Stroke

American
Heart
Association

American
Stroke
Association®

Age and National Institutes of Health Stroke Scale Score Within 6 Hours After Onset Are Accurate Predictors of Outcome After Cerebral Ischemia: Development and External Validation of Prognostic Models

C. Weimar, I.R. König, K. Kraywinkel, A. Ziegler and H.C. Diener

The online version of this article, along with updated information and services, is located on the
World Wide Web at:
http://stroke.ahajournals.org/content/35/1/158

Age and National Institutes of Health Stroke Scale Score Within 6 Hours After Onset Are Accurate Predictors of Outcome After Cerebral Ischemia

Development and External Validation of Prognostic Models

C. Weimar, MD; I.R. König, PhD; K. Kraywinkel, MD, MSc; A. Ziegler, PhD; H.C. Diener, MD;
on behalf of the German Stroke Study Collaboration

Background and Purpose—To date, no validated, comprehensive, and practicable model exists to predict functional recovery within the first hours of cerebral ischemic symptoms. The purpose of this study was to externally validate 2 prognostic models predicting functional outcome and survival at 100 days within the first 6 hours after onset of acute cerebral ischemia.

Methods—On admission to a participating hospital, patients were registered prospectively and included according to defined criteria. Follow-up was performed 100 days after the event. With the use of prospectively collected data, 2 prognostic models were developed and internally calibrated in 1079 patients and externally validated in 1307 patients. By means of age and National Institutes of Health Stroke Scale (NIHSS) score as independent variables, model I predicts incomplete functional recovery (Barthel Index <95) versus complete functional recovery, and model II predicts mortality versus survival.

Results—In the validation data set, model I correctly predicted 62.9% of the patients who were incompletely restituted or had died and 83.2% of the completely restituted patients, and model II correctly predicted 57.9% of the patients who had died and 91.5% of the surviving patients. Both models performed better than the treating physicians' predictions made within 6 hours after admission.

Conclusions—The resulting prognostic models are useful to correctly stratify treatment groups in clinical trials and should guide inclusion criteria in clinical trials, which in turn increases the power to detect clinically relevant differences. (*Stroke*. 2004;35:158-162.)

Key Words: cerebral ischemia ■ models, statistical ■ outcome ■ prognosis ■ stroke assessment

While the number of acute stroke trials has increased considerably during the past decade, comprehensive knowledge about the impact of prognostic factors on outcome after acute ischemic symptoms is still scarce. However, this information is indispensable in designing randomized clinical trials and in controlling for case mix variations in nonrandomized trials. Furthermore, the inclusion of predictive variables can crucially increase the power to detect clinically relevant differences.[1] Most previous prognostic models, however, are neither comprehensive nor externally validated.[2] More importantly, no validated prognostic model thus far is widely applicable to unselected patients admitted to the hospital within the first hours after acute cerebral ischemia, which is believed to be the therapeutic window for neuroprotective drugs and intra-arterial thrombolysis. We therefore sought to externally validate 2 comprehensive models to predict functional outcome and mortality within 6 hours after onset of cerebral ischemia, which had previously been developed from the large hospital-based cohort of the stroke data bank of the German Stroke Foundation (Stiftung Deutsche Schlaganfall-Hilfe).

Methods

Model Development

We developed 2 binomial logistic regression models for the prediction of complete recovery and mortality. Model I predicts complete functional recovery versus incomplete recovery or mortality, and model II predicts mortality versus survival.

In a first step, we identified possibly predictive variables in a systematic search of the literature (details are available from http://www.uni-essen.de/neurologie/stroke/free/lit_eng1.html). To allow for a very early prediction, only 16 variables displayed in Table 1

TABLE 1.   Clinical Variables Considered in Model Development

Age at event (in years)
Sex
History of:
Stroke
Myocardial infarction
Arterial hypertension: elevated blood pressure above 160/95 mm Hg at 2 independent readings before stroke or antihypertensive medication
Diabetes mellitus: history of elevated blood glucose at 2 independent readings or elevated HbA1c >7.5% at admission or antidiabetic medication
Baseline neurologic impairments at admission as rated on the National Institutes of Health Stroke Scale with single items (level of consciousness, gaze, motor right arm, motor left arm, motor right leg, motor left leg, dysarthria, extinction, and inattention) and overall score
Atrial fibrillation at admission on first ECG at admission

were taken into account, which can be assessed routinely on admission.

