Poor utility of grading scales in acute intracerebral hemorrhage: results from the INTERACT2 trial

Emma Heeley[1], Craig S. Anderson[1,2]*, Mark Woodward[1,3], Hisatomi Arima[1,4], Thompson Robinson[5], Christian Stapf[6], Mark Parsons[7], Pablo M. Lavados[8,9], Yining Huang[10], Yanxia Wang[11], Sophie Crozier[12], Adrian Parry-Jones[13], Jiguang Wang[14], and John Chalmers[1] for the INTERACT2 Investigators†

Background **Several simple clinical grading scores have been developed for intracerebral hemorrhage, primarily to predict 30-day mortality.**

Aims **We aimed to determine the accuracy of three popular scores (original intracerebral hemorrhage, modified intracerebral hemorrhage, and intracerebral hemorrhage grading scale) on 30-day mortality and 90-day death or major disability, and whether the magnitude of benefit varies according to prognosis graded by the three predictive scores.**

Methods **Data from the Intensive Blood Pressure Reduction in Acute Cerebral Hemorrhage Trial which included 2839 intracerebral hemorrhage patients (<6 hours) and elevated systolic blood pressure (150–220 mmHg), randomized to intensive (target systolic blood pressure <140 mmHg) or guideline-based (<180 mmHg) blood pressure management. Discrimination of scales for predicting death and poor outcome (modified Rankin scale 3–6) was evaluated in area under receiver operator characteristic curves.**

Results **Among 2556 (90%) participants with available data, the modified intracerebral hemorrhage had the highest discrimination (receiver operator characteristic 0·75) for 90-day poor outcome compared with the original intracerebral hemorrhage (receiver operator characteristic 0·68) and intracerebral hemorrhage grading scale (receiver operator characteristic 0·69). All scores had good positive predictive value (approximately 80–90%) for poor outcome but poor sensitivity and positive predictive value for death. The scores do not clearly discriminate a patient group most likely to benefit from blood pressure lowering.**

Conclusions **Intracerebral hemorrhage prognostic scores are not useful in defining patients at high probability of early death, but they are reliable for predicting poor outcome, defined by death or major disability. Potential benefits of early intensive blood pressure lowering are broadly applicable across grades of severity defined by such scores.**

Key words: BP lowering treatment, ICH scales, INTERACT2, intracerebral hemorrhage, prognosis

Correspondence: Craig S. Anderson*, The George Institute for Global Health, University of Sydney, PO Box M201, Missenden Road, Camperdown, Sydney, NSW 2050, Australia.
E-mail: canderson@georgeinstitute.org.au
[1]The George Institute for Global Health, University of Sydney, Sydney, NSW, Australia
[2]Royal Prince Alfred Hospital, Sydney, NSW, Australia
[3]The George Institute for Global Health, University of Oxford, Oxford, UK
[4]Seta University of Medical Science, Tsukinowa-cho, Otsu, Shiga, Japan
[5]Department of Cardiovascular Sciences and NIHR Biomedical Research Unit in Cardiovascular Disease, University of Leicester, Leicester, UK
[6]Department of Neurology, APHP – Hôpital Lariboisière and DHU NeuroVasc Paris – Sorbonne, Univ Paris Diderot – Sorbonne Paris Cité, Paris, France
[7]Department of Neurology, John Hunter Hospital, University of Newcastle, Newcastle, NSW, Australia
[8]Servicio de Neurología, Departamento de Medicina, Clínica Alemana, Centro de Neurociencias, Universidad del Desarrollo, Santiago, Chile
[9]Departamento de Ciencias Neurológicas, Universidad de Chile, Santiago, Chile
[10]Department of Neurology, Peking University First Hospital, Beijing, China
[11]Hejian City People's Hospital, Hebei, China
[12]Stroke unit, Pitié-Salpêtriere Hospital, Paris, AP-HP, France
[13]Salford Royal NHS Foundation Trust, Salford, UK
[14]The Shanghai Institute of Hypertension, Rui Jin Hospital, Shanghai Jiaotong University, Shanghai, China

Received: 9 November 2014; Accepted: 18 February 2015; Published online 4 June 2015

†For a full list of investigators, see Anderson *et al.* (2).

