J Neurol (2009) 256:464–469
DOI 10.1007/s00415-009-0115-z

ORIGINAL COMMUNICATION

Christian Weimar
Andreas Ziegler
Ralph L. Sacco
Hans C. Diener
Inke R. König
on behalf of the VISTA investigators*

Predicting recovery after intracerebral hemorrhage – An external validation in patients from controlled clinical trials

Received: 31 March 2008
Received in revised form: 28 August 2008
Accepted: 25 September 2008
Published online: 18 March 2009

PD Dr. C. Weimar (✉) · H. C. Diener
Dept. of Neurology
University of Duisburg-Essen
Hufelandstr. 55
45122 Essen, Germany
Tel.: +49-201/723-2495
E-Mail: stroke.med@uni-due.de

A. Ziegler · I. R. König
Institut für Medizinische Biometrie und Statistik
Universität zu Lübeck
Ratzeburger Allee 160
23538 Lübeck, Germany

R. L. Sacco
Miller School of Medicine
University of Miami
Miami, FL, USA

Abstract *Background* An early and reliable prognostic indication in stroke patients is potentially useful for initiation of individual treatment and for informing patients and relatives. We recently developed a regression model as well as a simple 11-point predictive score (Essen ICH score) for functional recovery within three months after acute intracerebral hemorrhage (ICH) based on age and the National Institutes of Health Stroke Scale (NIH-SS). Here, we demonstrate the applicability of our models in an independent sample of ICH patients from controlled clinical trials. *Methods* The prognostic models were used to predict functional recovery in 564 patients from the Virtual International Stroke Trials Archive (VISTA). Furthermore, we tried to improve the accuracy by re-calibration and estimating new model parameters. *Findings* The logistic regression model and the Essen ICH score were able to correctly classify 77.5 % and 76.4 % of patients, respectively. Re-calibration and novel estimation of parameters yielded only a slight improvement of overall predictive accuracy. *Interpretation* For acute ICH patients included in controlled trials, our predictive models based on age and the NIH-SS correctly predict functional recovery after three months and could be useful for future trial design.

Key words ICH · outcome · NIHSS · prediction · clinical trials

Introduction

While intracerebral hemorrhage (ICH) accounts for only 10–15 % of all strokes, it is associated with considerably higher fatality and greater post-stroke disability compared to ischemic stroke [1–3]. An early and reliable prognosis for potential outcome after stroke is important for clinicians as well as for patients and their family. In clinical trials, prediction models could also help to define individual clinical endpoints, select suitable patients, or reduce required sample sizes [4–6]. In addition to predicting a clinically meaningful endpoint, a prognostic model needs to be validated [7], and easy to implement, i.e., contain only a few variables which are readily available in all patients, to be useful and applicable to clinical practice [8]. We have previously developed and validated a regression model as well as a simple additive score for prediction of independent functional status after ICH in consecutive patients admitted to German neurology departments with acute stroke units [9, 10]. The purpose of this study was to demonstrate more stringently the utility of our models by applying them to patients in the data set of the Virtual International Stroke

* VISTA steering committee members: Lees KR (chair), Hacke W, Sacco RL, Diener HC, Grotta J, Lyden P, Donnan GA, Davis SM, Bath PMW, Wahlgren NG, Hennerici M, Kaste M, Hommel M, Fisher M, Warach S, Curram J, Teal P, Gregson B, Marler J, Claesson L, Bluhmki E, K. Muir and A. Shuaib

JON 3115

Trials Archive (VISTA, www.vista.gla.ac.uk/) [11]. Originating from diverse clinical trials in various countries, the ICH patients in VISTA differ from patients in our original German Stroke Study databank in terms of selection, level of stroke care, recruitment, and nationality. Based on our prognostic models, we investigated the adequacy of our previous models in this different population and analyzed whether the derivation of new model parameters could improve the predictive accuracy.

