPV are presented for scores selected for further analysis and for 3-month mortality only.

Score name	In-hospital mortality, c-statistic	3-month mortality, c-statistic	12-month mortality, c-statistic	Youden index	Cut-off (points)	Sensitivity	Specificity	PPV	NPV
Cincinnati model [24]	0.7898 (0.754–0.825)	0.7664 (0.734–0.799)	0.7393 (0.708–0.771)						
Masé model [25]	0.8570 (0.827–0.888)	0.8480 (0.819–0.877)	0.8230 (0.793–0.853)						
Tuhrim [11]	0.6580 (0.611–0.705)	0.6293 (0.587–0.672)	0.6187 (0.578–0.672)						
oICH [3]	0.8414 (0.811–0.872)	0.8454 (0.818–0.873)	0.8163 (0.787–0.846)	0.54	>2	77%	77%	65%	86%
mICH [16]	0.8497 (0.821–0.879)	0.8661 (0.841–0.891)	0.8453 (0.819–0.872)						
nICH [16]	0.7931 (0.759–0.827)	0.8010 (0.770–0.832)	0.7881 (0.757–0.819)						
mICH-A [26]	0.8498 (0.818–0.882)	0.8584 (0.831–0.886)	0.8392 (0.811–0.867)						
mICH-B [26]	0.8500 (0.818–0.882)	0.8574 (0.830–0.885)	0.8381 (0.810–0.866)						
Essen score [17]	0.8387 (0.807–0.870)	0.8539 (0.826–0.882)	0.8470 (0.820–0.875)						
GPoS [38]	0.8520 (0.821–0.884)	0.8370 (0.808–0.867)	0.8100 (0.799–0.841)						
ICH-GS [19]	0.8429 (0.811–0.875)	0.8419 (0.814–0.870)	0.8156 (0.786–0.845)						
FUNC [8]	0.8086 (0.771–0.843)	0.8126 (0.781–0.845)	0.7858 (0.753–0.818)						
Cho's MICH [22]	0.8503 (0.820–0.880)	0.8395 (0.811–0.868)	0.8125 (0.783–0.842)						
ICHOS [29]	0.7570 (0.716–0.798)	0.765 (0.729–0.801)	0.7620 (0.728–0.796)						
sICH [27]	0.7897 (0.754–0.826)	0.7781 (0.745–0.812)	0.7654 (0.733–0.798)						
LSICH [28]	0.8280 (0.794–0.862)	0.8170 (0.785–0.849)	0.7890 (0.757–0.822)						
ICHI [34]	0.7858 (0.751–0.821)	0.7688 (0.735–0.803)	0.7462 (0.712–0.780)						
ICH-FOS [39]	0.8661 (0.838–0.896)	0.8802 (0.855–0.906)	0.8642 (0.838–0.891)	0.63	>8	73%	90%	76%	88%
GWTG Stroke [18]	0.8540 (0.824–0.883)	0.860 (0.833–0.887)	0.8450 (0.818–0.872)						
NIHSS	0.8515 (0.821–0.882)	0.8475 (0.820–0.876)	0.8260 (0.797–0.855)	0.57	>14	81%	76%	60%	90%
GCS	0.8250 (0.790–0.860)	0.7900 (0.755–0.824)	0.7670 (0.732–0.801)	0.46	<11	57%	89%	70%	82%
Reg. model (derivation)	0.9090 (0.875–0.944)	0.9030 (0.871–0.934)	0.8830 (0.849–0.917)	0.67	n/a	78%	89%	75%	90%
Reg. model (validation)	0.8350 (0.792–0.879)	0.8660 (0.830–0.902)	0.8660 (0.830–0.901)	0.59	n/a	81%	78%	64%	90%
Reg. model (all)	0.8730 (0.844–0.901)	0.8900 (0.867–0.913)	0.8740 (0.850–0.899)	0.62	n/a	77%	85%	70%	89%

PPV, positive predictive value; NPV, negative predictive value.

weak association to mortality in our study. This aligns with previous studies, where age was not a very strong predictor for mortality. Therefore, the authors state, clinicians must remember not to stress the patients age too much when considering the possibilities of recovery after an ICH [4].

Choosing variables only based on univariate analyses can also lead to statistical over-fitting – the regression model may suggest leaving out covariates that are generally significant but not so for the derivation cohort [31]. In addition, majority of the scores treat continuous variables by categorization or dichotomization, which may cause loss of information and statistical power [32]. We found that the c-statistics of the best performing score, ICH-FOS, for one-year mortality were higher in our study than in the original study (0.8642 vs. 0.830) [14]. Our tailored regression model was not statistically better, and it seems that the ICH-FOS score was not statistically over-fitted to the original derivation population.