Conflict of interest: Dr Huang reports receiving reimbursement for travel expenses from Osaka Pharmaceuticals. Dr Jiguang Wang is receiving consulting fees from Novartis, Omron Healthcare, Pfizer, and Takeda; grant support from Novartis, Omron Healthcare, and Pfizer; lecture fees from A&D Pharma, Omron Healthcare, Novartis, Pfizer, and Servier; and reimbursement for travel expenses from Pfizer and Takeda. Dr Anderson reports receiving speaker fees from Takeda China and Covidien, and Advisory Board fees from The Medicine Company and Pfizer. The other authors declare that they have no conflicts of interest.

Clinical Trial Registration-URL: http://clinicaltrials.gov/show/NCT00716079. Unique identifier: NCT00716079.

Funding: The INTERACT2 study was supported by grants from the National Health and Medical Research Council (NHMRC) of Australia. Professor Anderson holds an NHMRC senior principal research fellowship.

DOI: 10.1111/ijs.12518

Nontraumatic intracerebral hemorrhage (ICH) is the least treatable and most devastating form of acute stroke, causing more deaths than ischemic stroke worldwide (1). The recently completed Intensive Blood Pressure Reduction In Acute Cerebral Hemorrhage Trial (INTERACT2) demonstrated that early intensive blood pressure (BP) lowering to a systolic (SBP) target of <140 mmHg within one-hour improved functional recovery and health-related quality of life in survivors of ICH (2). Although the results provide evidence supporting a safe and widely applicable management policy for this condition, the treatment had only a modest effect, did not influence mortality, and was derived from patients with predominantly mild forms of ICH. Thus, uncertainty persists over the benefits of early intensive BP lowering in patients with severe ICH or in those with the gravest prognosis.

Several simple clinical scoring systems have been developed for estimating prognosis in patients with ICH, mainly for predicting early death (3–5) but also functional outcome (Table 1) (6). One of the first simple and the best known is the original ICH scale

Table 1 ICH grading scales

Original ICH scale		Modified ICH scale		ICH-GS scale	
Characteristic	Points	Characteristic	Points	Characteristic	Points
Age	Years	Age	Years	Age	Years
<80	0	<80	0	<45	1
≥80	1	≥80	1	45–64	2
				≥65	3
GCS score at admission		NIHSS score at admission		GCS score at admission	
13–15	0	0–10	0	13–15	1
5–12	1	11–20	1	9–12	2
3–4	2	21–40	2	3–8	3
ICH location		ICH location		ICH location	
Supratentorial	0	Supratentorial	0	Supratentorial	1
Infratentorial	1	Infratentorial	1	Infratentorial	2
ICH volume, ml		ICH volume, ml		ICH volume, ml	
<30	0	<30	0	Supratentorial	
≥30	1	≥30	1	<40	1
				40–70	2
				>70	3
				Infratentorial	
				<10	1
				10–20	2
				>20	3
Extension into ventricles		Extension into ventricles		Extension into ventricles	
No	0	No	0	No	1
Yes	1	Yes	1	Yes	2

GCS, Glasgow Coma Scale; ICH, intracerebral hemorrhage; IGS-GS, ICH grading scale; NIHSS, National Institute Health stroke scale.

(oICH), developed in 152 patients in the United States in 2001 (3), which was subsequently modified by replacing the Glasgow Coma Scale (GCS) (7) with the National Institute Health stroke scale (NIHSS) (8) in 141 patients in Hong Kong, producing the modified ICH scale (mICH) in 2003 (4). An advantage of mICH scale is that it overcomes some limitations of the GCS in terms of poor sensitivity and use in aphasic patients (9), but a disadvantage is that it may overestimate clinical severity as the NIHSS is biased toward neurological function of the left hemisphere (10). Another scoring system with more granularity is the ICH grading scale (ICH-GS), developed in 310 patients in Mexico and is reported to perform better than the oICH (5).

The aims of this study were to determine the validity, reliability, and potential utility of these three popular scales in predicting poor outcome in participants of the INTERACT2 study, and whether the magnitude of benefit of intensive BP lowering varies according to prognosis grading.

Methods

Design

INTERACT2 was an international, multicentre, open, blinded endpoint assessed, randomized controlled trial, as outlined elsewhere (2,11). Briefly, 2839 patients with spontaneous ICH within six-hours of onset and elevated SBP (150–220 mmHg) were included from 144 hospitals in 21 countries between October 2008 and August 2012. Excluded were patients with a definite indication for, or contraindication to, intensive BP lowering treatment; a structural cerebral cause for the ICH; deep coma (scores 3–5 on the GCS) or massive hematoma with a poor prognosis; or if early surgery to evacuate the hematoma was planned. The study was approved by the ethics committees for each site, and informed consent was obtained from all patients or relevant surrogates. The study is registered with ClinicalTrials.gov (NCT00716079).