Methods

Model development

A description of the development of the prognostic models has been previously published [9, 10]. Functional independence was assessed on the Barthel Index (BI) as one of the most widely used measures of functional independence. To identify patients with recovery as advocated in guidelines for controlled clinical trials [12], a cut-off value of BI ≥ 95 versus < 95 was used. Specifically, a logistic regression model and an 11-point score were developed to predict functional recovery (BI ≥ 95) versus no functional recovery (BI < 95 or death) 3 months after index event. For development of the regression model we considered all variables on previous history, stroke severity and imaging information which could be routinely assessed within the first few hours after admission. The resulting logistic regression model, shown in Table 1, predicted a lower probability for functional recovery in patients with higher age and greater overall neurological impairments as measured on the overall score of the National Institutes of Health Stroke Scale (NIH-SS), assessed within 6 hours after symptom onset [10].

Additionally, we developed the Essen ICH score which included level of consciousness as a third variable despite overlapping with the NIH-SS total score, to account for the highly adverse prognosis of intubated or comatose patients [9]. Multiple models with different subscores for each of the three predictive variables were run with different cut-offs for fatality and functional recovery. The best fit in the development data set was obtained by summing subscores for age, NIH-SS level of consciousness, and NIH-SS total score as presented in Table 1. On a range from 0–10, an Essen ICH score of > 7 best predicted fatality and a score of < 3 best predicted functional recovery.

VISTA data set

VISTA (www.vista.gla.ac.uk/) was created with the aim of providing access to patient trial data in order to perform exploratory analyses and hypothesis testing [11]. All included trials were approved by institutional review boards. At the time of data extraction (15 March 2006), VISTA encompassed data of over 15,000 patients from 21 acute stroke randomized controlled trials. Most trials in VISTA have been targeted at ischemic stroke. For the purpose of this analysis, data on 564 patients were extracted from VISTA which met the following criteria:

- Documented entry criteria,
- Baseline assessment within 24 hours of stroke onset, including recording of neurological deficit by NIH-SS,
- Confirmation of ICH diagnosis by cerebral imaging,
- Outcome assessed three months after stroke onset, including recording of BI or fatality, and
- Monitoring procedures in practice to validate data.

These data comprised a total of 564 patients from treatment and placebo groups from the GAIN Americas (132 centers) and GAIN International trials (173 centers in 21 countries) [13, 14]. VISTA does not normally permit the unmasking of the trial source from which data were extracted; however in this case, an application was made to the principal investigators of the named trials in accordance with VISTA regulations. Because both trials had a neutral outcome, we do not believe that this would impact our results. A total of 16 VISTA patients enrolled in Germany between 1998–1999 were included in this analysis. None of these patients was included in the data set for development or validation of the prediction models.

Statistical analyses

For all analyses, patients with missing outcome data were excluded to allow for an evaluation of the resulting predictions. Patients in the original data set and in the VISTA data set are described with regard to age, NIH-SS overall score, NIH-SS level of consciousness, and gender. Differences in the data sets are presented as mean differences with 95% confidence intervals (CIs) based on a t-distribution of the difference (age and NIH-SS overall score) and as difference in proportion with 95% CIs as proposed by Newcombe (method 10) [15]. To investigate the applicability of our prognostic models, the algorithms as described in Table 1 were applied to all patients in the VISTA data set for whom complete information on age, the overall NIH-SS and level of consciousness (Essen ICH score only) was available. The resulting numbers of correct classifications overall and in each outcome group were determined. In addition, a Receiver Operating Characteristic (ROC) was drawn for both models plotting specificity against sensitivity. The ability of the models to discriminate between outcomes across the entire range of predicted probabilities was determined from the area under the curve (AUC) varying from 0.5 for a model that correctly predicts outcome no better than chance to 1.0 for a model that perfectly discriminates between the outcomes.