Table 5
Statistical significance of the differences in prognostic performance at the three time points. The results for the regression model are presented for both derivation and validation cohorts.

	NIHSS p	GCS p	oICH [21] p	ICH-FOS [14] p
In-hospital mortality				
GCS	0.3827	–	–	–
oICH	0.4858	0.5997	–	–
ICH-FOS	0.1522	0.1600	0.0092	–
Regression mod. (all patients)	0.1669	0.1037	0.0316	0.7058
3-month mortality				
GCS	0.0411	–	–	–
oICH	0.8786	0.0506	–	–
ICH-FOS	0.0003	0.0008	0.0002	–
Regression mod. (all patients)	0.0020	0.0003	<0.0001	0.6241
12-month mortality				
GCS	0.0366	–	–	–
oICH	0.4928	0.0854	–	–
ICH-FOS	<0.0001	0.0003	<0.0001	–
Regression mod. (all patients)	<0.0001	<0.0001	<0.0001	0.2588

Each new score aims at improving prognostication in all or some subgroups of ICH patients – patients in hemodialysis, [33] patients in developing countries, [34] or a highly selected cohort for surgical intervention [22]. However, the majority of the scores are based on the original ICH score [3] and include the same variables with slight differences in the point assignments, possibly reflecting the differences in the small derivation cohorts. In our analyses, majority of the scores performed reasonably well, with only the Tuhrim equation giving a c-statistic < 0.7 for 3-month mortality. Although there are no published data on the clinical usage of the different scores, to our knowledge, only the ICH score is in wide-spread clinical use.

Many of the authors share the concern of "self-fulfilling prophecies", the end-of-life decisions in the derivation cohorts having an impact on the prognostic value of the scores [2,3,8,35,36]. The data on do-not-resuscitate orders in our cohort have been published elsewhere [37]. In our opinion, as the ICH treatment policies vary significantly between the different centers and different physicians, not only the decisions to withdraw care but all the treatment decisions in the score derivation cohort patients have inevitably been incorporated in the scores and may act as potential confounding factors.

The strengths of our study include the large sample size and the fact that none of the scores was used for clinical prognostication, and the prognosis calculations could not therefore affect the treatment decisions. Our study has its limitations as well. The database did not include data on long-term functional outcome. Being a retrospective study, our data may suffer of minor selection bias, as the patients with the worst prognosis may have died before arrival to hospital. Some patients may have been left in primary admitting hospitals due to suspected grim prognosis. This was a single center study and the performance of the prognostic scores may differ at institutions with different treatment resources or approaches.

5. Conclusion

A plethora of scores exists to help clinicians estimate the prognosis of an acute ICH patient. The ICH-FOS performed best, but NIHSS is a good alternative for prediction of in-hospital mortality. None of these

scores is perfect, but they perform well as tools for quantifying the risk of death while communicating with patients and their relatives.

Conflicts of interest and funding

We report no conflicts of interest. The research was supported by Helsinki University Central Hospital Research Funds (EVO). Dr. Satopää has received research support from Maire Taponen Foundation. Dr. Meretoja has received research support and the National Health and Medical Research Council of Australia Centre for Research Excellence.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jns.2017.05.034.