Assessments

Participants allocated to intensive BP lowering were to commence intravenous treatment and oral agent(s), according to prespecified treatment protocols based on locally available agents, with the goal of achieving a SBP level <140 mmHg within one-hour of randomization and to maintain this level whilst in hospital over the next seven-days. A SBP of <130 mmHg was used as the threshold to cease any treatment. Symptomatic episodes of severe hypotension were treated with intravenous fluids or vasopressor agents. Participants allocated to the guideline group were to receive BP treatment if their SBP was >180 mmHg; no lower level was stipulated. All participants were to receive oral antihypertensive agents (or topical nitrates) within seven-days (or discharge from hospital if sooner) including via a nasogastric tube if required; if not contraindicated and no other drugs were specifically required, combination treatment with an angiotensin converting enzyme inhibitor and diuretic was recommended, with the goal of achieving an SBP level of <140 mmHg during follow-up for the prevention of recurrent stroke.

Demographic and clinical characteristics were recorded at the time of enrolment. Stroke severity was measured using the GCS and NIHSS at baseline, 24 hours, and at day 7 (or earlier upon

discharge from hospital). A prespecified subgroup of 964 participants had repeat CT scans conducted at 24 hours after the baseline scan according to standardized techniques. Hematoma volumes were calculated centrally by trained scientists, who were blind to clinical data, treatment, and date and sequence of the scan, using computer-assisted multislice planimetric and voxel threshold techniques in MIStar Version 3.2 (Apollo Medical Imaging Technology, Melbourne, Australia). Functional outcome was assessed using the modified Rankin scale (mRS) (12) at 7, 28, and 90 days postrandomization. The primary measure of poor clinical outcome was death or major disability (mRS scores 3–6) at 90 days. Secondary outcome for these analyses was death at 30 days.

Selection of scores

The scores chosen had to contain at least four characteristics relevant to the variables in the INTEARCT2 dataset. The oICH was selected as it is the most cited score (3), the modification of the oICH replacing the GCS with the NIHSS was selected to test if it overcomes the limitations of the GCS (4), and the ICH-GS was reported to perform better than the oICH score (5).

Analysis

Baseline characteristics of participants were used to derive scores on the oICH (3), mICH (4), and ICH-GS (5). Discrimination on each of the scales for the primary outcome (mRS 3–6) at 90 days and mortality at 30 days was evaluated using the area under the receiver operator characteristic (ROC) curves. Sensitivity, specificity, positive predictive value (PPV), and negative predictive value were calculated using cutoff values that generated the best Youden index, the maximum vertical distance between the ROC curve, and the diagonal or chance line (i.e. it occurs at the cut point that optimizes the scores differentiating ability when equal weight is given to sensitivity and specificity) (13). The nonparametric method of DeLong *et al.* (14) was used to compare the ROC curves for the scales with $P < 0{\cdot}01$ considered statistically different. Sensitivity analyses were conducted to determine if the discrimination of the scores was different, first in patients allocated intensive BP lowering compared with those in the guideline group, and second in patients randomized in China compared with the rest of the world. Finally, the DeLong method was used to compare discrimination of the scores calculated at baseline and 24 hours among participants of the CT substudy, who had repeated CT scans where the ICH volume was calculated centrally at baseline and 24 hours.

Differences in baseline characteristics of patients included and excluded in analyses were tested using the χ^2 test for categorical variables, Students *t*-test for differences between means, and the Mann–Whitney test for differences in medians. Heterogeneity of the effect of intensive BP lowering treatment on the primary outcome across each of the scales was assessed, unadjusted, and reported as odds ratios (ORs) with 95% confidence intervals (CIs). Data manipulations were carried out using SAS version 9.2 (SAS Institute, Cary, NC, USA), and statistical analyses were conducted using STATA version 12.1 (Stata Corporation, College Station, TX, USA).

