JCN Open Access ORIGINAL ARTICLE
pISSN 1738-6586 / eISSN 2005-5013 / J Clin Neurol 2015

Prognostic Tools for Early Mortality in Hemorrhagic Stroke: Systematic Review and Meta-Analysis

Katharina Mattishent[a]
Chun Shing Kwok[b]
Liban Ashkir[a]
Kelum Pelpola[c]
Phyo Kyaw Myint[d]
Yoon Kong Loke[a]

[a]Health Evidence Synthesis Group, Norwich Medical School, University of East Anglia, Norwich Research Park, Norwich, UK
[b]Institute of Cardiovascular Sciences, University of Manchester, Manchester, UK
[c]Department of Elderly Medicine, Southend University Hospital Trust, Westcliff-on-Sea, Essex, UK
[d]Epidemiology Group, Institute of Applied Health Sciences, School of Medicine & Dentistry, University of Aberdeen, Aberdeen, Scotland, UK

Background and Purpose Several risk scores have been developed to predict mortality in intracerebral hemorrhage (ICH). We aimed to systematically determine the performance of published prognostic tools.

Methods We searched MEDLINE and EMBASE for prognostic models (published between 2004 and April 2014) used in predicting early mortality (<6 months) after ICH. We evaluated the discrimination performance of the tools through a random-effects meta-analysis of the area under the receiver operating characteristic curve (AUC) or c-statistic. We evaluated the following components of the study validity: study design, collection of prognostic variables, treatment pathways, and missing data.

Results We identified 11 articles (involving 41,555 patients) reporting on the accuracy of 12 different tools for predicting mortality in ICH. Most studies were either retrospective or post-hoc analyses of prospectively collected data; all but one produced validation data. The Hemphill-ICH score had the largest number of validation cohorts (9 studies involving 3,819 patients) within our systematic review and showed good performance in 4 countries, with a pooled AUC of 0.80 [95% confidence interval (CI)=0.77−0.85]. We identified several modified versions of the Hemphill-ICH score, with the ICH-Grading Scale (GS) score appearing to be the most promising variant, with a pooled AUC across four studies of 0.87 (95% CI=0.84−0.90). Subgroup testing found statistically significant differences between the AUCs obtained in studies involving Hemphill-ICH and ICH-GS scores (p=0.01).

Conclusions Our meta-analysis evaluated the performance of 12 ICH prognostic tools and found greater supporting evidence for 2 models (Hemphill-ICH and ICH-GS), with generally good performance overall.

Key Words stroke, prognostic scores, risk prediction model, mortality.

INTRODUCTION

Strokes are an important cause of mortality and morbidity worldwide. The consequences of stroke can be severe, leading annually to 5 million deaths and another 5 million people being left permanently disabled.[1] While hemorrhagic stroke/intracerebral hemorrhage (ICH) is less common than ischemic stroke, the prognosis of ICH is substantially worse than those conditions with an ischemic etiology. The proportion of stroke patients with ICH was 14.5% in an Australian study, with a 28-day mortality of 45%, similar to data obtained in Europe and the US.[2] The threat from ICH appears to be growing (perhaps due to an aging population), as indicated by a 47% increase in its incidence and a 20% increase in the number of deaths during 1990−2010 in the Global Burden Disease Study.[3]

Several studies in recent years have therefore focused on deriving and validating prognostic scores for detecting early mortality after an ICH in the acute setting. This is particu-

Received September 23, 2014
Revised March 7, 2015
Accepted March 9, 2015

Correspondence
Yoon Kong Loke, MD
Health Evidence Synthesis Group, Norwich Medical School, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK
Tel +44 1603 591234
Fax +44 1603 593752
E-mail y.loke@uea.ac.uk

larly pertinent given that the risk of a poor outcome is higher for ICH than for the other stroke subtypes,[4] and the use of a prognostic model has been found to confer greater accuracy than merely relying on clinical judgment.[5] In the absence of well-established interventions to reduce deaths from ICH, accurate prognostic tools may prove useful for informed decision-making in the acute phase of ICH, including the options of transferring to intensive care, rehabilitation, and palliation. In the research setting, prognostic scores may also prove useful for the risk stratification of participants in clinical trials of interventions for ICH.

Published systematic reviews of prognostic models in ICH date back at least 10 years,[6,7] and the only recent systematic review that we are aware of was reported in a conference abstract in 2010, and has not been reported elsewhere in more detail.[8] A comprehensive update seems timely given 1) the recent publication of new studies that have evaluated different prognostic scores and 2) the absence of a unified system that is accepted in routine clinical practice.

Hence, in the present study we aimed to synthesize the recent evidence on prognostic tools in patients presenting with ICH, and to determine the comparative performances of different scores.

METHODS

Eligibility criteria

We selected studies that collected clinical variables (or sets of these variables) used to calculate risk scores predicting early mortality (<6 months) in adult patients at the time of presentation with ICH. We stipulated that studies had to have a sample size of >100 participants, with the main focus being on those presenting with primary ICH. Our specific interest was the reporting of the discrimination ability of the tool, measured based on the area under the receiver operating characteristic curve (AUC) or c-statistic. We aimed to base our systematic review on more-up-to-date evidence, and as such restricted our selection to the past decade; that is, studies published from 2004 to April 2014.

