Accuracy of Serial National Institutes of Health Stroke Scale Scores to Identify Artery Status in Acute Ischemic Stroke

Robert Mikulik, MD; Marc Ribo, MD; Michael D. Hill, MD; James C. Grotta, MD; Marc Malkoff, MD; Carlos Molina, MD; Marta Rubiera, MD; Raquel Delgado-Mederos, MD; Jose Alvarez-Sabin, MD; Andrei V. Alexandrov, MD; for the CLOTBUST Investigators

Background—Early recovery after intravenous thrombolysis can be observed in stroke; however, the utility of measuring clinical improvement to assess artery status has not been established. We sought to determine the accuracy of serial National Institutes of Health Stroke Scale (NIHSS) scores to detect complete early recanalization of the middle cerebral artery.

Methods and Results—Data from the CLOTBUST trial (Combined Lysis of Thrombus in Brain Ischemia Using Transcranial Ultrasound and Systemic tPA) were used to determine the most sensitive and specific NIHSS-derived parameter to identify complete recanalization. Then, reproducibility was tested against a separate patient population (Barcelona data set). NIHSS scores were determined before tissue plasminogen activator bolus and at 60 and 120 minutes in both data sets. Receiver operating characteristic curves were used to compare test performance. The accuracy of individual cutoffs was demonstrated by sensitivity, specificity, and positive and negative predictive values. A total of 122 patients in the CLOTBUST data set and 98 in the Barcelona data set received 0.9 mg/kg intravenous tissue plasminogen activator [mean age 69±12 versus 72±12 years, 57% male versus 51% male, median NIHSS 16 versus 17 points, mean time from onset to treatment 140±32 versus 177±59 minutes, and complete recanalization of the middle cerebral artery in 19% versus 17%). For identification of recanalization, an NIHSS score reduction of ≥40% offered the best tradeoff, with sensitivity, specificity, positive predictive value, and negative predictive value of 65%, 85%, 50%, and 91% at 60 minutes and 74%, 80%, 58%, and 89% at 120 minutes, respectively. Test performance was equal in the Barcelona data set.

Conclusions—Relative changes in serial NIHSS scores can serve as a simple clinical indicator of arterial status after intravenous thrombolysis. Accuracy parameters are affected by the process of recanalization and its varying clinical significance. (Circulation. 2007;115:2660-2665.)

Key Words: stroke ■ reperfusion ■ ultrasonography ■ thrombolysis

Post hoc exploratory analysis of the National Institute of Neurological Disorders and Stroke (NINDS) rt-PA Stroke Study1 showed that a neurological improvement by ≥5 points on the National Institutes of Health Stroke Scale (NIHSS) at 24 hours was observed significantly more often in patients treated with tissue plasminogen activator (tPA) than in placebo-treated patients.2 In addition, dramatic early clinical improvement was observed in stroke patients who achieved complete recanalization after treatment with intravenous thrombolysis.3-5 Such evidence suggests that serial NIHSS examinations within the first few hours are a useful and simple clinical indicator of early recanalization. The wide-ranging impact of such a simple clinical tool is that it can be used in place of, or in addition to, diagnostic tests such as transcranial ultrasound, computed tomographic angiography, magnetic resonance angiography, or digital subtraction angiography, which are limited by their availability, procedural risk, length of time to perform, and sometimes their cost.

The utility of a clinical surrogate marker for detecting recanalization is 2-fold. First, it can be used to identify a persistent occlusion after intravenous thrombolysis and thus select patients for additional rescue therapies such as intra-
arterial thrombolysis or mechanical thrombectomy. Second, it can be used as a suitable end point in designing clinical trials that would identify the biological activity of an intervention aimed at achieving early recanalization.

The NIHSS is the most widely used neurological deficit scale, with documented reliability, validity, and outcome predictive ability.11–13 We sought to determine the accuracy of serial NIHSS score measurements to detect complete recanalization of the middle cerebral artery (MCA) and to validate our results against a separate data set.