Results

A total of 2556 (90%) participants were included in analyses; excluded were 216 patients with missing data on ICH volume and 57 with missing 90-day outcomes. Table 2 shows that those excluded tended to be younger (mean age 61 vs. 64 years), male

Table 2 Baseline characteristics of participants with intracerebral hemorrhage excluded and included in analyses

Parameter	Excluded ($n = 273$)	Included ($n = 2556$)	*P* value
Time from onset of ICH to randomization, hours	3·9 (3·1–4·8)	3·7 (2·8–4·7)	0·040
Age, years	61 ± 13	64 ± 13	0·001
Male	189/273 (69)	1591/2556 (62)	0·023
Chinese region	202/273 (74)	1718/2556 (67)	0·023
Systolic blood pressure, mmHg	180 ± 18	179 ± 17	0·374
Diastolic blood pressure, mmHg	103 ± 16	101 ± 15	0·011
NIHSS score*	10 (6–15)	11 (6–16)	0·968
GCS score†	14 (12–15)	14 (13–15)	0·101
History of hypertension	200/272 (74)	1848/2554 (72)	0·681
Current use of antihypertensive drugs	132/272 (49)	1142/2554 (45)	0·229
Prior intracerebral hemorrhage	23/272 (9)	206/2554 (8)	0·822
Prior ischemic or undifferentiated stroke	33/273 (12)	290/2554 (11)	0·702
Prior acute coronary event	4/272 (2)	77/2554 (3)	0·147
Diabetes mellitus	27/272 (10)	278/2554 (11)	0·628
Baseline hematoma volume, ml	10 (5–16)	11 (6–20)	0·234
Deep location of hematoma‡	43/57 (75)	2127/2556 (83)	0·122
Left hemisphere site of hematoma	25/57 (44)	1288/2556 (50)	0·329
Intraventricular hemorrhage	21/57 (37)	719/2556 (28)	0·149

Values are *n* (%), mean (SD), or median (interquartile range).
*NIHSS, National Institutes of Health Stroke Scale, scores range from 0 (normal) to 42 (coma with quadriplegia).
†GCS, Glasgow Coma Scale, scores range from 15 (normal) to 3 (deep coma).
‡Location in the basal ganglia or thalamus.

(69% vs. 62%), randomized slightly later from the onset of symptoms (median 3·9 vs. 3·7 hours), and from China (74% vs. 67%).

At the 90-day outcome assessment, 311 (12%) patents had died, 1071 (42%) had major disability, and 1174 (46%) were independent. The distribution of poor outcomes in patients according to scores on each of the scales is shown in Fig. 1. The frequency of poor outcome increased with increasing scores on both the mICH and ICH-GS scales. For the oICH, no patients received the highest score of 5, due to the trial criterion excluding patients with GCS scores of 3–5 at baseline.

The mICH score had the best discrimination for 30-day mortality (ROC 0·780 mICH, 0·753 oICH, 0·745 ICH-GS, $P = 0·006$) and 90-day death or major disability (0·749 mICH, 0·677 oICH, 0·685 ICH-GS, $P < 0·001$) (Fig. 2). Sensitivity analysis of the ROC curves indicates the mICH was marginally different for participants in China compared with those in rest of the world (0·730 vs. 0·773, $P = 0·02$) but was similar for the two treatment groups (0·749 vs. 0·749, $P = 0·99$). In the 897 (35%) of participants of CT substudy, the mICH scale at 24 hours (using NIHSS and ICH volumes at this time) gave greater discrimination for both 30-day mortality (0·851 at 24 hours vs. 0·761 at admission, $P < 0·001$) and 90-day death or major disability (0·804 vs. 0·752, $P < 0·001$) than the mICH scale at baseline (Fig. S1).

Table 3 shows the cut point values for each of the scales that generated the highest Youden index (i.e. maximum sensitivity and specificity) for 30-day mortality and 90-day poor outcome. All scales showed poor sensitivity and PPV for predicting death alone, but they were good at predicting 90-day poor outcome, as it includes a measure of disability, with the mICH having the highest PPV (91%). Overall, though, the scales were only slightly better than chance in discriminating patients around the dichotomy of 'good' or 'poor' outcome, with c statistic/ROC range 0·57–0·68.

Figure 3 shows nonsignificant variation in the effect of intensive BP lowering on the primary outcome in INTERACT2 when stratified by grading scale. Stratifying by ICH-GS showed no variation in the treatment effect. The figure shows a trend for better outcome in patients with milder severity, in particular for oICH score of 1 (OR 0·65; 95% CI 0·50–0·85) and mICH score of 1 (OR 0·71; 95% CI 0·55–0·92). Although the point estimates suggest poor outcomes for patients with high oICH or mICH scores on these scales, the associated 95%CI are wide reflecting small numbers of poor prognosis participants in the study.