We excluded studies that were aimed solely at determining correlations between mortality and single laboratory (e.g., albumin or troponin) or radiological (e.g., lesion volume) variables. We did not include studies of only functional outcomes. Since our main focus was on stroke patients presenting to healthcare facilities and we wanted to maximize the generalizability of the findings, we excluded studies involving narrow subgroups of ICH patients who had been deemed to require admission to intensive care. We also excluded studies that focused on mortality in specific subsets of patients (e.g., following certain interventions such as thrombolysis) or those that specifically evaluated the prognosis of a stroke affecting a particular brain area (e.g., basal ganglia).

Search strategy

We searched MEDLINE and EMBASE (in April 2014, using the OvidSP interface) using the search terms listed in Supplementary (in the online-only Data Supplement), without any language restriction. We also checked the bibliographies of the included studies for other potentially suitable studies.

Study selection and data extraction

Study screening and data extraction were performed by pairs of reviewers (selected from K.M., C.S.K., K.P., and Y.K.L.) who independently scanned all titles and abstracts for potentially relevant articles, whose full-text versions were retrieved for further detailed evaluation. Any uncertainties and discrepancies were resolved through discussion and with a third reviewer. We also contacted authors if any aspects of their articles required further clarification.

We used a standardized form for data collection that included details of the setting and date of the study, geographical location, selection criteria, and other characteristics of the participants, and outcome measures.

Assessment of study validity

Study validity was assessed by pairs of reviewers independently checking whether there was clear reporting of the times of patient assessments, missing or incomplete data, use of standardized treatment protocols, and whether the study involved a derivation or validation cohort.

Data analysis

Statistical analysis was conducted by an experienced meta-analyst (Y.K.L.) using Cochrane Collaboration RevMan 5.3 software (Nordic Cochrane Centre, Kobenhavn, Denmark).

We chose to base our analysis on the AUC or c-statistic since these are equivalent measures of the discrimination ability for binary outcomes.[9] In the present context the discrimination ability refers to how well the model separates patients who subsequently die from those who are survivors. For studies that investigated both derivation and validation components, we chose to analyze data relating to the validation portion. If different mortality time points were used in a particular study, we used 30 days as the first choice and inpatient mortality as the second choice, and where neither was available we accepted a time point of <6 months for analysis. If multiple AUC values were available for a particular prognostic tool, we calculated a weighted pooled average using a random-effects inverse-variance meta-analysis. If the AUCs were listed without standard errors, we deter-

mined these values through Hanley's method and the 95% confidence intervals (CIs).[10]

We assessed heterogeneity using the I^2 statistic and by visual inspection of Forest plots. The performance of the prognostic score was judged according to the following AUC thresholds that have been described by other researchers: excellent (AUC ≥0.90), good (AUC ≥0.80 and <0.90), fair (AUC ≥0.70 and <0.80), and poor (AUC <0.70).[11]

RESULTS

We selected 11 relevant studies from 2,603 articles identified by searching the electronic databases (the flow chart of study selection is shown in Fig. 1).[4,5,12-20] The characteristics and results of the included studies are reported in Table 1, and our appraisal of study validity is presented in Table 2. Variables required for the calculation of each prognostic model are listed in Table 3.

The included studies involved 41,555 participants (sample sizes ranged from 154 to 37,509 in the 11 studies) with a mean age of 67 years, while 55% of them were male. Six of the studies addressed the 30-day mortality, three addressed inpatient mortality,[13,19,20] and two addressed mortality at 90 or 120 days.[4,5] Four studies recruited patients from two or more healthcare sites.[4,5,12,19] The geographical locations were diverse, and included North America, Europe, Mexico, and East Asia. Data from the study performed in Taiwan were reported in two separate articles, with Hemphill-ICH scores available from one and ICH-Grading Scale (GS) scores from the other, with substantial overlap in the included patients.[16,21]

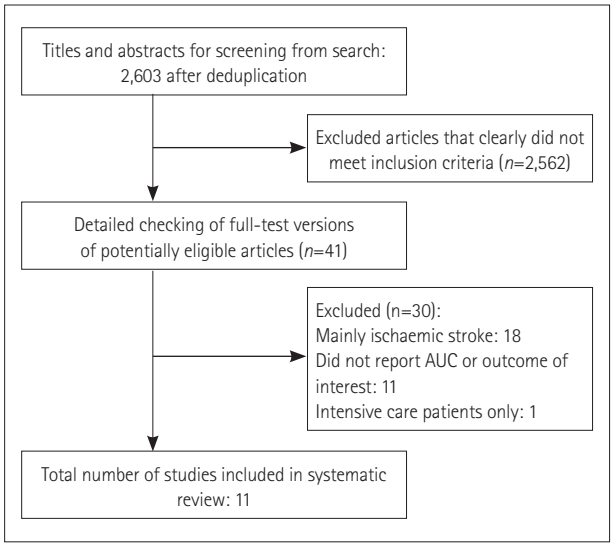

Fig. 1. Flow chart of study selection. AUC: area under receiver operating characteristic curve.