We selected the Barthel Index (BI) as the most widely used measure of functional independence.[3] This scale evaluates individual abilities in feeding, dressing, mobility (walking on a level surface and ascending/descending stairs), and personal hygiene (grooming, toileting, bathing, and control of bodily functions). It therefore adequately reflects functional consequences for daily activities that are immediately important to the patient. To identify patients with complete recovery as advocated for clinical trials, a cutoff value of BI ≥95 versus <95 was used.

In a second step, we developed and internally validated prognostic models using data from the stroke data bank of the German Stroke Foundation. In 1998 and 1999, 7238 patients with acute cerebral ischemic symptoms at admission were included in the database. Seven centers (Minden, München-Harlaching [only 1998], Essen, Benjamin Franklin Berlin, München-Groβhadern [only 1998], Frechen, and Leipzig) met the specified quality criteria, registering a total of 3575 patients. Details on data assessment and management have been previously published.[4,5] Of the 3575 registered patients, only patients with a Rankin Scale score <3 before the event (n=3281), patients admitted within 6 hours after the onset of the stroke symptoms (n=1346), and those who survived during the first 6 hours (n=1344) were included. Of the remaining patients, 1079 patients were interviewed between 80 and 150 days after admission or were found to have died by the time of the interview. Mean age of these patients was 67.0 years (SD 12.3), and 39.5% were women. After 100 days, 644 patients (59.7%) were completely restituted (BI ≥95), 311 patients (28.8%) were incompletely restituted (BI <95), and 124 patients (11.5%) had died.

Descriptive statistics were obtained for all 16 variables and the recruiting center. The ordinal variable National Institutes of Health Stroke Scale (NIHSS) total score was treated as a linear variable in the regression models. To model the relationship between this score and outcome as well as the relationship between the continuous variable age and outcome, fractional polynomials were used on a randomly selected 25% of the total sample. The best fit was obtained with inclusion of only the linear term. Because of substantive correlations with other variables and less predictive value or reliability than the respective correlated variable, 2 single variables were eliminated (NIHSS motor left arm and NIHSS motor right arm). The remaining 14 variables were fitted into the logistic regression models via forward, backward, and stepwise selection. For model I, the number of events per variable was >30. Nevertheless, variables were retained only if their resulting probability value was ≤0.005. For model II, because of the lower events per variable of 9, all variables with probability values >0.001 were excluded. From models with all variables that resulted from any of the selection procedures, any variable with $P>0.005$ (model I) or $P>0.001$ (model II) was

eliminated stepwise. To the remaining set of variables, every previously eliminated variable was again added and kept in the model if it fulfilled the same criteria. Finally, all 2-way interactions of the resulting variables were investigated and kept if $P≤0.005$ (model I) or $P≤0.001$ (model II). In addition, the proportion of explained variance R^2 was calculated for each model.[6] Leave-1-out cross-validation was used to estimate the shrinkage factor γ in both models.[7] The threshold for classification with the use of the logistic distribution function was set so that the predicted proportion of events was equal to the observed. Finally, the calibrated percentage of correctly classified patients was calculated.

We assessed the discrimination of the 2 models by calculating the area under the receiver operating characteristic (ROC) curve, which is a plot of sensitivity of predictions against 1−specificity of predictions. An area under the ROC curve of 0.5 indicates no discrimination (ie, the line follows the 45° diagonal), and an area of 1.0 (ie, the line includes the entire area within the horizontal and vertical axes) indicates perfect discrimination.
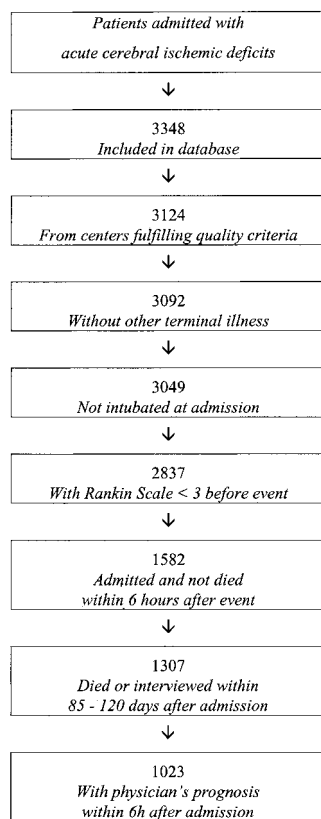
External Validation Study