Methods

The present study used 2 independent patient data sets. The first (CLOTBUST data set) originates from the Combined Lysis of Thrombus in Brain Ischemia Using Transcranial Ultrasound and Systemic tPA (CLOTBUST) trial.16 This data set was used to determine the accuracy of serial NIHSS score measurements for detecting recanalization (development data set). The second set (Barcelona data set), obtained in Barcelona, Spain, apart from the CLOTBUST trial, validated the results against a separate patient population (validation data set).

Both data sets included patients with acute MCA occlusions (including isolated M1 occlusion, proximal M2 occlusion, tandem internal carotid artery–MCA occlusion, or terminal internal carotid artery occlusion) who were treated with intravenous tPA (0.9 mg/kg body weight, with 10% given as a bolus). CLOTBUST was a phase II multicenter, randomized clinical trial that determined the safety and signal-of-efficacy of adjuvant therapy with continuous transcra-nial Doppler (TCD) monitoring versus sham TCD monitoring.16 The Barcelona data set was an open-label series of consecutive patients with MCA occlusions diagnosed by TCD between January 2001 and June 2005 (excluding those enrolled in the CLOTBUST study). In CLOTBUST, patients were treated within a 0- to 3-hour window of symptom onset; in the Barcelona data set, they were treated within a 0- to 6-hour window.17

In both data sets, complete recanalization of the MCA was determined with the previously validated Thrombolysis in Brain Ischemia (TIBI) flow grading system (which consists of a 6-point scale, with a score of 0 indicating no flow and 5 indicating completely normal flow). TIBI demonstrated >90% accuracy compared with angiography.18–20 Complete recanalization was defined as TIBI 5 flow in the symptomatic artery. The NIHSS score was determined at baseline and at 60 and 120 minutes in both data sets by physicians trained and certified in NIHSS scoring.

In the CLOTBUST trial, treating physicians determined the NIHSS scores without knowledge of vessel recanalization. In the Barcelona data set, physicians were not blinded to the diagnosis made by TCD. In both data sets, however, physicians were not aware of the purposes of this analysis. In the CLOTBUST data set, patients with unknown NIHSS score at 60 minutes (4 patients) and at 120 minutes (11 patients) were excluded from the analysis. In the Barcelona data set, 3 patients with missing data at 120 minutes were excluded.

Statistical Analysis

Development (CLOTBUST) Data Set

Serial NIHSS scoring offers 3 options for describing clinical improvement: (1) the absolute value of the NIHSS score at a certain time point (ie, NIHSS60 at 60 minutes), or (2) the absolute improvement, ie, the difference between the NIHSS score at baseline (NIHSSbaseline) and at a certain time point (ΔNIHSS60=NIHSSbaseline minus NIHSS60), or (3) the relative improvement, ie, percent reduction from baseline to a certain time point (%NIHSS60=ΔNIHSS60 divided by NIHSSbaseline).

Our first step was to select the most “valuable” NIHSS-derived parameter for identifying complete MCA recanalization. This was gauged by its accuracy and performance regardless of the baseline NIHSS scores.

To measure the accuracy of different NIHSS-derived parameters, we constructed an empirical (nonparametric) receiver operating characteristic curve for each parameter and calculated the area under the receiver operating characteristic curve (AUC) with 95% CIs.21 The AUC is a measure of the accuracy of a diagnostic test that ranges from 0.5 (no diagnostic ability) to 1.0 (perfect diagnostic ability).

Because response to stroke treatment is governed by initial stroke severity, we then tested the performance of the NIHSS-derived parameter irrespective of NIHSSbaseline. This ensures that the accuracy of the selected parameter was similar across the range of NIHSSbaseline. To address this issue, we calculated whether there was an association (Spearman rank correlation coefficient [r]) between the value of the NIHSS-derived parameter and NIHSSbaseline in patients with complete recanalization. If such an association was absent, it meant that the same proportion of change (ie, NIHSS60 or ΔNIHSS60 or %NIHSS60) was achieved with different NIHSSbaseline scores (eg, 5 points or 50% NIHSS reduction similarly predicted recanalization whether the NIHSSbaseline was 10 or 20 points).