Validity assessment

Most of the studies had a retrospective design, or performed post-hoc analyses of prospectively collected clinical data. We were able to determine that the prognostic variables were collected early in the course of the presentation in five studies.[5,16-19] Treatment pathways were seldom reported, with only one study explicitly stating that all participants received similar care.[5] Details on losses to follow-up or missing data were reported for seven studies (Table 2).[4,5,12,14,15,17,18] Eight studies aimed to perform model validation, two had a mixed derivation-validation design,[18,20] and one that had a purely derivation design.[13] We considered the findings of these derivation studies to be less robust than those that had been submitted for external validation.

Quantitative comparison of AUC

We were able to evaluate the following prognostic models in the comparative quantitative analysis: Hemphill-ICH[22] (nine cohorts)[4,5,12,14-18,20] and ICH-GS (four cohorts).[4,15,16,18]

The AUCs from individual studies and the pooled mean AUCs across studies are shown in Fig. 2.

Hemphill-ICH score

The predictive accuracy of the Hemphill-ICH model for mortality has been evaluated in 9 cohorts comprising 3,819 participants worldwide.[4,5,12,14-18,20] Point estimates of the AUC ranged from 0.72 to 0.88, with a weighted pooled average of 0.80 (95% CI=0.77−0.85) across all studies. Subgroup evaluation according to the mortality time point found that data from six cohorts were for the 30-day mortality (pooled AUC=0.81, 95% CI=0.76−0.86),[12,14-18] while two were for the 90- or 120-day mortality (pooled AUC=0.79, 95% CI=0.70−0.88).[4,5] When we excluded the study with the lowest 30-day mortality rate, the overall AUC remained good at 0.82 (95% CI=0.78−0.85).

We conducted subgroup analyses looking at the prognostic value of the Hemphill-ICH model according to study design and patient characteristics (e.g., age and geographical location). We found that Hemphill-ICH scores generally performed well across different subgroups (Fig. 3), but there was a possible slight decrease in performance in those studies conducted outside of North America and Europe, or in those where the participants were on average younger than 70 years.

ICH-GS score

The performance of the ICH-GS score in predicting mortality has been evaluated in four cohorts comprising participants in the US, Spain, Taiwan, and the UK. Point estimates of the AUC ranged from 0.74 to 0.88, with a weighted pooled aver-

Table 1. Characteristics of included studies

Study ID	Study setting; period	Study design; name of score	Patients, *n*	Age, years	Males, %	Mortality rate	AUC and AUC category (excellent/good/fair/poor)
Clarke 2004[12]	Two centers, USA; 1998–2000	Retrospective, validation; ICH	175	70	54.3	30 days: 40%	Hemphill-ICH: AUC=0.88 (good)
Matchett 2006[14]	Single hospital, USA; 1998–2002	Retrospective, validation; ICH	241	72	52	30 days: 33%	Hemphill-ICH: AUC=0.814 (good) (SE=0.031), sens=66%, spec=87%; Broderick: AUC=0.773 (fair) (SE=0.036), sens=45%, spec=92%
Takahashi 2006[20]	Single hospital, Japan; 1998–2001	Retrospective, derivation, validation; ICH, CART	347	71.7	49	Inpatient: 20.2%	CART: AUC=0.86 (good); Hemphill-ICH: AUC=0.83 (good)
Weimar 2006[5]	Multiple centers, Germany; 2000–2002	Prospective, validation; Essen-ICH, ICH, Cheung-ICH	371	67	56	120 days: 29.1%	Essen-ICH: AUC=0.831 (good) (95% CI=0.784–0.878), sens=43.9%, spec=97.7%; Hemphill-ICH: AUC=0.831 (good) (95% CI=0.783–0.878), sens=58.5%, spec=93.1%; Cheung-ICH: AUC=0.835 (good) (95% CI=0.787–0.882), sens=64.2%, spec=85.8%
Romano 2007[17]	Single hospital, Spain; 2003–2006	Prospective, validation; ICH	154	Unclear	48	30 days: 41%	Hemphill-ICH: AUC=0.736 (fair) (SE=0.042); Score ≥3: sens=73%, spec=91%
Ruiz-Sandoval 2007[18]	Single hospital, Mexico; 1999–2003	Prospective, derivation, validation; ICH, ICH-GS	378	64.2	50	30 days: 57%	ICH-GS: AUC=0.88 (good) (95% CI=0.85–0.92); ICH: AUC=0.83 (good) (95% CI=0.79–0.88); ICH-GS: sens=78.2% Hemphill-ICH: sens=63.8%
Peng 2010[16] and Chuang 2009[21]	Single hospital, Taiwan; 2006–2008	Retrospective, validation; ICH variants	423	61	67	30 days: 14.7%	Hemphill-ICH: AUC=0.72 (fair) (95% CI=0.68–0.76), sens=64.5%, spec=71.2%; ICH-GS: AUC=0.74 (fair) (95% CI=0.65–0.83)
Garrett et al. 2013[4]	Emergency departments, USA; 2009–2011	Retrospective, validation; ICH-GS, ICH, FUNC	366	Median <70	51	90 days: 38%	FUNC: AUC=0.87 (good) (95% CI=0.83–0.91); ICH-GS: AUC=0.88 (good) (95% CI=0.85–0.92); Hemphill-ICH: AUC=0.74 (fair) (95% CI=0.69–0.79)
Li 2012[13]	Single hospital, China; 2008-2009	Retrospective, derivation; ICH Index	227	64 (deceased), 58 (alive)	63	Inpatient: 21.6%	ICH Index: AUC=0.923 (excellent) (95% CI=0.883–0.963), sens=65%, spec=95%
Parry-Jones 2013[15]	Single hospital, UK; 2008–2010	Prospective, validation; GCS, ICH variants	1,364	73	53	30 days: 41.1%	Hemphill-ICH: AUC=0.861 (good) (95% CI=0.840–0.880); ICH-GS: AUC=0.874 (good) (95% CI=0.853–0.892); Modified ICH: AUC=0.824 (good) (95% CI=0.801–0.845); GCS: AUC=0.874 (good) (95% CI=0.853–0.892)
Smith 2013[19]	Multiple centers, US and Canada; 2001–2007	Retrospective, validation; GWTG, NIHSS	37,509 NIHSS scores available: 10,352	73	49	Inpatient: 27%	GWTG alone: AUC=0.66 (poor); GWTG+NIHSS: AUC=0.82 (good)