Discussion

In this analysis of a large group of initially noncomatose patients with hypertension early after the onset of ICH, we have shown that three popular ICH clinical grading scales (oICH, mICH, and ICH-GS) cannot reliably identify those patients most likely to die over the next month. As such, we cannot recommend their use as a basis for decisions over the withdrawal of acute management or care in this patient group. However, the scales were more effective in defining the severity of this illness and in grading prognosis with regard to the likelihood of a poor outcome, either death or residual major disability at 90-days, especially when calculated using data 24 hours after stroke onset.

Although a greater benefit of acute BP lowering was suggested in patients with low scores of these scales, there was no statistically significant heterogeneity in the treatment effect. Taken together with other data indicating consistency of the treatment effect and safety in patients with different levels of presenting SBP, we consider that early intensive BP lowering should be applied as routine in all noncomatose ICH patients with elevated SBP irrespective of prognostic score to improve their chances of a better functional recovery.

In replacing the GCS with the NIHSS, our study indicates that the mICH improves upon the oICH in prediction of early death and subsequent poor outcome. In the derivation cohort of the ICH-GS, all patients with a score of ≥11 were dead at 30 days, whereas in our cohort, the figure was only 55% (11 deaths out of 20) (5). Previous validation studies have reported higher ROC for the oICH (0·861 (15), 0·882 (16)) and ICH-GS (0·874) (15), but they included a broader range of ICH patients who presented to hospital up to 24 hours after the onset of symptoms where as our clinical trial cohort was more restricted by time and characteristics, including the exclusion of patients in coma (GCS scores 3–5) or assessed as having a high likelihood of death in the next 24 hours.

A clinical grading scale that can accurately and reliably predict outcome in ICH could assist clinicians in the triage of patients in the emergency room, by identifying those who would benefit from intensive treatment, and in counseling about prognosis and prospect for recovery (17). However, our study has shown that not all patients with the worst prognosis, defined by high scores on these scales, die or remain disabled by 90 days. Reasons why these scores are not perfect in predicting outcome in the hyperacute phase of ICH include the complexity of this condition with associated co-morbid and other factors contributing to neurological status. Furthermore, the pattern of recovery from ICH is particularly difficult to model early after the onset of symptoms as it often follows a nonlinear trajectory, with the greatest gains in recovery being in the first few weeks, and then, the recovery continues at a slower rate until it plateaus beyond several months (18,19). It is not surprising that there are so many different prognostic ICH scales, with none widely accepted for routine care or clinical research (17). Moreover, concerns have been raised over the use of these scores as self-fulfilling prophecies, such that they lead to inappropriate withdrawal or palliative care of ICH patients in the acute phase (20,21). Our data support this view by showing that these prognostic scales are inadequate in defining those patients who are likely to die from ICH.

Strengths of this study include the use of a large and broad patient population who was prospectively assessed with rigorous data collection, including of the volume and location of the ICH being analyzed centrally blind to other information. As the scales were assessed for both death and functional outcomes, at an early and late time points in 2556 patients, this is by far the largest evaluation of the reliability of these grading scales to date. However, the study is limited by being undertaken in a selected clinical trial population of initially noncomatose patients with