After the most valuable NIHSS-derived parameter was selected, we used the receiver operating characteristic curve to determine accuracy at varying thresholds, namely, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and likelihood ratios (positive likelihood ratio and negative likelihood ratio) with 95% CIs (Pearson χ²). We aimed to detect thresholds with the best tradeoffs between sensitivity and specificity, as well as PPV and NPV.

Validation (Barcelona) Data Set

The overall accuracy of the selected NIHSS-derived parameter was tested by comparing the AUC for the Barcelona data set with the CLOTBUST data set.21 Next, performance of the selected thresholds of the most valuable NIHSS-derived parameter from the CLOTBUST data set was assessed on the Barcelona set.

Statistical significance for intergroup differences was assessed by χ² test for categorical variables. For continuous variables, the 2-sample Student t test was used to compare the means and the Mann-Whitney U test to compare the medians. Repeated-measures ANOVA was used to compare time-event occurrence data. For comparison of AUC and NIHSS-derived parameters between the CLOTBUST and Barcelona data sets, a maximum difference was set to 0.01 to ensure comparative validity. A value of P<0.05 was considered significant.

The authors had full access to and take full responsibility for the integrity of the data. All authors have read and agree to the manuscript as written.

Results

In the CLOTBUST data set, we analyzed 122 patients at 60 minutes and 115 patients at 120 minutes. In the Barcelona data set, we analyzed 98 patients at 60 minutes and 95 at 120 minutes. Baseline characteristics of both data sets are shown in Table 1. There was a difference in the onset-to-treatment time. The overall mean NIHSS scores in the CLOTBUST data set in patients with early complete recanalization versus partial or no recanalization, shown in Figure 1, were significantly different (P<0.001, repeated-measures ANOVA).

The performance of different NIHSS-derived parameters (absolute value versus absolute improvement versus relative improvement) for diagnosing complete recanalization in the CLOTBUST data set is shown in Figure 2 and Table 2. There was no difference in accuracy between different NIHSS-derived parameters at 60 or 120 minutes as measured by AUC. The percent of NIHSS reduction at 60 and 120 minutes, however, was the only parameter that did not correlate with and was independent of the pretreatment stroke
severity, and therefore, it was selected for further analysis and validation.

The accuracy parameters at different thresholds of % NIHSS reduction at 60 and 120 minutes are listed in Table 3. The best tradeoff between sensitivity and specificity was achieved with a cutoff of ≥40% NIHSS reduction at 60 and 120 minutes. Such a threshold offered sensitivity of 65% (95% CI, 45% to 81%) and specificity of 85% (95% CI, 76% to 91%) and 80% (95% CI, 70% to 87%) at 60 and 120 minutes, respectively. In addition, a ≥40% NIHSS reduction at 60 and 120 minutes offered a positive likelihood ratio of 4.3 (95% CI, 2.5 to 6.7) and 3.7 (95% CI, 2.4 to 5.2) and a negative likelihood ratio of 0.4 (95% CI, 0.2 to 0.6) and 0.3 (95% CI, 0.2 to 0.5), respectively.

The best tradeoff between PPV and NPV is achieved with a cutoff of ≥80% NIHSS reduction at 60 and 120 minutes. Such a threshold offers a PPV of 67% (95% CI, 30% to 90%) and 86% (95% CI, 49% to 97%) and an NPV of 84% (95% CI, 76% to 89%) and 77% (95% CI, 68% to 84%) at 60 and 120 minutes, respectively. The association between ≥80% NIHSS reduction and recanalization was statistically significant (P<0.01 for 60 and 120 minutes). If NIHSS reduction was ≥80% at 2 hours, then 85% of patients achieved a modified Rankin score of 0 or 1 at 3 months; if NIHSS reduction was <80% at 2 hours, then 30% of patients achieved similar results at 3 months (P=0.003). In the Barcelona data set, the overall accuracy (AUC) of % NIHSS reduction and of selected thresholds (sensitivity, specificity, PPV, and NPV) did not differ from the CLOTBUST data set (Table 4).