AUC thresholds: excellent (AUC ≥0.90), good (AUC ≥0.80 and <0.90), fair (AUC ≥0.70 and <0.80), and poor (AUC <0.70).
AUC: area under receiver operating characteristic curve, CART: classification and regression trees, CI: confidence interval, FUNC: functional outcome risk stratification scale, GCS: Glasgow Coma Scale, GTWG: Get With The Guidelines, NIHSS: national Institutes of Health Stroke Scale, SE: standard error, sens: sensitivity, spec: specificity.

Table 2. Validity assessment of included studies

Study ID	Was the index/score obtained early in the presentation course?	Did authors give numbers of patients lost to follow-up or withdrawals, and the reasons?	Did authors state if the patients were treated on a standardized or similar care pathway?	Validation study	Amount of missing data
Clarke 2004[12]	No	Yes	No	Yes	4 patients with no follow-up
Garrett 2013[4]	No	Yes	No	Yes	31 patients with no follow-up
Li 2011[13]	No	No	No	No	
Matchett 2007[14]	No	Yes	No	Yes	66 patients with missing data
Parry-Jones 2013[15]	No	Yes	No	Yes	13 patients could not be traced
Peng 2010[16] and Chuang 2009[21]	Yes	No	No	Yes	
Romano 2007[17]	Yes	Yes	No	Yes	7 patients excluded
Ruiz-Sandoval 2007[18]	Yes	Yes	No	Derivation of ICH-GS	68 patients with no follow-up at 30 days
Smith 2013[19]	Yes	No	No	Yes	
Takahashi 2006[20]	No	No	No	Derivation of CART	
Weimar 2006[5]	Yes	Yes	Yes	Yes	75 patients with no follow-up

CART: classification and regression trees, ICH-GS: intracerebral hemorrhage-Grading Scale.

age of 0.87 (95% CI=0.84–0.90).[4,15,16,18]

Hemphill-ICH versus ICH-GS scores

An overall comparison of the two subgroups using the Hemphill-ICH and ICH-GS scores showed that the AUC differed significantly (p=0.01) between these two subgroups, with the ICH-GS score exhibiting better overall discrimination performance (Fig. 2). We also assessed four studies that evaluated both the Hemphill-ICH and ICH-GS scores in the same sample of participants.[4,15,16,18] The greatest difference in the comparative predictive accuracies of the Hemphill-ICH and ICH-GS models was seen in the cohort of Garrett et al.,[4] with a reported difference of 0.14 in the AUC, favoring the ICH-GS model. In contrast, the other three studies demonstrated far smaller absolute differences in AUC, with an average difference of 0.03 that was also in favor of the ICH-GS model.

Get With The Guidelines (GWTG) model with or without National Institutes of Health Stroke Scale (NIHSS)

Only one study analyzed the performance of the GWTG score for predicting inpatient mortality,[19] that study enrolled 37,509 participants in US and Canada. The GWTG alone (based on age, vascular risk factors, comorbid conditions, and mode of arrival at the hospital) does not require a detailed clinical examination or neuroimaging, but in that study it demonstrated a relatively poor predictive accuracy with an AUC of 0.66. However, combining the GWTG model with the NIHSS (in 10,352 participants) resulted in a markedly improved pooled AUC of 0.82.

Glasgow Coma Scale (GCS)

We identified only one recent study that involved a prognostic validation of the GCS.[15] That study recruited 1,364 participants in the UK and found an AUC of 0.87 (95% CI=0.85–0.89), which was similar to the AUC values obtained by applying the Hemphill-ICH and ICH-GS models to the same participants.

Functional outcome risk stratification scale (FUNC)

Garrett et al.[4] validated the prognostic accuracy of the FUNC score in 366 patients in the US, and found an AUC of 0.87 (95% CI=0.83–0.90) for 90-day mortality. The AUC for the FUNC indicated similarly good predictive accuracy to that reported with the ICH-GS score in the same participants, whereas the Hemphill-ICH score only exhibited fair performance in that cohort.