The 13 neurological departments listed in the Appendix participated in this study. Enrollment of patients started on February 1, 2001, and was terminated on March 15, 2002, after the predefined number of patients according to our study protocol had been reached. Details on data collection and management have been previously described.[4,8] On admission, the treating physician reported the admission of every stroke patient via fax to the coordinating center at the University Hospital of Essen. Additionally, the admitting physician's prediction of outcome after 100 days, together with delay from admission, was assessed into 1 of the following categories: death, severe dependence (BI score <70), moderate dependence (BI score 70 to 90), and functional independence (BI score ≥95). Patients were informed about study participation, and informed written consent was obtained to forward personal data to the coordinating center. Imaging studies were performed to exclude patients with hemorrhages and causes other than cerebral ischemia. Patients were treated according to best current knowledge in clinical routine.

A central follow-up was performed via telephone interview by the coordinating center (84.4%) or by the treating hospital itself (15.6%) if the patient did not consent that personal data be forwarded. The outcome of the patient was assessed on the BI within 85 to 120 days after the event or by confirmation of death within 120 days after initial stroke. Otherwise, follow-up data were considered missing for analysis.

The study was approved by the Ethics Committee of the University of Essen, and aspects of data safety were approved by the responsible data protection state representative. According to our study protocol,[4] we excluded all patients from those centers with <75% follow-up or a dropout rate of >10%. The dropout rate was defined as the proportion of initially reported patients who could not be included in the validation study because of missing baseline information. The remaining patients were included if they met the following criteria: no serious functional impairment (Rankin Scale score <3) before the event to ensure that patients were functionally independent to a certain degree and not intubated at admission to allow for a valid assessment of neurological deficits. We furthermore included only those patients with complete follow-up information obtained between 85 and 120 days after admission. Of 275 patients without a valid follow-up, 44 (16%) refused to participate, 108 (39.2%) were interviewed outside the defined time window, and 123 (44.7%) could not be tracked via their primary care physician or the local citizen registry. These patients were not significantly different regarding sex, age, and NIHSS score at admission compared with the patients included in the validation analysis. The flow chart of patient inclusion is depicted in the Figure.

Statistical Analysis

Both models were validated in the whole cohort of patients for whom complete data on the predictive variables and outcome were obtainable.

Patients admitted with
acute cerebral ischemic deficits

↓

3348
Included in database

↓

3124
From centers fulfilling quality criteria

↓

3092
Without other terminal illness

↓

3049
Not intubated at admission

↓

2837
With Rankin Scale < 3 before event

↓

1582
*Admitted and not died
within 6 hours after event*

↓

1307
*Died or interviewed within
85 - 120 days after admission*

↓

1023
*With physician's prognosis
within 6h after admission*

Flow diagram of patients included in the validation analysis.

To confirm the statistical prognostic significance for all variables included in model I, a binary logistic regression analysis was performed with the following prognostic model: logit(BI <95)=Intercept+β_1 · (Age)+β_2 · (Neurological Impairment on NIHSS).

Using likelihood ratio test statistics, we tested the null hypotheses that both variables equal zero versus the alternative hypotheses that they differ from zero. The global significance level was given by $\alpha=0.05$. For each hypothesis, the significance level was adjusted according to Šidak.[9]

Similarly, the prognostic model II was calculated to obtain a validated estimate of the predictive quality. For all variables, odds ratio estimates with corresponding 95% Wald CIs are presented. For both models, patients were classified with the use of the estimates of the previously developed regression models.

Results

Model Development
Model I found an increased risk of not attaining complete recovery (BI <95 or death) in older patients and in patients with a more severe level of neurological impairments at admission (NIHSS total score) (Table 2). With the use of the

threshold 0.402, 79.0% of all patients could be correctly classified. The final model explained $R^2=51.42\%$ of the complete variation. A shrinkage factor of $\gamma=0.99$ was obtained.