**Discussion**

The present study showed that a serial neurological examination is a valuable measure of arterial status after intravenous thrombolysis, with similar performance in 2 different patient data sets. We demonstrated that considerable change in the NIHSS scores happens within the first 60 minutes of treatment after complete recanalization is achieved and thus can serve as an indicator of early complete recanalization. This is in agreement with other studies that showed that recanalization is associated with short-term improvement and long-term outcome.

The present study developed and validated a parameter from serial NIHSS score measurements suitable for identifying arterial status irrespective of pretreatment stroke severity. We showed that clinical improvement, measured as the NIHSS percentage reduction, performs independently of baseline stroke severity. We therefore consider NIHSS per-

---

**Table 1. Demographic Characteristics of CLOTBUST Data Set and Barcelona Data Set**

<table>
<thead>
<tr>
<th></th>
<th>CLOTBUST (n=126)</th>
<th>Barcelona (n=98)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, y</td>
<td>69±12</td>
<td>72±12</td>
<td>0.07</td>
</tr>
<tr>
<td>Female, n (%)</td>
<td>52 (43)</td>
<td>48 (49)</td>
<td>0.22</td>
</tr>
<tr>
<td>Onset to treatment time, min</td>
<td>140±32</td>
<td>177±59</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>Systolic blood pressure, mm Hg</td>
<td>160±22</td>
<td>155±26</td>
<td>0.2</td>
</tr>
<tr>
<td>Serum glucose, mg/dL</td>
<td>143±66</td>
<td>137±55</td>
<td>0.43</td>
</tr>
<tr>
<td>Baseline NIHSS, median (minimum–maximum)</td>
<td>16 (4–34)</td>
<td>17 (5–23)</td>
<td>0.96</td>
</tr>
<tr>
<td>Complete recanalization at 60 min, n (%)</td>
<td>23 (19)</td>
<td>17 (17)</td>
<td>0.77</td>
</tr>
<tr>
<td>Complete recanalization at 120 min, n (%)</td>
<td>31 (27)</td>
<td>29 (31)</td>
<td>0.58</td>
</tr>
</tbody>
</table>

Values are mean±SD unless otherwise noted.
percentage reduction as a better way to gauge the early recanalization process than the absolute numbers of the NIHSS scores or NIHSS point reduction. This is because patients with complete recanalization are more likely to achieve the same degree of relative reduction of the deficit rather than the same absolute improvement.

This somewhat contradicts the most sensitive measure of successful thrombolysis at 2 hours from treatment onset (total NIHSS score ≤5 points), derived from the NINDS rt-PA Stroke Study. If such a criterion is applied to the present data, however, sensitivity to detect recanalization for patients with an NIHSS baseline score <14 (77%) would be significantly different ($P<0.01$) from sensitivity for patients with NIHSS baseline score ≥14 (28%; not shown in results).

The present data have important implications for the design of clinical trials, because the absolute values of NIHSS score or NIHSS point reduction may not be accurate measures of early clinical improvement unless adjusted for baseline NIHSS score. Likewise, other studies have argued for a similar approach that adjusted 3- and 12-month clinical outcome to the baseline NIHSS score, which better evaluates the effect of treatment.

### TABLE 3. Diagnostic Properties at Different Thresholds of NIHSS60 and NIHSS120 Reduction in the CLOTBUST Data Set

<table>
<thead>
<tr>
<th>%NIHSS60</th>
<th>%NIHSS120</th>
<th>AUC (95% CI)</th>
<th>$r_1$ (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>≥0</td>
<td>91</td>
<td>0.74 (0.63 to 0.86)</td>
<td>0.048 (0.09 to 0.74)</td>
</tr>
<tr>
<td>≥10</td>
<td>83</td>
<td>0.74 (0.62 to 0.88)</td>
<td>0.49 (0.1 to 0.74)</td>
</tr>
<tr>
<td>≥20</td>
<td>74</td>
<td>0.75 (0.62 to 0.88)</td>
<td>0.06 (-0.36 to 0.45)</td>
</tr>
<tr>
<td>≥30</td>
<td>70</td>
<td>0.78 (0.68 to 0.89)</td>
<td>0.52 (0.21 to 0.73)</td>
</tr>
<tr>
<td>≥40</td>
<td>65</td>
<td>0.79 (0.71 to 0.88)</td>
<td>0.61 (0.32 to 0.79)</td>
</tr>
<tr>
<td>≥50</td>
<td>52</td>
<td>0.74 (0.63 to 0.85)</td>
<td>0.09 (-0.43 to 0.26)</td>
</tr>
</tbody>
</table>

Se indicates sensitivity; Sp, specificity; PPV, positive predictive value; NPV, negative predictive value; and LR+ or LR−, likelihood ratio positive or negative, respectively.