To optimize the fit of the logistic regression model in the VISTA data set, we additionally adjusted the original model as previously described in detail for a model predicting patients with ischemic stroke [16]. Specifically, we aimed at re-calibrating the model, i.e., at adjusting the value of the intercept. For this, a logistic regression model was developed to predict the observed outcomes in the VISTA data from the linear predictor of the original logistic regression

Table 1 Algorithms for prediction of favorable outcome after 3 months

	Original model $\hat{\beta}$	Re-calibrated model $\hat{\beta}$	Novel model $\hat{\beta}$	Essen ICH score point scores
Age (years)	−0.089	−0.089	−0.060	< 60 = 0; 60–69 = 1; 70–79 = 2; ≥ 80 = 3
NIH-SS (total score)	−0.244	−0.244	−0.205	0–5 = 0; 6–10 = 1; 11–15 = 2; 16–20 = 3; > 20 or coma = 4
NIH-SS (consciousness)	–	–	–	alert = 0; drowsy = 1; stupor = 2; coma = 3
Intercept	7.543	8.156	5.782	
Cut-off to predict independence	> 0.42	> 0.59	> 0.59	< 3

$\hat{\beta}$ estimated logistic regression coefficients

model. To keep the slope fixed but estimate the intercept, the linear predictor was used as an offset variable, thus fixing the coefficient at unity, so that the intercept was the only free parameter. The resulting estimate for the intercept indicates the deviation from the original one, and this model renders re-calibrated predicted probabilities.

Finally, a novel logistic regression model was developed based on the variables which had been selected for the previous models, namely, age and the NIH-SS overall score. The threshold for categorization of patients was determined anew based on the outcome frequencies in the validation data set in order to compare the resulting numbers of correct classifications with those from the previous approaches.

Because in the different approaches, the outcome of the same patients is predicted using different prognostic models, we tested differences in the overall accuracies using McNemar's test and present the estimated differences with 95 % CIs according to Zhou and Qin [17]. The analyses were performed using the R software environment, version 2.3.1, with the Design package by Harrell.

Results

Characteristics of the 564 patients in the VISTA data set meeting the specified inclusion criteria are presented in Table 2. The BI after 3 months was recorded in 449 patients, and in 115 patients without recorded BI, information on fatality within this time frame was available, so that these were additionally classified as not functionally recovered. Therefore, data from 564 patients were available, of whom 171 had recovered and 393 had not recovered or died after 3 months. As shown in Table 2, age was slightly higher than in the original data sets (mean difference = 2.24 years, 95 % CI = –0.34; 4.15); similarly, neurological impairment was less severe in the original sample (mean difference = 3.76, 95 % CI = 2.82; 4.70). The proportion of women was similar in both data sets (difference in proportions = 0.8 %, 95 % CI = –6.9 %; 8.7 %).

The original regression model and the Essen ICH score were applied to predict the patients' outcome, and the ROCs are shown in Fig. 1. Using the Essen ICH score with the predefined cut-off for functional recovery, 430 patients (76.4 %) were classified correctly overall, with a more correct prediction of not recovered patients than recovered patients (88.3 % and 48.8 %, respectively). The

Table 2 Patients' characteristics in original and VISTA data sets

		Original data (N = 207)	VISTA data (N = 564)
Gender (n, %)	female	85 (41.1 %)	227 (40.2 %)
	male	122 (58.9 %)	337 (59.8 %)
Age (years)	M (SD)	65.9 (11.3)	68.1 (12.2)
National Institutes of Health Stroke Scale (NIH-SS) at baseline	M (SD)	10.2 (6.2)	14.0 (5.8)
NIH-SS level of consciousness at baseline	alert	129 (62.3 %)	383 (68.0 %)
	not alert	78 (37.7 %)	180 (31.6 %)
Barthel Index at 3 months (n, %)	≥ 95	78 (37.7 %)	171 (30.3 %)
	< 95	89 (43.0 %)	278 (49.3 %)
	died	40 (19.3 %)	115 (20.4 %)

