Konig, Inke and Ziegler, Andreas and Bluhmki, Enrich and Hacke, Werner and Bath, Philip M.W. and Sacco, Ralph L. and Diener, Hans-Christoph and Weimar, Christian (2008) Predicting long-term outcome after acute ischemic stroke: a simple index works in patients from controlled clinical trials. Stroke . ISSN 0039-2499

Access from the University of Nottingham repository:
http://eprints.nottingham.ac.uk/890/2/predictinglongterm.pdf

Copyright and reuse:
The Nottingham ePrints service makes this work by researchers of the University of Nottingham available open access under the following conditions.

This article is made available under the University of Nottingham End User licence and may be reused according to the conditions of the licence. For more details see:
http://eprints.nottingham.ac.uk/end_user_agreement.pdf

A note on versions:
The version presented here may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the repository url above for details on accessing the published version and note that access may require a subscription.

For more information, please contact eprints@nottingham.ac.uk
Predicting Long-Term Outcome After Acute Ischemic Stroke 
A Simple Index Works in Patients From Controlled Clinical Trials

Inke R. König, PhD; Andreas Ziegler, PhD; Erich Bluhmki, PhD; Werner Hacke, MD; 
Philip M.W. Bath, MD; Ralph L. Sacco, MD, MS; Hans C. Diener, MD; Christian Weimar, MD; 
on behalf of the Virtual International Stroke Trials Archive (VISTA) Investigators*

Background and Purpose—An early and reliable prognosis for recovery in stroke patients is important for initiation of 
individual treatment and for informing patients and relatives. We recently developed and validated models for predicting 
survival and functional independence within 3 months after acute stroke, based on age and the National Institutes of 
Health Stroke Scale score assessed within 6 hours after stroke. Herein we demonstrate the applicability of our models 
in an independent sample of patients from controlled clinical trials.

Methods—The prognostic models were used to predict survival and functional recovery in 5419 patients from the Virtual 
International Stroke Trials Archive (VISTA). Furthermore, we tried to improve the accuracy by adapting intercepts and 
estimating new model parameters.

Results—The original models were able to correctly classify 70.4% (survival) and 72.9% (functional recovery) of patients. 
Because the prediction was slightly pessimistic for patients in the controlled trials, adapting the intercept improved the 
accuracy to 74.8% (survival) and 74.0% (functional recovery). Novel estimation of parameters, however, yielded no 
relevant further improvement.

Conclusions—For acute ischemic stroke patients included in controlled trials, our easy-to-apply prognostic models based 
on age and National Institutes of Health Stroke Scale score correctly predicted survival and functional recovery after 3 
months. Furthermore, a simple adaptation helps to adjust for a different prognosis and is recommended if a large data 
set is available. (Stroke. 2008;39:000-000.)

Key Words: stroke ■ outcome ■ National Institutes of Health Stroke Scale ■ prediction ■ clinical trials

The importance of an early and reliable prognosis for 
recovery in patients with acute stroke is undisputed.1,2 In 
addition to clinical reasons, such as information for patients and family as well as adapting treatment and rehabilitation options, inclusion of prognostic information in controlled clinical trials helps to define individual clinical end points, to select suitable patients, and to reduce required sample sizes.3-5

To be useful and applicable to clinical practice, a prognostic model needs to be validated and easy to implement; ie, it should contain only a few variables that are readily available for all patients.6 A systematic review that included studies until 1997 showed that the methodology of most reported prognostic models for stroke recovery was poor, and none of the models was recommended for clinical practice or research.7 Since then, additional prognostic models have been developed. Of these, the validated models by Baird et al8 and Johnston et al9-11 that predicted recovery within 3 months relied on imaging variables, which may not be available for all patients and in all settings. In contrast, the models by Counsell et al12 and our own group13 included a few simple clinical variables and were also subsequently validated.14-16

Because in these models prognostic variables were assessed within a delay of 4 (median) and 2 to 3 days after stroke, respectively, timely prediction for initiating acute treatment was precluded. To allow for an almost immediate prognosis based on a few simple variables, we recently developed and externally validated models for predicting survival and functional independence within 3 months.17 These models are based on age and neurologic impairment as measured on the National Institutes of Health Stroke Scale (NIHSS), assessed within 6 hours after stroke onset. In the external validation sample, ≈75% of patients were classified correctly for functional independence and >85% with regard to survival.