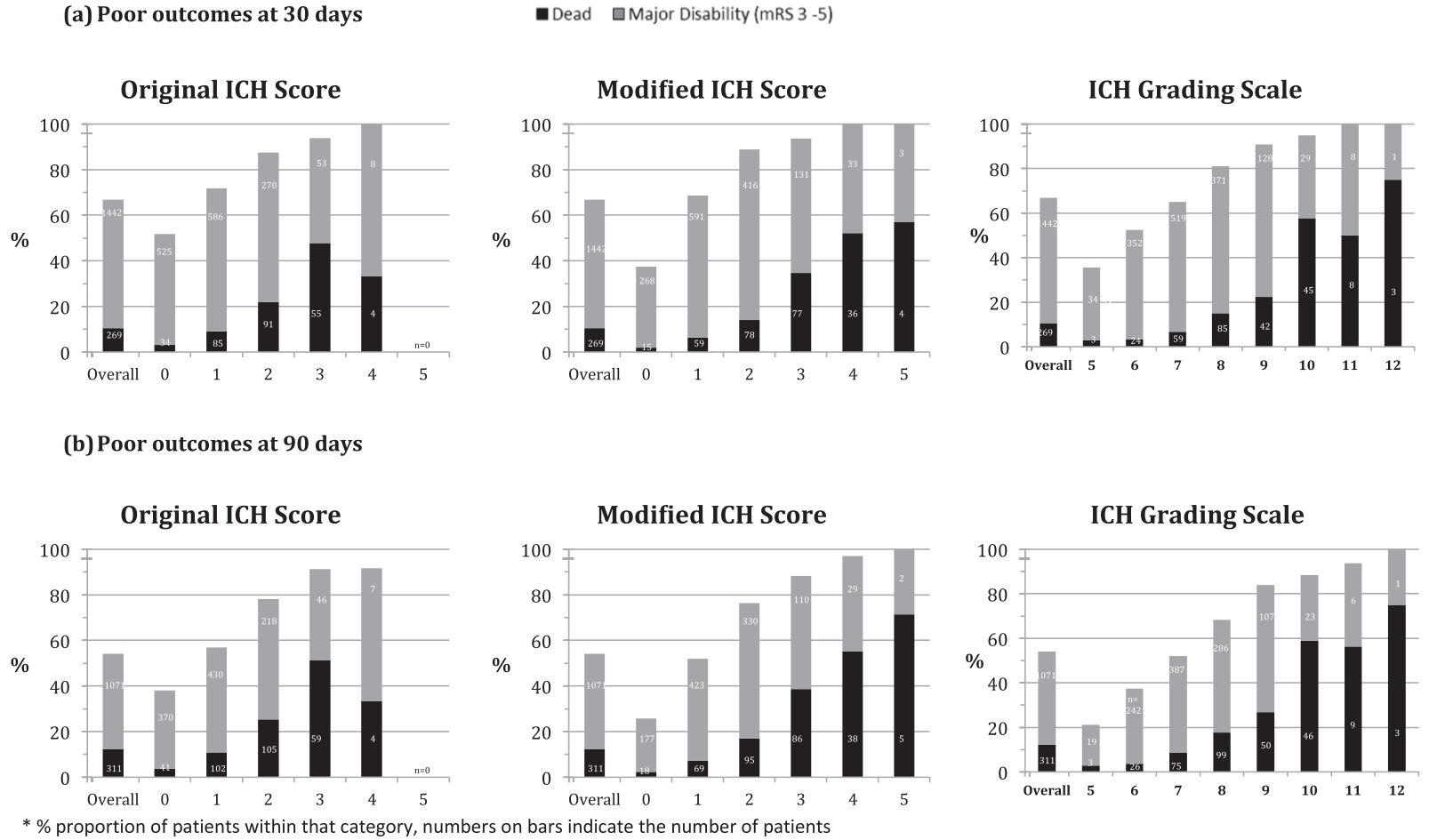

Fig. 1 Distribution of outcome at (a) 30 days and (b) 90 days, on the original intracerebral hemorrhage (ICH) score, modified ICH score, or ICH grading scale.