Essen-ICH score

The Essen-ICH score was validated in 1 study involving 371

Table 3. Variables required for estimating the prognostic score

Predictor	Hemphill-ICH	Essen-ICH	ICH-GS	FUNC
Age, years	≥80=1 <80=0	<60=0 60–69=1 70–79=2 ≥80=3	<45=1 45–64=2 ≥65=3	<70=2 70–79=1 ≥80=0
Pre-ICH cognitive impairment		-	-	No=1 Yes=0
GCS score	3 or 4=2 5–12=1 13–15=0	-	13–15=1 9–12=2 3–8=3	≥9=2 ≤8=0
Hemorrhage volume, mL	≥30=1 <30=0	-	Supratentorial <40=1 40–70=2 >70=3 Infratentorial <10=1 10–20=2 >20=3	<30=4 30–60=2 >60=0
ICH location	Infratentorial Yes=1 No=0	-	Supratentorial=1 Infratentorial=2	Lobar=2 Deep=1 Infratentorial=0
Extension into ventricles	Yes=1 No=0	-	Yes=1 No=0	-
NIHSS neurological examination score		0–5=0 6–10=1 11–15=2 16–20=3 >20 or coma=4	-	-
NIHSS level of consciousness		Alert=0 Drowsy=1 Stupor=2 Coma=3		
Maximum score	6	10	13	11

FUNC: functional outcome risk stratification scale, GCS: Glasgow Coma Scale, ICH: intracerebral hemorrhage, NIHSS: National Institutes of Health Stroke Scale.

German patients.[5] The AUC of 0.83 (95% CI=0.78–0.88) for the Essen-ICH model was similar to those obtained when applying the Hemphill-ICH and Cheung-ICH models to the same sample of patients.

Other prognostic models

We identified data for two prognostic tools [classification and regression trees (CART) and ICH Index] that have yet to be validated but which showed good predictive value (AUCs of 0.86 and 0.92, respectively) during derivation.[13,20] The CART model is based only on three variables—size of hemorrhage, age, and score on the Japan Coma Scale[20]—whereas the ICH Index was constructed based on age, GCS score, glucose level, and white cell count but without the use of neuroimaging.[13]

DISCUSSION

ICH is associated with the highest morbidity and mortality of all types of stroke. We have systematically evaluated recent data obtained by applying several models to predict mortality in ICH, and found the Hemphill-ICH score to have the broadest evidence base. This score has exhibited generally consistent predictive accuracy throughout several studies worldwide covering thousands of patients in eight countries (China, Japan, Germany, the US, Mexico, Argen-

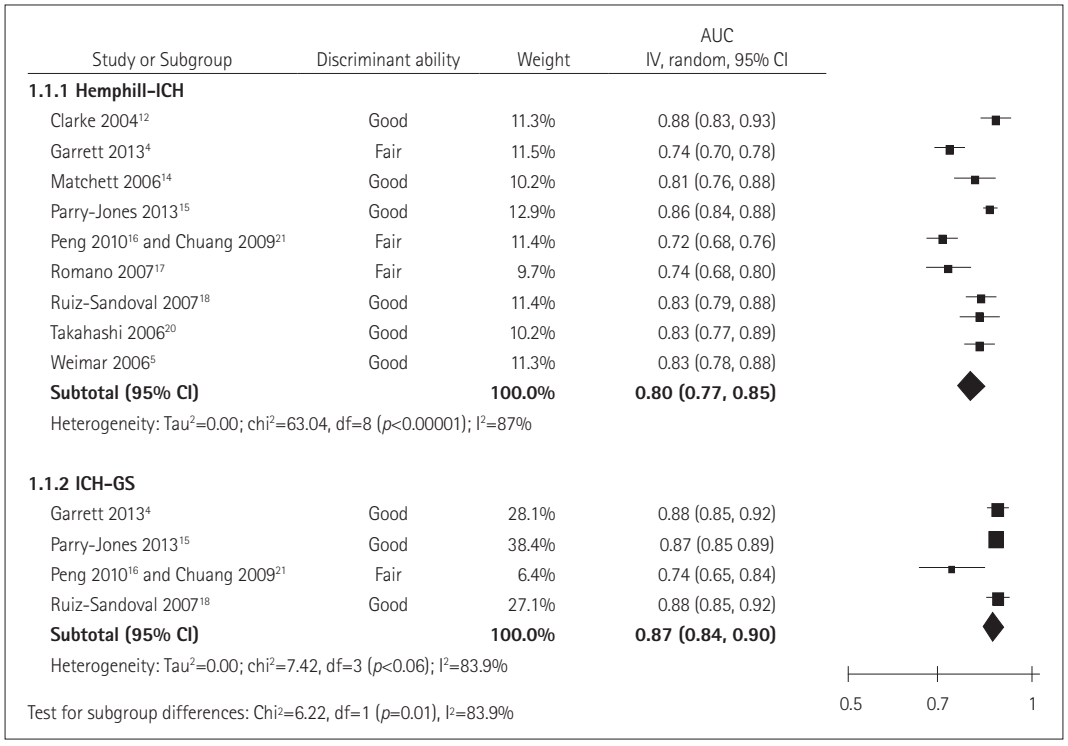

Fig. 2. Meta-analysis of the areas under the receiver operating characteristic curve (AUC) for various prognostic models. CI: confidence interval, ICH: intracerebral hemorrhage.

tina, Taiwan, and the UK). Although the Hemphill-ICH score was introduced more than 10 years ago, it is not yet widely adopted in clinical practice. Instead, we found numerous instances where researchers have modified the Hemphill-ICH score to try and improve its predictive accuracy, with varying degrees of success. The availability of several versions of the ICH score can seem bewildering—an important finding of our systematic review is that the ICH-GS score seems the one most likely to offer some consistent advantage over the original Hemphill-ICH score. The slightly improved performance when using the ICH-GS score may stem from the greater detail with which the site and size of the hemorrhage are considered, as well as the inclusion of additional age categories (Table 3). However, we recognize that these changes may make the ICH-GS score more complicated to calculate in practice.