In some respects, patients in our previous training and validation samples were highly comparable: all patients were admitted consecutively to German neurology departments with an acute stroke unit. We were therefore able to test the transportability of the models, which constitutes accuracy in different
but similar populations. The aim of the current 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 Trials Archive (VISTA, www.vista.gla.ac.uk/). Originating from diverse clinical trials in various countries, the stroke patients in VISTA differ from our original German Stroke Study data bank in terms of selection, level of stroke care, recruitment, and nationality. On the basis of our prognostic models, we will address the following questions: (1) Are the original data set and in the VISTA data set are described with regard to gender, NIHSS, and sex, and differences in the data sets are presented as mean differences and 95% CIs based on a t distribution of the difference (age and NIHSS) and as the difference in proportion with 95% CI, as proposed by Newcombe (method 10).

To compare the applicability of our prognostic models, 3 different approaches were taken. In the most stringent approach, the algorithms as described in Figure 1 and Table 1 were applied to all patients in the VISTA data set for whom complete information on age and overall NIHSS score was available. The resulting numbers of correct classifications overall and in each outcome group were determined. In addition, a receiver operating characteristic was drawn for each model, which plotted specificity versus sensitivity, and the area under the curve is presented with its SE.

In the second approach, we aimed to adjust the original models to optimize the fit in the VISTA data set. For this, it should be remembered that a logistic regression model for prognosis principally consists of 2 components, the intercept and a set of slope coefficients. If the intercept is valid in the new data, the model is well calibrated. In contrast, if it is misspecified, the resulting predicted probabilities are systematically either too high or too low. On the other hand, if the slope is incorrectly estimated, the model shows insufficient discrimination in the new data, and the spread of the predicted probabilities is either too extreme or not extreme enough, so that the model cannot differentiate between patients with a more or less favorable outcome. In this approach, we wanted to allow for a different prognosis but assumed that the effects of predictors would be similar. Therefore, we only aimed at recalibrating the models, i.e., adjusting the values of the intercept. This was solved technically by developing a logistic regression model to predict the observed outcomes in the VISTA data from the linear predictor of the original logistic regression model. The aims of this new regression model were to keep the slope fixed but to estimate the intercept. To meet this end, 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 recalibrated predicted probabilities. Only data from patients with complete information were used, and the resulting numbers of correct classifications overall and in each outcome group were determined.

Finally, novel logistic regression models were developed on the basis of the variables that had been selected for the previous models, namely, age and NIHSS score. The thresholds for categorization of patients were estimated by the Barthel Index (BI) as one of the most widely used models, we will address the following questions: (1) Are the original, recalibrated, and novel models predicting functional recovery and survival? (2) Can models be relevantly improved by developing novel models, i.e., deriving new parameter estimates? Subjects and Methods Model Development A description of the development of the models has been detailed elsewhere. In brief, functional independence of the patients was assessed by the Barthel Index (BI) as one of the most widely used measures of functional independence. To identify patients who recovered, as advocated in the guidelines for controlled clinical trials, a cutoff BI value ≥95 versus <95 was used. Specifically, the following 2 models were developed with data from the stroke data bank of the German Stroke Collaborators: (1) model I to predict functional recovery, i.e., BI ≥95 versus BI <95 or dead and (2) model II to predict survival versus death (all causes).

From a set of 16 possible predictive variables that had been identified in a systematic literature search and included single items as well as the overall score of the NIHSS, logistic regression models were fitted by forward, backward, and stepwise selection. To model the relation with continuous variables, fractional polynomials were applied, and possible model components, the intercept and a set of slope coefficients. If the intercept is valid in the new data, the model is well calibrated. In contrast, if it is misspecified, the resulting predicted probabilities are systematically either too high or too low. On the other hand, if the slope is incorrectly estimated, the model shows insufficient discrimination in the new data, and the spread of the predicted probabilities is either too extreme or not extreme enough, so that the model cannot differentiate between patients with a more or less favorable outcome.

To apply the resulting models, nomograms were created that provide the estimated probabilities for outcome and the resulting classifications based on age and NIHSS score of a single patient (see Figure 1). To forecast the outcome for a specific patient, the values of each variable are marked on the respective lines. For instance, for the functional recovery model, age is marked on the second line of Figure 1A. Then, a straight line is drawn upward to determine the points for the variable “age.” This is repeated for the NIHSS, and the points are summed and marked in the second to last line “Total Points.” Drawing a line downward to the lowest line then gives the predicted probability for this patient to become functionally independent. For example, a patient aged 66 years (15 points) and with an overall NIHSS score of 7 (82 points) receives a total point score of 97. This corresponds to an estimated probability of ~65% for functional independence. On the basis of the classification threshold from previous studies, this patient is therefore predicted to recover functionally. Similarly, application of the survival model (Figure 1B) results in 109 total points for this patient, corresponding to a 94% probability of survival.