References

[1] C.J. van Asch, M.J. Luitse, G.J. Rinkel, I. van der Tweel, A. Algra, C.J. Klijn, 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. 9 (2) (2010 Feb) 167–176.
[2] D.B. Zahuranec, L.B. Morgenstern, B.N. Sánchez, K. Resnicow, D.B. White, J.C. Hemphill, Do-not-resuscitate orders and predictive models after intracerebral hemorrhage, Neurology 75 (7) (Aug 2010) 626–633 Lippincott Williams & Wilkins.
[3] J.C. Hemphill, D.C. Bonovich, L. Besmertis, G.T. Manley, S.C. Johnston, S. Tuhrim, The ICH score: a simple, reliable grading scale for intracerebral hemorrhage, Stroke 32 (4) (Apr 1, 2001) 891–897.
[4] A.R. Parry-Jones, K.A. Abid, M. Di Napoli, C.J. Smith, A. Vail, H.C. Patel, et al., Accuracy and clinical usefulness of intracerebral hemorrhage grading scores: a direct comparison in a UK population, Stroke 44 (7) (2013 Jul) 1840–1845.
[5] C.M. Cheung, T.H. Tsoi, S.F.K. Hon, M. Au-Yeung, K.L. Shiu, C.N. Lee, et al., Using the National Institutes of Health Stroke Scale (NIHSS) to predict the mortality and outcome of patients with intracerebral haemorrhage, Hong Kong Med J. 14 (5) (Oct 2008) 367–370.
[6] A. Meretoja, D. Strbian, J. Putaala, S. Curtze, E. Haapaniemi, S. Mustanoja, et al., SMASH-U: a proposal for etiologic classification of intracerebral hemorrhage, Stroke 43 (10) (Oct 2012) 2592–2597.
[7] T. Steiner, R. Al-Shahi Salman, R. Beer, H. Christensen, C. Cordonnier, L. Csiba, et al., European Stroke Organisation (ESO) guidelines for the management of spontaneous intracerebral hemorrhage, Int. J. Stroke 9 (2014) 840–855.
[8] N.S. Rost, E.E. Smith, Y. Chang, R.W. Snider, R. Chanderraj, K. Schwab, et al., Prediction of functional outcome in patients with primary intracerebral hemorrhage: the FUNC score, Stroke 39 (8) (Aug 2008) 2304–2309.
[9] D.A. Graeb, W.D. Robertson, J.S. Lapointe, R.A. Nugent, P.B. Harrison, Computed tomographic diagnosis of intraventricular hemorrhage. Etiology and prognosis, Radiology 143 (1) (Apr 1982) 91–96.
[10] H. Hallevi, N.S. Dar, A.D. Barreto, M.M. Morales, S. Martin-Schild, A.T. Abraham, et al., The IVH score: a novel tool for estimating intraventricular hemorrhage volume: clinical and research implications, Crit. Care Med. 37 (3) (Mar 2009) (969–74–e1).
[11] S. Tuhrim, D.R. Horowitz, M. Sacher, J.H. Godbold, Volume of ventricular blood is an important determinant of outcome in supratentorial intracerebral hemorrhage, Crit. Care Med. 27 (3) (Mar 1999) 617–621.
[12] C.E. Metz, B.A. Herman, J.H. Shen, Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data, Stat. Med. 17 (9) (May 1998) 1033–1053.
[13] X.-H. Zhou, S.M. Li, C.A. Gatsonis, Wilcoxon-based group sequential designs for comparison of areas under two correlated ROC curves, Stat. Med. 27 (2) (Jan 2008) 213–223.
[14] R. Ji, H. Shen, Y. Pan, P. Wang, G. Liu, Y. Wang, et al., A novel risk score to predict 1-year functional outcome after intracerebral hemorrhage and comparison with existing scores, Crit. Care 17 (6) (2013) R275.
[15] A.V. Specogna, S.B. Patten, T.C. Turin, M.D. Hill, The reliability and sensitivity of the National Institutes of Health Stroke Scale for spontaneous intracerebral hemorrhage in an uncontrolled setting, PLoS One 8 (12) (2013) e84702.
[16] R.T.F. Cheung, L.-Y. Zou, Use of the original, modified, or new intracerebral hemorrhage score to predict mortality and morbidity after intracerebral hemorrhage, Stroke 34 (7) (Jul 2003) 1717–1722.
[17] C. Weimar, J. Benemann, H.-C. Diener, German Stroke Study Collaboration, Development and validation of the Essen Intracerebral Haemorrhage Score, J. Neurol. Neurosurg. Psychiatry 77 (5) (May 2006) 601–605.
[18] E.E. Smith, N. Shobha, D. Dai, D.M. Olson, M.J. Reeves, J.L. Saver, et al., A risk score for in-hospital death in patients admitted with ischemic or hemorrhagic stroke, J. Am. Heart Assoc. 2 (1) (Feb 2013) e005207.
[19] J.L. Ruiz-Sandoval, E. Chiquete, S. Romero-Vargas, J.J. Padilla-Martínez, S. González-Cornejo, Grading scale for prediction of outcome in primary intracerebral hemorrhages, Stroke 38 (5) (May 2007) 1641–1644.