Predicting mortality versus survival, model II likewise included age and level of neurological impairments at admission (NIHSS total score) (Table 3). A total of 86.9% of patients were classified correctly when the threshold 0.267 was used. The proportion of variance explained by this model was $R^2=29.9\%$, and the shrinkage factor was estimated to be $\gamma=0.99$. The area under the ROC curve was 0.856 in model I and 0.832 in model II. The ROC curves for development of both models are available online at http://stroke.ahajournals.org (XFigures I and II).

External Validation Study
Mean age of the 1307 patients was 68.2 years (SD 12.5), and 43.5% were women. The mean NIHSS score at admission was 7.6, with SD of 6.9 and a median of 5. In addition to the inclusion in this study, 49 patients (3.7%) had participated in clinical trials, and 178 patients (13.6%) had received systemic or intra-arterial thrombolysis. After 100 days, 722 patients (55.2%) were completely restituted (BI ≥95), 445 patients (34.0%) were incompletely restituted (BI <95), and 140 patients (10.7%) had died.

Model I was validated to predict incomplete recovery (BI <95 or death) in patients with higher age and greater overall neurological impairments at admission on the NIHSS. This model explained $R^2=44.3\%$ of the complete variation. With the use of the original β estimates and the predefined threshold 0.402, 74.1% of all patients could be correctly classified. Details of the classification correctness in each group are given in Table 4. According to the admitting physician's prediction, only 68.9% of the patients were predicted correctly, with 83.5% of patients who had died or were incompletely restituted and 56.7% of completely restituted patients.

In model II, a total of 87.9% of patients were classified correctly when the original β estimates and the threshold 0.289 as presented in Table 5 were used. The proportion of variance explained by this model was $R^2=38.9\%$. According to the admitting physician's prediction, 90.0% were predicted correctly, with only 9.0% of patients who died, in contrast to 99.9% of surviving patients.

Discussion
To our knowledge, this is the first study to externally validate 2 comprehensive models predicting functional outcome

TABLE 2.   Development of Model I to Predict Incomplete Functional Recovery or Mortality Versus Complete Recovery (BI <95 or dead versus BI ≥95)

Variable	β	SE	Odds Ratio	95% CI
Intercept	−5.782	0.537		
Age (difference of 1 year)	0.049	0.007	1.051	1.036 to 1.065
NIHSS total score at admission (difference of 1 scale score)	0.272	0.018	1.313	1.268 to 1.360

β=regression coefficient; SE indicates standard error; CI, confidence interval; NIHSS, National Institutes of Health Stroke Scale at admission.

TABLE 3.   Development of Model II to Predict Mortality Versus Survival

Variable	β	SE	Odds Ratio	95% CI
Intercept	−7.040	0.778		
Age (difference of 1 year)	0.049	0.010	1.050	1.029 to 1.071
NIHSS total score at admission (difference of 1 scale score)	0.155	0.015	1.168	1.134 to 1.203

β=regression coefficient; SE indicates standard error; CI, confidence interval; NIHSS, National Institutes of Health Stroke Scale.

within the first hours after onset of cerebral ischemia, which were developed with consideration of all practicable and previously identified prognostic variables. As advocated by international guidelines[10] and a review of prognostic models,[2] we focused on complete recovery and mortality 100 days after the ischemic event as end points of primary interest. By performing a systematic literature search before model development, we were able to consider previously suggested factors simultaneously to estimate their relative influence on the outcome variables. Through a predominantly central follow-up, we were able to ensure a standardized outcome assessment based on consistent criteria. Although the follow-up rate of 83.8% does not preclude a possible bias, this seems unlikely to affect the validity of our findings because the main characteristics of the patients lost to follow-up were not significantly different from those of the patients included in this analysis.

While our models have many strengths and seem more widely applicable than any previously presented prognostic model in acute ischemic stroke, several limitations apply. Because both study populations represent hospital-based cohorts, unselected patients in different care settings might have a different prognosis than suggested in our models. Our models therefore can only be considered validated for patients on acute German Stroke Units and cannot be transferred readily to stroke registers or other stroke care institutions. To make a more accurate prognosis of the population at interest, we excluded patients with little or no chance of reaching the primary outcome variable. We also excluded patients who were intubated at admission because this precluded a meaningful assessment of the neurological deficits from the ischemic event. We included patients who were only mildly affected in both development and validation of the models. Excluding patients with NIHSS score <3 at admission from the validation data set reduces the percentage of correctly classified patients in model I from 74.1% to 71.7%