### Table 1. Estimated Regression Coefficients of the Original, Recalibrated, and Novel Models Predicting Functional Recovery and Survival

<table>
<thead>
<tr>
<th>Model</th>
<th>Original</th>
<th>Recalibrated</th>
<th>Novel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>−5.782</td>
<td>−6.148</td>
<td>−5.112</td>
</tr>
<tr>
<td>Age</td>
<td>0.049</td>
<td>0.049</td>
<td>0.046</td>
</tr>
<tr>
<td>NIHSS*</td>
<td>0.272</td>
<td>0.272</td>
<td>0.196</td>
</tr>
<tr>
<td>Survival</td>
<td>−7.040</td>
<td>−7.373</td>
<td>−5.445</td>
</tr>
<tr>
<td>Age</td>
<td>0.049</td>
<td>0.049</td>
<td>0.037</td>
</tr>
<tr>
<td>NIHSS*</td>
<td>0.155</td>
<td>0.155</td>
<td>0.092</td>
</tr>
</tbody>
</table>

*Overall score on the NIHSS.
determined anew on the basis of the outcome frequencies in the data sets as before\textsuperscript{17} 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 by different prognostic models, we tested differences in the overall accuracies by McNemar’s test, and we present the estimated differences with 95% CIs according to Zhou and Qin.\textsuperscript{25} The analyses were performed with the R software environment, version 2.3.1, with the Design package by Harrell.\textsuperscript{26}

**Results**

Characteristics of the 5843 patients in the VISTA data set meeting the specified inclusion criteria are presented in Table 2. The BI after 90 days was recorded in 4441 patients, and 1970 of these had recovered, whereas 2471 had not functionally recovered. In another 978 patients without a recorded BI, information on mortality within this time frame was available. Of these, 607 (62.1%) had died, so that these were additionally classified as not functionally recovered. Therefore, for the functional recovery model, data from 5048 patients were available, of whom 1970 had recovered and 2471 + 607 = 3078 had either not recovered or had died. Independent of the availability of the BI, there was information on mortality for 5419 patients, of whom 4441 had survived and 978 were deceased. As shown in Table 2, age was slightly higher than in the original data sets (mean difference = 0.71 years, 95% CI = 0.04 to 1.38); similarly, neurologic impairment was less severe in the original sample (mean difference = 6.50, 95% CI = 6.15 to 6.84). In the VISTA data set, the proportion of women was higher than in the original data set (difference in proportions = 3.4%, 95% CI = 0.7% to 6.0%). Further details on patients’ characteristics are given in the original publications.\textsuperscript{17,19}

In the most stringent approach, the original models were applied to predict the patients’ outcomes, and the receiver

![Figure 1](image-url)

**Figure 1.** Nomograms for (A) the model predicting functional recovery (BI ≥ 95) vs no functional recovery (BI < 95) or mortality and for (B) the model predicting survival vs mortality.
operating characteristics are shown in Figure 2. According to the original thresholds, 3678 patients (72.9%) were classified correctly overall in the functional recovery model, with a more correct prediction of patients who did not recover than in those who did (90.8% and 44.9%, respectively). According to the survival model, 3815 patients (70.4%) were classified correctly overall, with 56.3% of patients who died and 73.5% of patients who survived.

In the second approach, the original models were recalibrated by using an adapted intercept. The deviations of the original and the new intercepts were estimated to be 0.36 (95% CI 0.29 to 0.44) for the functional recovery model and 0.33 (95% CI 0.26 to 0.41) for the survival model, showing that the predicted probabilities for favorable outcome in both models were systematically too low (see Figures 3 A and 3B).