[20] J.S. Garrett, M. Zarghouni, K.F. Layton, D. Graybeal, Y.A. Daoud, Validation of clinical prediction scores in patients with primary intracerebral hemorrhage, Neurocrit. Care. 19 (3) (Dec 2013) 329–335.
[21] J.C. Hemphill, D.C. Bonovich, L. Besmertis, G.T. Manley, S.C. Johnston, The ICH score: a simple, reliable grading scale for intracerebral hemorrhage, Stroke 32 (4) (Apr 2001) 891–897.
[22] D.-Y. Cho, C.-C. Chen, W.-Y. Lee, H.-C. Lee, L.-H. Ho, A new Modified Intracerebral Hemorrhage score for treatment decisions in basal ganglia hemorrhage—a randomized trial, Crit. Care Med. 36 (7) (Jul 2008) 2151–2156.
[23] G. Saposnik, R. Cote, M. Mamdani, S. Raptis, K.E. Thorpe, J. Fang, et al., JURaSSiC: accuracy of clinician vs risk score prediction of ischemic stroke outcomes, Neurology 81 (5) (Jul 2013) 448–455.
[24] J.P. Broderick, T.G. Brott, J.E. Duldner, T. Tomsick, G. Huster, Volume of intracerebral hemorrhage. A powerful and easy-to-use predictor of 30-day mortality, Stroke 24 (7) (Jul 1993) 987–993.
[25] G. Masè, M. Zorzon, E. Biasutti, G. Tasca, B. Vitrani, G. Cazzato, Immediate prognosis of primary intracerebral hemorrhage using an easy model for the prediction of survival, Acta Neurol. Scand. 91 (4) (Apr 1995) 306–309.
[26] D.A. Godoy, G. Piñero, M. Di Napoli, Predicting mortality in spontaneous intracerebral hemorrhage: can modification to original score improve the prediction? Stroke 37 (4) (Apr 2006) 1038–1044.
[27] Y.-C. Chuang, Y.-M. Chen, S.-K. Peng, S.-Y. Peng, Risk stratification for predicting 30-day mortality of intracerebral hemorrhage, Int. J. Qual. Health Care 21 (6) (Dec 2009) 441–447.
[28] H.-S. Chen, C.-F. Hsieh, T.-T. Chau, C.-D. Yang, Y.-W. Chen, Risk factors of in-hospital mortality of intracerebral hemorrhage and comparison of ICH scores in a Taiwanese population, Eur. Neurol. 66 (1) (2011) 59–63.
[29] B.-R. Huang, C.-C. Liao, W.-H. Huang, Y.-H. Hsu, J.-C. Hsu, H.-C. Yen, et al., Prognostic factors of spontaneous intracerebral haemorrhage in haemodialysis patients and predictors of 30-day mortality, Intern. Med. J. 38 (7) (Jul 2008) 568–574.
[30] M.N. Diringer, D.F. Edwards, A.R. Zazulia, Hydrocephalus: a previously unrecognized predictor of poor outcome from supratentorial intracerebral hemorrhage, Stroke 29 (7) (Jul 1998) 1352–1357.
[31] S. Mallett, P. Royston, S. Dutton, R. Waters, D.G. Altman, Reporting methods in studies developing prognostic models in cancer: a review, BMC Med. 8 (1) (2010) 20.
[32] A.D. Hingorani, D.A.V.D. Windt, R.D. Riley, K. Abrams, K.G.M. Moons, E.W. Steyerberg, et al., Prognosis research strategy (PROGRESS) 4: stratified medicine research, BMJ (2013) 346e5793.
[33] B.-R. Huang, Y.-L. Lo, C.-H. Chang, T.-Y. Chen, Testing the outcome score of spontaneous intracerebral haemorrhage in haemodialysis patients, Intern. Med. J. 39 (10) (Oct 2009) 692–695.
[34] Y.-F. Li, J. Luo, Q. Li, Y.-J. Jing, R.-Y. Wang, R.-S. Li, A new simple model for prediction of hospital mortality in patients with intracerebral hemorrhage, CNS Neurosci. Ther. 18 (6) (Jun 2012) 482–486.
[35] K.J. Becker, A.B. Baxter, W.A. Cohen, H.M. Bybee, D.L. Tirschwell, D.W. Newell, et al., Withdrawal of support in intracerebral hemorrhage may lead to self-fulfilling prophecies, Neurology 56 (6) (Mar 27, 2001) 766–772.
[36] C. Weimar, A. Ziegler, R.L. Sacco, H.C. Diener, I.R. König, VISTA investigators. Predicting recovery after intracerebral hemorrhage—an external validation in patients from controlled clinical trials, J. Neurol. 256 (3) (Mar 2009) 464–469.
[37] K. Silvennoinen, A. Meretoja, D. Strbian, J. Putaala, M. Kaste, T. Tatlisumak, Do-not-resuscitate (DNR) orders in patients with intracerebral hemorrhage, Int. J. Stroke 9 (1) (Jan 2014) 53–58.
[38] N. Poungvarin, N.C. Suwanwela, N. Venketasubramanian, L.K.S. Wong, J.C. Navarro, E. Bitanga, et al., Grave prognosis on spontaneous intracerebral haemorrhage: GP on STAGE score, J. Med. Assoc. Thail. 89 (Suppl. 5) (Nov 2006) S84–S93.
[39] D.-M. Wang, R. Ji, H. Shen, J. Li, J.-R. Liu, Y. Pan, et al., Diffusion tensor imaging predicts long-term motor functional outcome in patients with acute supratentorial intracranial hemorrhage, Cerebrovasc. Dis. 34 (3) (2012) 199–205.