TABLE 4.   Classification of Patients in Model I to Predict Incomplete Functional Recovery Versus Complete Recovery Using 0.402 as Threshold*

		Observed Frequency, %		
		BI <95	BI ≥95	Total
Predicted frequency, %	BI <95	368 (62.9)	121 (16.8)	489
	BI ≥95	217 (37.1)	601 (83.2)	818
	Total	585	722	1307

*Incomplete functional recovery=Barthel Index (BI) <95; complete recovery=BI ≥95.

and the R^2 from 44.3% to 38.3%. Still, our models reflect the whole range of patients who would be included in an acute intervention trial.

Before model development, we decided to refrain from considering specific treatment methods as possible predictors for 2 reasons. First, treatment decisions in our sample were based on clinical judgment, which would differ in the context of a clinical trial. Second, we surmised that no specific effective treatment is commonly applied to a considerable number of patients. Our data, however, showed that 13.6% of all patients received arterial or systemic thrombolysis. Excluding those patients from the validation sample increased the percentage of correctly classified patients in model I to 75.3% and the R^2 to 49.0%.

Several prior analyses from randomized clinical trials as well as observational studies agree on the early predictive value of both age and initial stroke severity.[11–19] However, no study thus far has both exclusively relied on information obtained within 6 hours after onset of ischemic symptoms and been externally validated. With the large sample size available for our model development, we were furthermore able to test several possibly predictive variables without overfitting the resulting models. However, neither the information on risk factors and comorbidity nor any specific neurological deficits in addition to the overall score on the NIHSS proved to be independent predictors for our chosen end points. This supports the predictive accuracy of the NIHSS without a need to correct for imbalances in scale composition.

Both variables in model I proved to be independent predictors of outcome in the validation data set. In model II, both previously identified prognostic variables were likewise independent predictors of outcome in the validation data set. The overall prognostic accuracy in this model was very high (87.9%), although only 57.9% of the deceased patients were classified correctly. Still, the sensitivity of the model was substantially better than the prediction of the physicians, in which merely 9.0% of the deceased patients were predicted correctly. Although they are better than the clinicians' estimates, the predictive values of the models are still less than ideal. A less rigorous approach to inclusion of other prognostic factors may possibly have led to more accurate models, albeit at the expense of greater complexity and lower stability. However, we opted for stringent selection criteria for the prognostic factors to increase the stability of our models, which was substantiated by the high shrinkage factor of γ=0.99 for both models. We were unable to include information from cerebral imaging in our models because this would have required a very tight time frame for the examination as well as a highly standardized evaluation protocol. In

TABLE 5.   Classification of Patients in Model II to Predict Mortality Versus Survival Using 0.289 as Threshold

		Observed Frequency, %		
		Mortality	Survival	Total
Predicted frequency, %	Mortality	81 (57.9)	99 (8.5)	180
	Survival	59 (42.1)	1068 (91.5)	1127
	Total	140	1167	1307

contrast, a previous study assessing diffusion imaging in acute stroke patients not only had a far smaller number of patients included but also used a longer time window of 24 hours.[14] Instead of other technical or laboratory investigations as surrogate markers of stroke severity, we decided to focus our models on anamnestic and clinical variables at admission because these variables are more readily accessible and do not require a sophisticated technique or rigorous time frame.

In conclusion, our models are based on 2 uniquely large, multicenter cohorts of stroke patients and have proven to be valid for patients with acute cerebral ischemic deficits admitted to acute stroke units in Germany. They will provide a valuable tool in the design of randomized trials and be helpful to control for case mix variations in nonrandomized trials. The inclusion of these predictive variables for stratifying treatment groups should be considered in clinical trials; this in turn would increase the power to detect clinically relevant differences.

Appendix

Members of the German Stroke Study Collaboration include the following neurology departments and responsible study investigators: Charité Berlin (N. Amberger), Krankenanstalten Gilead Bielefeld (C. Hagemeister), Rheinische Kliniken Bonn (C. Kley), University of Saarland (P. Kostopoulos), University of Jena (V. Willig), University of Magdeburg (M. Goertler), Klinikum Minden (J. Glahn), Städtisches Krankenhaus München Harlaching (K. Aulich), Klinikum München Groβhadern (A. Müllner), University of Rostock (A. Kloth), Bürgerhospital Stuttgart (T. Mieck), University of Ulm (M. Riepe), University of Essen (G. Mörger-Kiefer).