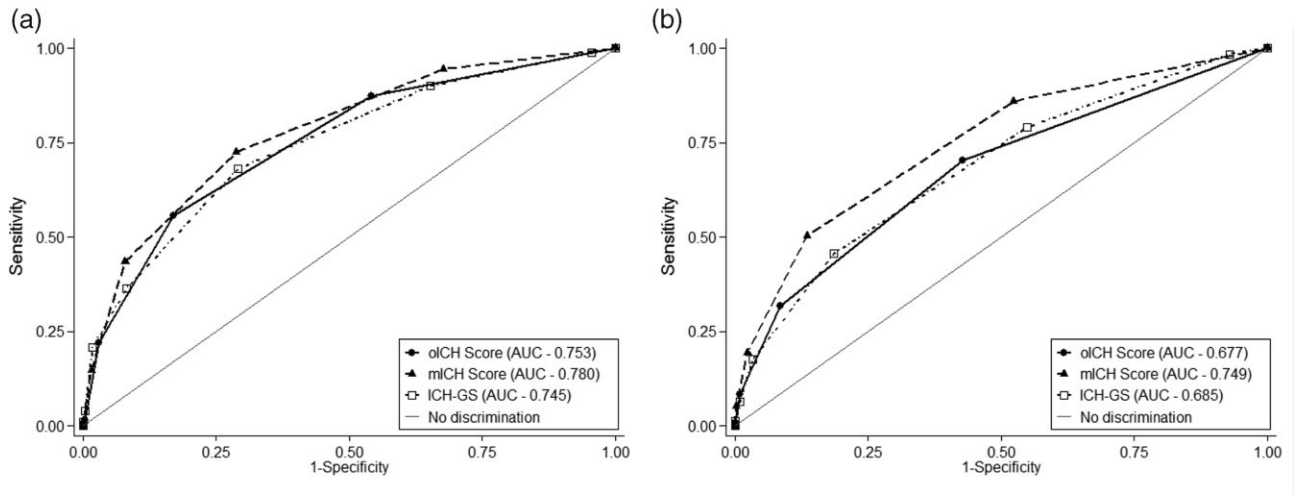

Fig. 2 Receiver operator characteristic (ROC) curves for each prognostic scale calculated at presentation (baseline) for predicting (a) 30-day mortality and (b) 90-day poor outcome (death or major disability).

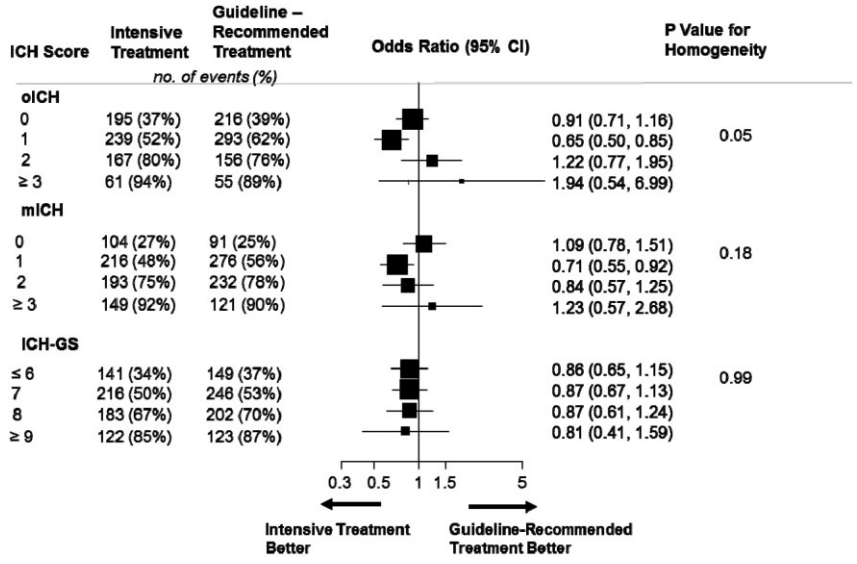

Fig. 3 Effect of intensive blood pressure (BP) lowering treatment on 90-day poor outcome by prognostic scales.

Table 3 Sensitivity, specificity, and positive and negative predictive values, using the cut point determined by the Youden index, and discrimination of the oICH, mICH, and ICH-GS scales for 30-day mortality and 90-day poor outcome

	30-Day mortality			90-Day poor outcome		
	oICH	mICH	ICH-GS	oICH	mICH	ICH-GS
Cut off	>2	>2	>8	>1	>2	>8
Sensitivity, % (95% CI)	22 (17, 27)	44 (38, 50)	36 (31, 43)	32 (29, 34)	20 (18, 22)	18 (16, 20)
Specificity, % (95% CI)	97 (96, 98)	92 (91, 93)	92 (91, 93)	91 (90, 93)	98 (97, 98)	97 (95, 98)
Positive predictive value, % (95% CI)	46 (38, 56)	39 (34, 45)	34 (29, 40)	81 (78, 85)	91 (87, 94)	86 (81, 90)
Negative predictive value, % (95% CI)	92 (90, 92)	93 (92, 94)	93 (91, 94)	53 (51, 55)	51 (49, 53)	50 (48, 52)
c Statistic* (95% CI)	0·59 (0·57, 0·62)	0·68 (0·65, 0·71)	0·64 (0·61, 0·67)	0·62 (0·60, 0·63)	0·59 (0·57, 0·60)	0·57 (0·56, 0·58)

*Area under the receiver operating characteristic curve.
oICH, original intracerebral hemorrhage (ICH) scale; mICH, modified ICH scale; ICH-GS, ICH grading scale; CI, confidence interval.

generally mild grade ICH, explaining the low mortality of 12%. The low numbers of poor prognosis patients limits the ability to assess reliability of the effects of treatment across all grades of ICH.