We also identified variations in the complexity and in the requirement for specialist knowledge when using some of the tools (e.g., reproducibility when interpreting hemorrhage volume on CT scans and calculation of subscores such as the NIHSS). The need for specialist expertise may prove to be a barrier in emergency departments where clinicians may prefer a tool that is simply based on clinical variables, such as the GCS and the (as yet unvalidated) ICH Index.[13,15] Indeed, Parry-Jones et al.[15] found that the AUC of the GCS score was as good as that of the ICH score in a UK validation cohort, but we were unable to identify other recent data sets for confirming the generalizability of these findings. This is an interesting point, since the GCS score can be rapidly assessed at the initial presentation and does not require specialist neurological imaging procedures or expertise. Further validation studies of the GCS score and ICH Index would be useful, particularly in resource-poor areas or as initial triage tools in nonspecialized healthcare facilities.

Several prognostic models are associated with additional complexity due to them requiring a detailed neurological examination to estimate the NIHSS score.[5,19] For instance, the GWTG model exhibited a poor AUC score (<0.7) when it was applied alone, but this improved to a good AUC score when it was combined with the NIHSS.[19] Having to use both GWTG and NIHSS scores together may prove too laborious for clinicians, particularly given that dedicated online training is required for calculating the NIHSS score.[23] The Essen-ICH score also requires calculation of the NIHSS score and this might equally limit its acceptability, particularly given that a previous study found no marked improvement in AUC over that for the Hemphill-ICH score.[5]

Most of the available studies have not addressed the acceptability and uptake of current prognostic scores in the day-to-day management of stroke patients. While the availability of a prediction rule with good performance is an important prerequisite, patients will not benefit from the pro-