Acknowledgments

References

1. The European Agency for the Evaluation of Medicinal Products, Human Medicine Evaluation Unit. ICH Topic E 9: statistical principles for clinical trials. Available at: http://www.emea.eu.int/pdfs/human/ich/036396en.pdf. Accessed December 4, 2003.
2. Counsell C, Dennis M. Systematic review of prognostic models in patients with acute stroke. *Cerebrovasc Dis*. 2001;12:159–170.
3. Mahony FI, Barthel DW. Functional evaluation: the Barthel Index. *Md Med J*. 1965;14:61–65.
4. König IR, Weimar C, Diener HC, Ziegler A. Vorhersage des Funktionsstatus 100 Tage nach einem ischämischen Schlaganfall: Design einer prospektiven Studie zur externen Validierung eines prognostischen Modells. *Z Ärztl Fortbild Qualitätssich*. 2003;97:717–722.
5. Weimar C, Ziegler A, König IR, Diener HC. Predicting functional outcome and survival after acute ischemic stroke. *J Neurol*. 2002;249:888–895.
6. McKelvey RD, Zavoina W. A statistical model for the analysis of ordinal level dependent variables. *J Math Sociol*. 1975;4:103–120.
7. Verweij PJ, Van Houwelingen HC. Cross-validation in survival analysis. *Stat Med*. 1993;12:2305–2314.
8. The German Stroke Study Collaboration. Predicting outcome after acute ischemic stroke: an external validation of prognostic models. Neurology. In press.
9. Westfall PH, Young SS. *Resampling-Based Multiple Testing*. New York, NY: John Wiley; 1993.
10. The European Agency for the Evaluation of Medicinal Products, Human Medicine Evaluation Unit. Points to consider on clinical investigation of medicinal products for the treatment of acute stroke. Available at: http://www.emea.eu.int/pdfs/human/ewp/056098en.pdf. Accessed December 4, 2003.
11. Adams HP Jr, Davis PH, Leira EC, Chang KC, Bendixen BH, Clarke WR, Woolson RF, Hansen MD. Baseline NIH Stroke Scale score strongly predicts outcome after stroke: a report of the Trial of ORG 10172 in Acute Stroke Treatment (TOAST). *Neurology*. 1999;53:126–131.
12. Generalized efficacy of t-PA for acute stroke: subgroup analysis of the NINDS t-PA Stroke Trial. *Stroke*. 1997;28:2119–2125.
13. Allen CM. Predicting recovery after acute stroke. *Br J Hosp Med*. 1984;31:428,430,432–434.
14. Baird AE, Dambrosia J, Janket S, Eichbaum Q, Chaves C, Silver B, Barber PA, Parsons M, Darby D, Davis S, et al. A three-item scale for the early prediction of stroke recovery. *Lancet*. 2001;357:2095–2099.
15. Censori B, Camerlingo M, Casto L, Ferraro B, Gazzaniga GC, Cesana B, Mamoli A. Prognostic factors in first-ever stroke in the carotid artery territory seen within 6 hours after onset. *Stroke*. 1993;24:532–535.
16. Chambers BR, Norris JW, Shurvell BL, Hachinski VC. Prognosis of acute stroke. *Neurology*. 1987;37:221–225.
17. Fiorelli M, Alperovitch A, Argentino C, Sacchetti ML, Toni D, Sette G, Cavalletti C, Gori MC, Fieschi C, for the Italian Acute Stroke Study Group. Prediction of long-term outcome in the early hours following acute ischemic stroke. *Arch Neurol*. 1995;52:250–255.
18. Henon H, Godefroy O, Leys D, Mounier-Vehier F, Lucas C, Rondepierre P, Duhamel A, Pruvo JP. Early predictors of death and disability after acute cerebral ischemic event. *Stroke*. 1995;26:392–398.
19. Johnston KC, Connors AF Jr, Wagner DP, Haley EC Jr. Predicting outcome in ischemic stroke: external validation of predictive risk models. *Stroke*. 2003;34:200–202.