Using the recalibrated models (see Table 1 for estimated regression coefficients and Figures 3C and 3D for calibration plots) led to slightly altered classifications, with overall 3735 patients (74.0%) being classified correctly for the functional recovery model (86.1% who did not recover and 55.1% who did recover). For the survival model, 4054 patients (74.8%) were predicted correctly (46.7% of patients who died and 81.0% of patients who survived). Thus, the accuracies were higher than in the original model (for the functional recovery model, difference = 1.1%, 95% CI = 0.1% to 2.3%, 2-sided \( P = 0.0026 \), and for the survival model, \( \text{difference} = 4.5\%, \quad 95\% \text{ CI} = 3.7\% \text{ to} \ 4.5\% \), 2-sided \( P = 2.2 \times 10^{-16} \)).

Finally, novel logistic regression models were developed by estimating new regression coefficients of the previously identified parameters (Table 1, right column). Thereby the recovery model predicted 3736 patients (74.0%) correctly overall (78.8% of those who did not recover and 66.6% of those who did). In the survival model, 4212 patients (77.7%) were classified correctly (38.2% of those who died and 86.4% of those who survived). Compared with the original model, the accuracy was higher overall (for functional recovery, \( \text{difference} = 1.2\%, \quad 95\% \text{ CI} = 0.1\% \text{ to} \ 2.3\% \), 2-sided \( P = 0.0452 \); for survival, \( \text{difference} = 7.4\%, \quad 95\% \text{ CI} = 6.4\% \text{ to} \ 8.3\% \), 2-sided \( P = 2.2 \times 10^{-16} \)). Although there was no difference in accuracies for the functional recovery model between the novel and the recalibrated prognosis (\( \text{difference} = 0.0\%, \quad 95\% \text{ CI} = -0.8\% \text{ to} \ 0.9\% \), 2-sided \( P = 1.0000 \)), accuracy was higher in the novel than in the recalibrated survival model overall (\( \text{difference} = 2.9\%, \quad 95\% \text{ CI} = 2.3\% \text{ to} \ 3.6\% \), 2-sided \( P = 2.2 \times 10^{-16} \)). However, this was mostly due to a more correct classification of the patients who survived, whereas <40% of patients who died were predicted correctly.

### Discussion

An early, simple, and reliable model to calculate the prognosis of likely outcome in stroke patients is desirable and useful, for both clinical practice and research purposes. We previously developed and externally validated models that fulfilled
the original models predicting functional recovery, indicating a good discrimination of the original model. For the survival model, the higher overall accuracy of the new model was accompanied by a worse prediction of deceased patients, so that this is not recommended over the recalibrated model for clinical practice. To a greater extent than for the refined models, it should be considered that the latter models with novel parameter estimates need to be validated in different samples to guarantee generalization.

This study has some limitations. First, the predictive accuracies identified in the most stringent approach may not seem to justify relying on the given prediction over clinical judgment. However, we have previously shown that clinical judgment by the admitting neurology resident is inferior to our models and correctly predicted 70% of all patients. Also, a simple recalibration of the intercept considerably improved the overall correct prediction of our models in VISTA. Second, our models do not consider imaging or laboratory investigations, which were impossible to obtain in our large original cohorts within an early time frame and a standardized evaluation protocol. Instead, we decided to focus our models on variables that are readily accessible and require neither a sophisticated technique nor a rigorous time frame. However, other studies have shown the prognostic value of magnetic resonance imaging in acute stroke, which has also become an inclusion criterion in thrombolysis trials with desmoteplase. Third, we have now shown the external validity of our models in 2 different populations of stroke 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 stroke patients, and subsequent studies are required to investigate the validity in other stroke populations.
Simple prognostic models may play an important role in future randomized trials in acute stroke. Patients included in these studies should have a high chance of incomplete recovery but low probability of mortality, which is usually unaffected by new medical treatment options. Therefore, it may be desirable to exclude patients with a high chance of complete spontaneous recovery and those with a high chance of mortality because their data are unlikely to contribute to a measurable treatment effect. We have previously shown that overall sample size and trial time can be reduced by eliminating potential nonresponders and by increasing the number of eligible patients compared with conventional study designs.3 Alternatively, prognosis-adjusted end points could be defined for patients with a high probability of functional recovery, as have already been used in several acute stroke trials.32–34

In conclusion, our original models can readily be applied in clinical practice and research settings with sufficient predictive accuracy, even in different patient populations. For patients included in clinical trials, a simple recalibration helps to adjust for a different case mix and is indeed recommended if a large data set is available.

Appendix

VISTA Steering Committee Members


Disclosures

R.L.S. serves as a consultant and is on the Advisory Board for Boehringer Ingelheim. The remaining authors report no conflicts.

References