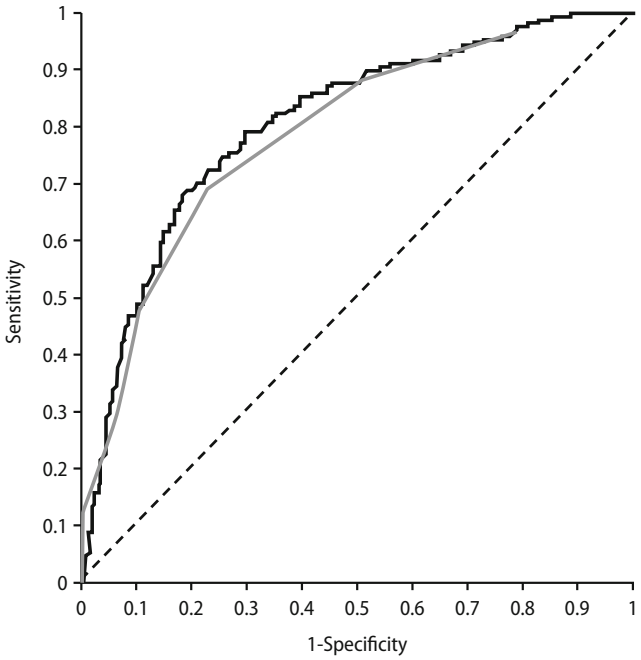

Fig. 1 Receiver Operating Characteristic for the original regression model (black) and the Essen ICH score (grey) predicting complete functional recovery (Barthel Index, BI ≥ 95) versus no functional recovery (BI < 95) or fatality

AUC was 0.786 (standard error (S.E.) = 0.021). Using the original parameter estimates and threshold for the regression model, 437 patients (77.5 %) were classified correctly overall, again with a more correct prediction of not recovered patients than recovered patients (92.1 % and 43.9 %, respectively). The AUC was 0.805 (S.E. = 0.020). Thus, the accuracy of the regression model was negligibly higher than that of the Essen ICH score (difference = 1.2 %, CI = –1.3 %; 3.7 %, two-sided p = 0.40). When re-calibrating the original regression model by using an adapted intercept, the deviation of the original and the new intercepts were estimated to be 0.63 (95 % CI = 0.39; 0.83), showing that the predicted probabilities for functional independence were systematically too low (see Fig. 2 a). Using the re-calibrated models (see Table 1 for estimated regression coefficients and Fig. 2 b for the calibration plot) led to slightly altered classifications with overall 77.3 % being classified correctly (83.0 % of not recovered and 64.3 % of recovered patients, respectively). Thus, the accuracy overall was comparable to that in the original model (difference = 0.2 %, 95 % CI = –2.8 %; 3.1 %, two-sided p = 1.00). However, the re-calibrated model predicted recovered patients better than the original model (difference = 20.5 %, 95 % CI = 14.5 %; 26.6 %, two-sided p < 0.01) but not recovered patients worse (difference = 9.2 %, 95 % CI = 6.3 %; 12.0 %, two-sided p < 0.01).

A novel logistic regression model was developed by estimating new regression coefficients of the previously

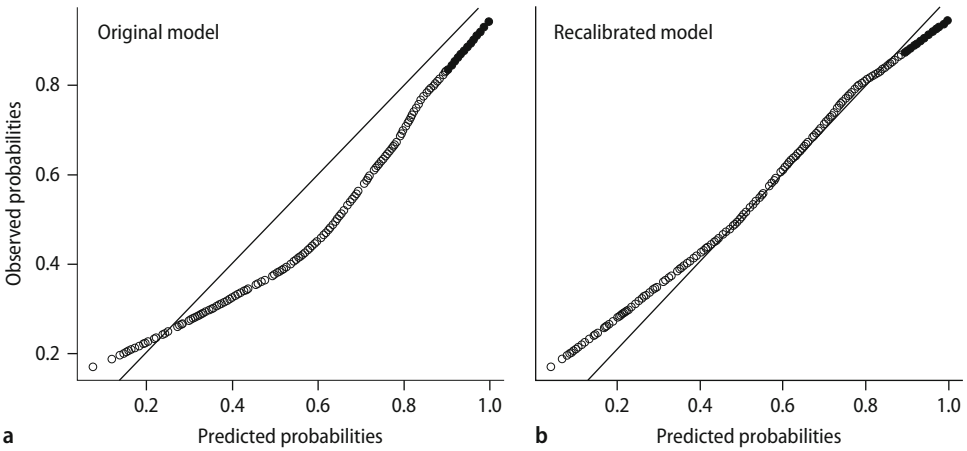

Fig. 2 Calibration plots delineating observed versus predicted outcome probabilities for functional dependence or fatality from (**a**) the original regression model and (**b**) from the recalibrated regression model. Dots represent the calibration curves created using lowess smoothers, and lines show the ideal calibration