The present study provides sensitivities and specificities at different thresholds of NIHSS percentage reduction. We showed that the optimal tradeoff between sensitivity and specificity is reached with ≥40% NIHSS score reduction compared with baseline. This means that if all patients after intravenous thrombolysis who have improvement in NIHSS score by <40% are sent for rescue interventional therapy, 85% of all occlusions will be indicated, and 65% of all recanalizations will be saved from intervention. In terms of likelihood, compared with patients who did not experience recanalization, patients whose occlusion recanalized had odds of 4.3 for ≥40% NIHSS score reduction and odds of 0.4 for <40% NIHSS score reduction.

For different purposes, different cutoffs may apply, ie, a cutoff of ≥60% NIHSS score reduction, because of a high specificity, may be a more appropriate end point for a clinician who does not have a TCD at bedside. The processes of either brain stunning (delayed improvement after recanalization) or clinically “silent” recanalization (ie, no improvement despite recanalization) may be responsible for limited sensitivity. In addition, limited specificity (ie, improvement despite lack of recanalization) may be explained by partial recanalization or the presence of good collateral flow.

We also showed that NIHSS score reduction by ≥80% at 2 hours after the start of intravenous thrombolytic therapy predicts recanalization with a PPV of 86% and an NPV of 77%. The present results confirm the findings from the NINDS trial, which suggested that major neurological im-

### TABLE 4. Comparison of Accuracy of NIHSS Reduction Between the CLOTBUST Data Set and Barcelona Data Set

<table>
<thead>
<tr>
<th></th>
<th>CLOTBUST</th>
<th>Barcelona</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUC</td>
<td>%NIHSS60</td>
<td>0.75 (0.62 to 0.88)</td>
<td>0.76 (0.63 to 0.88)</td>
</tr>
<tr>
<td></td>
<td>%NIHSS120</td>
<td>0.74 (0.63 to 0.85)</td>
<td>0.72 (0.57 to 0.87)</td>
</tr>
<tr>
<td>Sensitivity, %</td>
<td>≥40%NIHSS60</td>
<td>65 (45 to 81)</td>
<td>47 (26 to 69)</td>
</tr>
<tr>
<td></td>
<td>≥40%NIHSS120</td>
<td>74 (57 to 86)</td>
<td>59 (41 to 74)</td>
</tr>
<tr>
<td>Specificity, %</td>
<td>≥40%NIHSS60</td>
<td>85 (76 to 91)</td>
<td>85 (76 to 91)</td>
</tr>
<tr>
<td></td>
<td>≥40%NIHSS120</td>
<td>80 (70 to 87)</td>
<td>86 (76 to 93)</td>
</tr>
<tr>
<td>PPV, %</td>
<td>≥80%NIHSS60</td>
<td>67 (30 to 90)</td>
<td>50 (15 to 85)</td>
</tr>
<tr>
<td></td>
<td>≥80%NIHSS120</td>
<td>86 (49 to 97)</td>
<td>83 (44 to 97)</td>
</tr>
<tr>
<td>NPV, %</td>
<td>≥80%NIHSS60</td>
<td>84 (76 to 89)</td>
<td>84 (75 to 90)</td>
</tr>
<tr>
<td></td>
<td>≥80%NIHSS120</td>
<td>77 (63 to 84)</td>
<td>73 (63 to 81)</td>
</tr>
</tbody>
</table>

PPV indicates positive predictive value; NPV, negative predictive value. Numbers in parentheses represent 95% CIs.
Improvement at 24 hours may be a useful surrogate for thrombolytic activity. The present data are, however, distinct, because we related NIHSS scores directly to recanalization rather than to 3-month outcome.