In summary, the oICH, mICH, and the ICH-GS scores are not useful in defining patients with high probability of death but rather are reliable in predicting poor outcome as defined by either death or major disability at 90 days. The mICH, which incorporates the NIHSS, appears to be the best outcome predictor, with performance consistent between Chinese and non-Chinese ICH patients and was even better among patients who survive to 24 hours. As the outcome from ICH is difficult to determine with precision on initial assessment, and as early intensive BP lowering is safe and broadly beneficial at improving the chances of functional recovery, all noncomatose patients deserve active management which includes targeted control of BP.

References

1 Krishnamurthi RV, Feigin VL, Forouzanfar MH *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 Global Health* 2013; **1:**e259–81.

2 Anderson CS, Heeley E, Huang Y *et al*. Rapid blood-pressure lowering in patients with acute intracerebral hemorrhage. *N Engl J Med* 2013; **368:**2355–65.

3 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:**891–7.

4 Cheung RT, Zou LY. Use of the original, modified, or new intracerebral hemorrhage score to predict mortality and morbidity after intracerebral hemorrhage. *Stroke* 2003; **34:**1717–22.

5 Ruiz-Sandoval JL, Chiquete E, Romero-Vargas S, Padilla-Martinez JJ, Gonzalez-Cornejo S. Grading scale for prediction of outcome in primary intracerebral hemorrhages. *Stroke* 2007; **38:**1641–4.

6 Hwang BY, Appelboom G, Kellner CP *et al*. Clinical grading scales in intracerebral hemorrhage. *Neurocrit Care* 2010; **13:**141–51.

7 Teasdale G, Jennett B. Assessment of coma and impaired consciousness. A practical scale. *Lancet* 1974; **2:**81–4.

8 Brott T, Adams HP Jr, Olinger CP *et al*. Measurements of acute cerebral infarction: a clinical examination scale. *Stroke* 1989; **20:**864–70.

9 Weimar C, Roth M, Willig V, Kostopoulos P, Benemann J, Diener HC. Development and validation of a prognostic model to predict recovery following intracerebral hemorrhage. *J Neurol* 2006; **253:**788–93.

10 Woo D, Broderick JP, Kothari RU *et al*. Does the National Institutes of Health Stroke Scale favor left hemisphere strokes? NINDS t-PA Stroke Study Group. *Stroke* 1999; **30:**2355–9.

11 Delcourt C, Huang Y, Wang J *et al*. The second (main) phase of an open, randomised, multicentre study to investigate the effectiveness of an intensive blood pressure reduction in acute cerebral haemorrhage trial (INTERACT2). *Int J Stroke* 2010; **5:**110–6.

12 Rankin J. Cerebral vascular accidents in patients over the age of 60. II. Prognosis. *Scott Med J* 1957; **2:**200–15.

13 Youden WJ. Index for rating diagnostic tests. *Cancer* 1950; **3:**32–5.

14 DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988; **44:**837–45.

15 Parry-Jones AR, Abid KA, Di Napoli M *et al*. Accuracy and clinical usefulness of intracerebral hemorrhage grading scores: a direct comparison in a UK population. *Stroke* 2013; **44:**1840–5.

16 Godoy DA, Pinero G, Di Napoli M. Predicting mortality in spontaneous intracerebral hemorrhage: can modification to original score improve the prediction? *Stroke* 2006; **37:**1038–44.

17 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:**839–44.

18 Toschke AM, Tilling K, Cox AM, Rudd AG, Heuschmann PU, Wolfe CD. Patient-specific recovery patterns over time measured by dependence in activities of daily living after stroke and post-stroke care: the South London Stroke Register (SLSR). *Eur J Neurol* 2010; **17:**219–25.

19 Wei JW, Heeley EL, Wang JG *et al*. Comparison of recovery patterns and prognostic indicators for ischemic and hemorrhagic stroke in China: the ChinaQUEST (QUality Evaluation of Stroke Care and Treatment) Registry study. *Stroke* 2010; **41:**1877–83.

20 Becker KJ, Baxter AB, Cohen WA *et al*. Withdrawal of support in intracerebral hemorrhage may lead to self-fulfilling prophecies. *Neurology* 2001; **56:**766–72.

21 Zahuranec DB, Morgenstern LB, Sanchez BN, Resnicow K, White DB, Hemphill JC 3rd. Do-not-resuscitate orders and predictive models after intracerebral hemorrhage. *Neurology* 2010; **75:**626–33.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Figure S1. ROC curves using the modified ICH score calculated at presentation (baseline) and at 24 hours for predicting (a) 30-day mortality and (b) 90-day poor outcome.