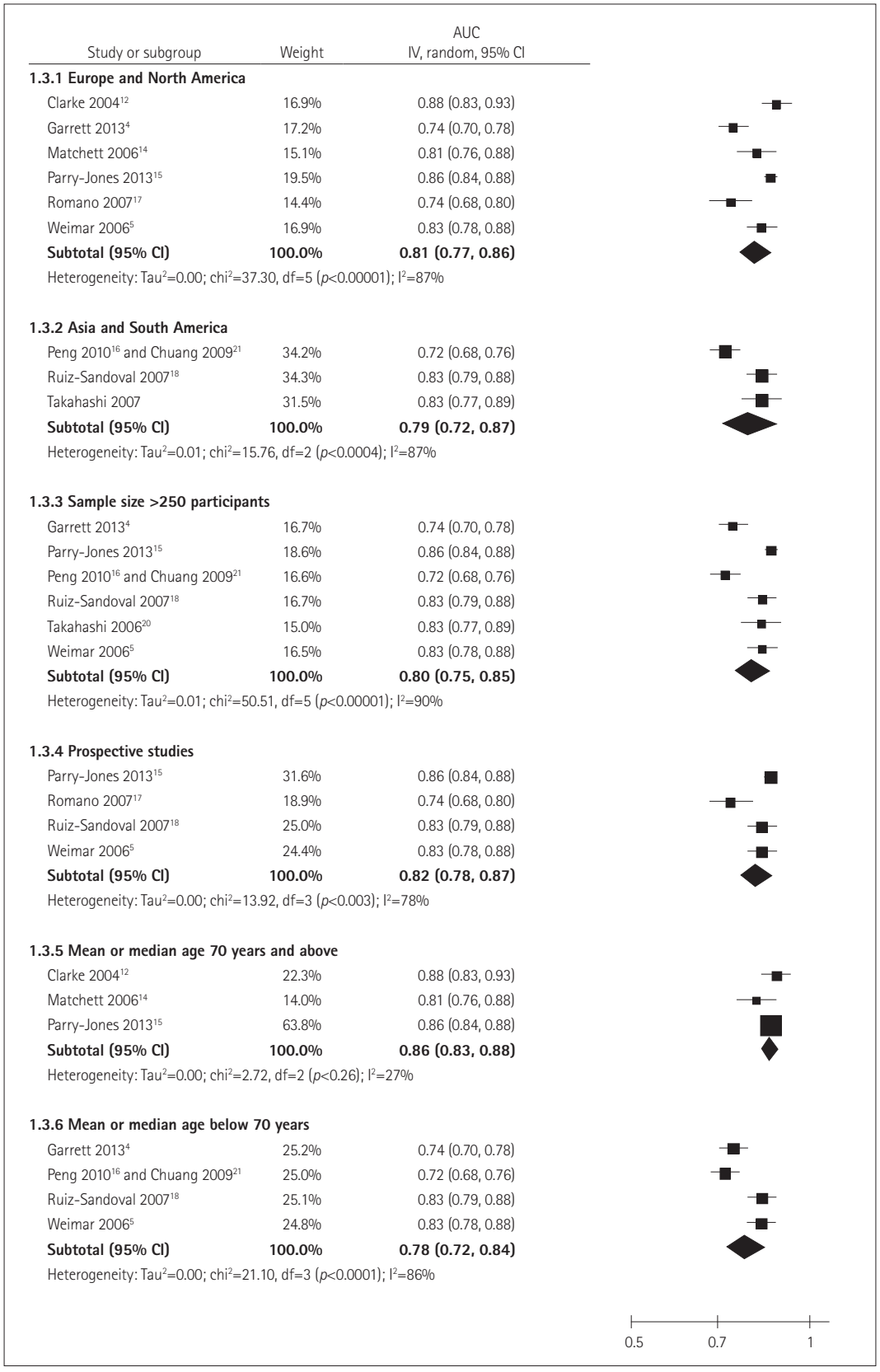

Fig. 3. Subgroup analyses of the Hemphill-intracerebral hemorrhage model according to the study design and characteristics of participants. CI: confidence interval.

liferation of prognostic scoring models if their uptake and implementation is patchy. It is important to determine what clinicians want or expect from a score and what factors would facilitate their use of it. Furthermore, the expectations of patients and their relatives also need to be considered, such as by determining whether prognostic scoring is acceptable and useful to interactions (as compared to relying on clinical judgment). Shared decision-making is pivotal in modern medicine, but our systematic review shows that none of the current prognostic models are able to achieve excellent performance, and thus the acceptability of imperfect results needs to be assessed. We note that a survey found that 96% of emergency physicians were prepared to use a prognostic tool for stroke or death in patients with transient ischemic attacks, but only if the tool achieved a sensitivity of >97%.[24]

Our systematic review has limitations. We focused only on larger studies (>100 participants) published during the last 10 years, and emphasized overall mortality—because of the high rate of early mortality in ICH—rather than the functional outcome. Most of the included studies had a retrospective design or were post-hoc analyses of prospectively collected clinical data, and we did not categorize the studies into high- and low-quality subgroups. We selected published studies that used the AUC or c-statistic as their primary measure, and it is possible that studies that found poor performance have not been reported on.

The strengths of our systematic review are that we conducted an exhaustive and up-to-date search of the current evidence, accompanied by critical appraisal and quantitative data analysis. To the best of our knowledge, none of the previous systematic reviews have performed a meta-analysis of discrimination ability. We have summarized the evidence for the relative performances from comprehensive data sets to help guide stroke researchers and clinicians as to which score to use, further develop, or test.

A key question to consider is whether we genuinely need further research that might involve only minor modifications to the Hemphill-ICH model, and which may not provide more than minor incremental benefits to the clinical accuracy. The proliferation of variants of the Hemphill-ICH model may simply cause greater confusion amongst clinicians and thereby have a detrimental effect on clinical implementation. Future studies should focus on the factors that influence the acceptability and adoption of scoring systems, and whether their implementation leads to consistent improvements in patient care relative to simply using subjective clinical judgment.

In conclusions, we have highlighted several prognostic scores that exhibit good performance in ICH, the front runners being the Hemphill-ICH score and the ICH-GS variant, which we believe can usefully guide clinicians in making better-informed treatment decisions. Although further validation studies are needed, the GCS and ICH Index may also be reasonable options in situations where simple and rapid clinical assessment is needed before neuroimaging results become available, such as during triage when a patient initially presents to a healthcare facility.

Supplementary Materials

The online-only Data Supplement is available with this article.

Conflicts of Interest

The authors have no financial conflicts of interest.

Acknowledgements

K.M., C.S.K., Y.K.L., and P.K.M. conceptualized the review and developed the protocols. K.M., C.S.K., Y.K.L., K.P., and L.A. selected studies and abstracted the data. K.M. and Y.K.L. carried out the synthesis of the data and wrote the manuscript while receiving critical input from all authors. Y.K.L. acts as guarantor for the paper.