identified parameters (Table 1). The novel model predicted 430 patients (76.2 %) correctly overall (83.0 % of not recovered and 60.8 % of recovered patients). The AUC was 0.805 (S.E. = 0.020). Compared with the original model, the overall accuracy was almost identical (difference = 1.2 %, 95 % CI = –1.6 %; 4.0 %, two-sided p = 0.46). As for the re-calibrated model, the novel model predicted recovered patients better (difference = 17.0 %, 95 % CI = 11.4 %; 22.7 %, two-sided p < 0.01) but not recovered patients worse (difference = 9.2 %, 95 % CI = 6.3 %; 12.0 %, two-sided p < 0.01).

Discussion

Our prediction models including age, NIH-SS total score and level of consciousness (Essen ICH score only) are valid in predicting functional recovery three months after ICH, with the classical regression model performing slightly better than the Essen ICH score. This prediction is highly clinically relevant to patients, clinicians and clinical researchers. To detect a possible treatment effect in acute stroke studies, patients included should have a high chance of incomplete recovery. Our models may therefore be helpful to exclude patients with high a chance of spontaneous recovery, because they are unlikely to contribute to a measurable treatment effect. An additional use of such a model to increase statistical power could be a sliding scale, i.e., if those predicted to be dependent actually are independent this is a good outcome, if those predicted to be independent who are not independent this is a bad outcome.

In contrast to previously developed ICH scores [18–25], our models are based on clinical variables only and have been externally validated for prediction of functional recovery in consecutively admitted ICH patients [9, 10]. By relying on the NIH-SS instead of the Glasgow Coma Scale as a clinical indicator of prognosis, we did not retain hematoma volume as an independent predictor which might be even easier to apply in clinical routine. Like in ischemic stroke [26], imaging parameters therefore may not add any relevant improvement to the prediction in addition to clinical stroke severity.

Although do-not-resuscitate orders in ICH patients may lead to self-fulfilling prophecies [27–29], this seems unlikely to have affected our population from clinical trials. Conversely, patients with pre-existing disability or highly adverse prognosis are usually excluded from clinical trials and therefore are likely to be underrepresented in the VISTA dataset. As a consequence, the Essen ICH score did not show any meaningful prediction of death in the VISTA data set (data not shown). Similar to patients with ischemic stroke [16], ICH patients from controlled studies in VISTA were systematically predicted too pessimistically by the original model. Interestingly, estimating the coefficients anew only marginally improves the overall prediction over the pre-specified logistic model, indicating a good discrimination of the original model. The same holds true for the re-calibrated model.

This study has some limitations. First, the predictive accuracies may not seem to justify relying on the given prediction over clinical judgement. However, we have previously shown that clinical judgement by the admitting neurology resident is inferior to our models and correctly predicted less than 70 % of all patients [10]. In contrast, the pre-specified models correctly predicted 76.4 and 77.5 %, respectively. Furthermore, the only other validated score has shown contradictory results concerning the predictive accuracy for favorable outcome [23, 30–32]. Second, imaging data in VISTA were not (yet) available to calculate and compare the performance of other ICH scores [18–24]. Third, we have now shown the external validity of our models in two different populations of ICH patients, namely, patients admitted to German neurology departments with an acute stroke unit and patients included in controlled clinical trials. This by no means represents the entire universe of

ICH patients, and subsequent studies are required to investigate the validity in further hemorrhagic stroke populations.