The main limitation of the present study is that the patient population was limited to MCA occlusions. The present results could not be broadened for the entire spectrum of stroke patients, especially those with non-MCA infarction or small-vessel stroke. We recognize, additionally, that the lack of binding of the TCD evaluation in the Barcelona cohort may have biased the assessment of the NIHSS score.

The main strength of the present study is that the performance of NIHSS percentage reduction was equivalent in 2 separate and characteristically different (ie, age and treatment-to-onset time) data sets. These results can therefore be generalized to patients with MCA occlusion within 6 hours from stroke onset.

In conclusion, changes in serial NIHSS scores can be used as a clinical indicator of arterial status, although accuracy is affected by the process of recanalization and its varying clinical significance. Ultimately, the present data may have an impact on the design of clinical trials, as well as a practical application in emergency settings. Further studies will be needed to determine the final implications for patient management.

**Sources of Funding**

This study was supported by NINDS grants 1K23NS02229-01 and 1P50NS44227. Canadian sites were supported by the Canadian Institutes of Health Research and the Alberta Heritage Foundation for Medical Research. The CLOTBUST trial is an investigator-sponsored trial (protocol A2207s, Genentech, Inc) that is exempt from investigational new drug status by the Food and Drug Administration. Spencer Technologies, Seattle, Wash, provided power-motion Doppler units and technical support for all participating study sites. DWL, Multigon, and Nicolet also provided portable equipment to hospitals in Houston, Tex.

**Disclosures**

None.

**References**


CLINICAL PERSPECTIVE
The present study is the first to provide data on the accuracy of serial National Institutes of Health Stroke Scale (NIHSS) scores for detecting recanalization of the middle cerebral artery in patients treated with intravenous tissue plasminogen activator. The utility of a clinical surrogate marker for detecting recanalization is 2-fold: (1) to identify a persistent occlusion after intravenous thrombolysis, which may enable the selection of patients suitable for rescue therapies, and (2) to serve as an end point in clinical trials that allow assessment of the efficacy of an intervention aimed at achieving early recanalization. The present study used data on 126 patients from the CLOTBUST trial to determine the accuracy of serial NIHSS score measurements for detecting recanalization and then validated the results against a separate patient population (98 patients from Barcelona). The best tradeoff between sensitivity and specificity was achieved with a cutoff of ≥40% reduction in NIHSS score at 60 and 120 minutes. Such a threshold offered sensitivity of 65% (95% CI, 45% to 81%) and 74% (95% CI, 57% to 86%) and specificity of 85% (95% CI, 76% to 91%) and 80% (95% CI, 70% to 87%) at 60 and 120 minutes, respectively. The best tradeoff between positive and negative predictive value was achieved with a cutoff of ≥80% NIHSS reduction at 60 and 120 minutes. Such a threshold offered a positive predictive value of 67% (95% CI, 30% to 90%) and 86% (95% CI, 49% to 97%) and a negative predictive value of 84% (95% CI, 76% to 89%) and 77% (95% CI, 68% to 84%) at 60 and 120 minutes, respectively. Thus, the present study showed that a serial neurological examination is a valuable measure of arterial patency after intravenous thrombolysis, with similar performance in 2 different patient data sets.

Go to http://cme.ahajournals.org to take the CME quiz for this article.
Accuracy of Serial National Institutes of Health Stroke Scale Scores to Identify Artery Status in Acute Ischemic Stroke

Robert Mikulik, Marc Ribo, Michael D. Hill, James C. Grotta, Marc Malkoff, Carlos Molina, Marta Rubiera, Raquel Delgado-Mederos, Jose Alvarez-Sabin and Andrei V. Alexandrov

for the CLOTBUST Investigators

Circulation. 2007;115:2660-2665; originally published online May 14, 2007;
doi: 10.1161/CIRCULATIONAHA.106.651026

Circulation is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2007 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7322. Online ISSN: 1524-4539

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circ.ahajournals.org/content/115/20/2660

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation is online at:
http://circ.ahajournals.org//subscriptions/