REFERENCES

1. Mackay J, Mensah GA. *Atlas of heart disease and stroke. 15 Global burden of stroke [Internet]*. Geneva: World Health Organization; 2004 [cited 2014 Apr 8]. Available from: URL: www.who.int/cardiovascular_diseases/en/cvd_atlas_15_burden_stroke.pdf.
2. Thrift AG, Dewey HM, Macdonell RA, McNeil JJ, Donnan GA. Incidence of the major stroke subtypes: initial findings from the North East Melbourne stroke incidence study (NEMESIS). *Stroke* 2001;32:1732-1738.
3. Krishnamurthi RV, Feigin VL, Forouzanfar MH, Mensah GA, Connor M, Bennett DA, et al. Global and regional burden of first-ever ischaemic and haemorrhagic stroke during 1990-2010: findings from the Global Burden of Disease Study 2010. *Lancet Glob Health* 2013;1:e259-e281.
4. Garrett JS, Zarghouni M, Layton KF, Graybeal D, Daoud YA. Validation of clinical prediction scores in patients with primary intracerebral hemorrhage. *Neurocrit Care* 2013;19:329-335.
5. Weimar C, Benemann J, Diener HC; German Stroke Study Collaboration. Development and validation of the Essen Intracerebral Haemorrhage Score. *J Neurol Neurosurg Psychiatry* 2006;77:601-605.
6. Ariesen MJ, Algra A, van der Worp HB, Rinkel GJ. Applicability and relevance of models that predict short term outcome after intracerebral haemorrhage. *J Neurol Neurosurg Psychiatry* 2005;76:839-844.
7. Counsell C, Dennis M. Systematic review of prognostic models in patients with acute stroke. *Cerebrovasc Dis* 2001;12:159-170.
8. Agarwal D, Surana A, Chandra S. P-152. Prognostic accuracy of the iCh score on initial evaluation of patients presenting with non-traumatic intracerebral hemorrhage; a systematic review and meta-analysis. *Can J Neurol Sci* 2010;37(3 Suppl 1):S83-S84.
9. Austin PC, Steyerberg EW. Interpreting the concordance statistic of a logistic regression model: relation to the variance and odds ratio of a continuous explanatory variable. *BMC Med Res Methodol* 2012;12:82.
10. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29-36.
11. Vellutino FR, Scanlon DM, Zhang H. 14. Identifying reading disability based on response to intervention: evidence from early intervention research. In: Jimerson SR, Burns MK, Van Der Heyden AM. *Handbook of response to intervention: the science and practice of assessment and intervention*. New York: Springer, 2007;194.

12. Clarke JL, Johnston SC, Farrant M, Bernstein R, Tong D, Hemphill JC 3rd. External validation of the ICH score. *Neurocrit Care* 2004;1:53-60.
13. Li YF, Luo J, Li Q, Jing YJ, Wang RY, Li RS. A new simple model for prediction of hospital mortality in patients with intracerebral hemorrhage. *CNS Neurosci Ther* 2012;18:482-486.
14. Matchett SC, Castaldo J, Wasser TE, Baker K, Mathiesen C, Rodgers J. Predicting mortality after intracerebral hemorrhage: comparison of scoring systems and influence of withdrawal of care. *J Stroke Cerebrovasc Dis* 2006;15:144-150.
15. Parry-Jones AR, Abid KA, Di Napoli M, Smith CJ, Vail A, Patel HC, et al. Accuracy and clinical usefulness of intracerebral hemorrhage grading scores: a direct comparison in a UK population. *Stroke* 2013;44:1840-1845.
16. Peng SY, Chuang YC, Kang TW, Tseng KH. Random forest can predict 30-day mortality of spontaneous intracerebral hemorrhage with remarkable discrimination. *Eur J Neurol* 2010;17:945-950.
17. Romano LM, Ioli P, Gonorazky S. Predicting mortality and externally validation of the original intracerebral haemorrhage score after spontaneous intracerebral haemorrhage. *Revista Neurológica Argentina* 2007;32:94-99.
18. Ruiz-Sandoval JL, Chiquete E, Romero-Vargas S, Padilla-Martínez JJ, González-Cornejo S. Grading scale for prediction of outcome in primary intracerebral hemorrhages. *Stroke* 2007;38:1641-1644.
19. Smith EE, Shobha N, Dai D, Olson DM, Reeves MJ, Saver JL, et al. A risk score for in-hospital death in patients admitted with ischemic or hemorrhagic stroke. *J Am Heart Assoc* 2013;2:e005207.
20. Takahashi O, Cook EF, Nakamura T, Saito J, Ikawa F, Fukui T. Risk stratification for in-hospital mortality in spontaneous intracerebral haemorrhage: a Classification and Regression Tree analysis. *QJM* 2006;99:743-750.
21. Chuang YC, Chen YM, Peng SK, Peng SY. Risk stratification for predicting 30-day mortality of intracerebral hemorrhage. *Int J Qual Health Care* 2009;21:441-447.
22. Hemphill JC 3rd, Bonovich DC, Besmertis L, Manley GT, Johnston SC. The ICH score: a simple, reliable grading scale for intracerebral hemorrhage. *Stroke* 2001;32:891-897.
23. Bushnell CD, Johnston DC, Goldstein LB. Retrospective assessment of initial stroke severity: comparison of the NIH Stroke Scale and the Canadian Neurological Scale. *Stroke* 2001;32:656-660.
24. Perry JJ, Goindi R, Brehaut J, Taljaard M, Schneider S, Stiell IG. Emergency physicians' management of transient ischemic attack and desired sensitivity of a clinical decision rule for stroke in three countries. *CJEM* 2011;13:19-27.

Supplementary Material

Search Strategy

Interface: OvidSP

[(stroke OR intracranial-hemorrhage OR intracerebral-hemorrhage) AND score AND (prognostic OR prognosis OR predicti*) AND (mortality OR death OR survival)].mp

For this search, .mp includes the fields of title, Abstract, subject headings, heading words, original title, drug or device manufacturer, trade name, keyword, keyword heading word, and unique identifier.

We checked the bibliographies of included articles for any additional relevant studies.

We contacted authors for more information if there were any uncertainties when reviewing the articles.

We used online translation tools if there were any foreign language articles that we were unable to translate ourselves.