In summary, we have demonstrated and validated a strong predictive relation between our pre-specified prognostic models and the three month functional outcome even in different patient populations. These models have the potential to guide both clinical and research decisions.

Conflict of interest VISTA is a not-for profit collaboration of researchers from academia and commercial organizations. The VISTA Steering Committee members have each contributed to the organization of contributing trials and where these involved industry support have acknowledged that within the original publications. No author has any additional conflict of interest to declare in relation to this work, which was not externally supported.

Acknowledgments We thank the VISTA collaborators who contributed data from previous clinical trials, and staff at the Robertson Centre for Biostatistics for their help in organizing and managing the data.

References

1. Broderick JP, Brott T, Tomsick T, Miller R, Huster G (1993) Intracerebral hemorrhage more than twice as common as subarachnoid hemorrhage. J Neurosurg 78(2):188–191
2. Weimar C, Weber C, Wagner M, Busse O, Haberl RL, Lauterbach KW, Diener HC (2003) Management patterns and health care use after intracerebral hemorrhage. A cost-of-illness study from a societal perspective in Germany. Cerebrovasc Dis 15:29–36
3. Dennis MS (2003) Outcome after brain haemorrhage. Cerebrovasc Dis 16 (Suppl 1):9–13
4. Weir CJ, Kaste M, Lees KR (2004) Targeting neuroprotection clinical trials to ischemic stroke patients with potential to benefit from therapy. Stroke 35(9):2111–2116
5. Young FB, Lees KR, Weir CJ (2005) Improving trial power through use of prognosis-adjusted end points. Stroke 36(3):597–601
6. Weimar C, Ho TW, Katsarava Z, Diener HC (2006) Improving patient selection for clinical acute stroke trials. Cerebrovasc Dis 21(5–6):386–392
7. Konig IR, Malley JD, Weimar C, Diener HC, Ziegler A (2007) Practical experiences on the necessity of external validation. Stat Med 26(30):5499–5511
8. Wyatt JC, Altman DG (1995) Commentary: Prognostic models: clinically useful or quickly forgotten? Br Med J 311(7019):1539–1541
9. Weimar C, Benemann J, Diener HC (2006) Development and validation of the Essen Intracerebral Haemorrhage Score. J Neurol Neurosurg Psychiatry 77(5):601–605
10. Weimar C, Roth M, Willig V, Kostopoulos P, Benemann J, Diener HC (2006) Development and validation of a prognostic model to predict recovery following intracerebral hemorrhage. J Neurol 253(6):788–793
11. Ali M, Bath PM, Curram J, Davis SM, Diener HC, Donnan GA, Fisher M, Gregson BA, Grotta J, Hacke W, Hennerici MG, Hommel M, Kaste M, Marler JR, Sacco RL, Teal P, Wahlgren NG, Warach S, Weir CJ, Lees KR (2007) The Virtual International Stroke Trials Archive. Stroke 38(6):1905–1910
12. Committee for Proprietary Medicinal Products (CPMP) (2000) Points to consider on clinical investigation of medicinal products for the treatment of acute stroke. The European Agency for the Evaluation of Medicinal Products. CPMP/EWP/560/98
13. Sacco RL, DeRosa JT, Haley EC Jr, Levin B, Ordronneau P, Phillips SJ, Rundek T, Snipes RG, Thompson JLP, for the GAIN Americas Investigators (2001) Glycine antagonist in neuroprotection for patients with acute stroke: GAIN Americas: a randomized controlled trial. Jama 285(13):1719–1728
14. Lees KR, Asplund K, Carolei A, Davis SM, Diener HC, Kaste M, Orgogozo JM, Whitehead J (2000) Glycine antagonist (gavestinel) in neuroprotection (GAIN International) in patients with acute stroke: a randomised controlled trial. GAIN International Investigators. Lancet 355(9219):1949–1954
15. Newcombe RG (1998) Interval estimation for the difference between independent proportions: comparison of eleven methods. Stat Med 17:873–890
16. Konig IR, Ziegler A, Bluhmki E, Hacke W, Bath PMW, Sacco RL, Diener H-C, Weimar C (2008) Predicting long-term outcome after acute ischemic stroke – a simple index works in patients from controlled clinical trials. Stroke 39(6): 1821–1826
17. Zhou XH, Qin GS (2005) A new confidence interval for the difference between two binomial proportions of paired data. J Statist Plann Inference 128:527–542
18. Ruiz-Sandoval JL, Chiquete E, Romero-Vargas S, Padilla-Martinez JJ, Gonzalez-Cornejo S (2007) Grading scale for prediction of outcome in primary intracerebral hemorrhages. Stroke 38(5):1641–1644
19. Hallevy C, Ifergane G, Kordysh E, Herishanu Y (2002) Spontaneous supratentorial intracerebral hemorrhage. Criteria for short-term functional outcome prediction. J Neurol 249(12):1704–1709
20. Schwarz S, Hafner K, Aschoff A, Schwab S (2000) Incidence and prognostic significance of fever following intracerebral hemorrhage. Neurology 54(2):354–361
21. Lisk DR, Pasteur W, Rhoades H, Putnam RD, Grotta JC (1994) Early presentation of hemispheric intracerebral hemorrhage: prediction of outcome and guidelines for treatment allocation. Neurology 44(1):133–139
22. Portenoy RK, Lipton RB, Berger AR, Lesser ML, Lantos G (1987) Intracerebral haemorrhage: a model for the prediction of outcome. J Neurol Neurosurg Psychiatry 50(8):976–979
23. Cheung RT, Zou LY (2003) Use of the original, modified, or new intracerebral hemorrhage score to predict mortality and morbidity after intracerebral hemorrhage. Stroke 34(7): 1717–1722
24. Castellanos M, Leira R, Tejada J, Gil-Peralta A, Davalos A, Castillo J (2005) Predictors of good outcome in medium to large spontaneous supratentorial intracerebral haemorrhages. J Neurol Neurosurg Psychiatry 76(5):691–695
25. Rost NS, Smith EE, Chang Y, Snider RW, Chanderraj R, Schwab K, Fitz-Maurice E, Wendell L, Goldstein JN, Greenberg SM, Rosand J (2008) Prediction of functional outcome in patients with primary intracerebral hemorrhage: the FUNC score. Stroke 39(8): 2304–2309

26. Johnston KC, Wagner DP, Wang XQ, Newman GC, Thijs V, Sen S, Warach S (2007) Validation of an acute ischemic stroke model: does diffusion-weighted imaging lesion volume offer a clinically significant improvement in prediction of outcome? Stroke 38(6): 1820–1825
27. Becker KJ, Baxter AB, Cohen WA, Bybee HM, Tirschwell DL, Newell DW, Winn HR, Longstreth WT Jr (2001) Withdrawal of support in intracerebral hemorrhage may lead to self-fulfilling prophecies. Neurology 56(6):766–772
28. Hemphill JC 3rd, Newman J, Zhao S, Johnston SC (2004) Hospital usage of early do-not-resuscitate orders and outcome after intracerebral hemorrhage. Stroke 35(5):1130–1134
29. Zahuranec DB, Brown DL, Lisabeth LD, Gonzales NR, Longwell PJ, Smith MA, Garcia NM, Morgenstern LB (2007) Early care limitations independently predict mortality after intracerebral hemorrhage. Neurology 68(20): 1651–1657
30. Hemphill JC, Bonovich DC, Besmertis L, Manley GT, Johnston SC, Tuhrim S (2001) The ICH Score : A simple, reliable grading scale for intracerebral hemorrhage. Stroke 32(4):891–897
31. Jamora RD, Kishi-Generao EM Jr, Bitanga ES, Gan RN, Apaga NE, San Jose MC (2003) The ICH score: predicting mortality and functional outcome in an Asian population. Stroke 34(1):6–7; author reply 6–7
32. Fernandes H, Gregson BA, Siddique MS, Mendelow AD (2002) Testing the ICH score. Stroke 33(6):1455–1456; author reply 1455–1456