Relationship Between Neurologic Deficit Severity and Final Functional Outcome Shifts and Strengthens During First Hours After Onset

Jeffrey L. Saver, MD; Hernan Altman, MBA

Background and Purpose—Early neurological deficit severity is the most important determinant of final functional outcome in acute ischemic stroke. However, deficit severity frequently changes during the first hours and days postonset.

Methods—Analysis of control group patients enrolled in the 2 National Institute of Neurological Disorders and Stroke tissue-type plasminogen activator trials. Neurological deficit severity was measured serially using the National Institutes of Health Stroke Scale (NIHSS) at 1 to 3 hours postonset, 3 to 5 hours, 24 hours, 7 to 10 days, and 90 days. Final global disability outcome was assessed at 90 days using the modified Rankin Scale.

Results—Among the 312 patients, median neurological deficit severity on the NIHSS improved throughout the 90-day observation period, from 15 (interquartile range, 9.5–20) at 1 to 3 hours, to 12 (interquartile range, 6–19) at 24 hours, to 7 (interquartile range, 2–19) at 90 days. Between 1-to-3-hours to 24 hours, more patients spontaneously improved than worsened: 39.1% versus 17.6% ($\textit{P} < 0.001$). NIHSS scores associated with individual final modified Rankin Scale global disability ranks shifted to lower values over time; e.g., patients with a final modified Rankin Scale of 2 had the following median NIHSS scores: 12 at 1 to 3 hours, 10 at 3 to 5 hours, 9 at 24 hours, and 3 at 90 days. Correlation coefficients between NIHSS and the final modified Rankin Scale increased over time, from 0.51 at 1 to 3 hours, to 0.72 at 24 hours, to 0.87 at 90 days.

Conclusions—During the first 24 hours after onset, spontaneous improvement occurs in 2 of 5 acute ischemic stroke patients. The NIHSS scores associated with individual global disability ranks decrease over time. Neurological deficit severity increasingly predicts final disability outcome, accounting for one quarter of the variance at 1 to 3 hours, one half at 24 hours, and three quarters at 90 days. (Stroke. 2012;43:1537-1541.)

Key Words: ischemic stroke • deficit • prognosis • disability • examination

Presenting stroke severity, most commonly assessed using the National Institutes of Health Stroke Scale (NIHSS), is the single most important predictor of final outcome in ischemic stroke patients.1–3 Accordingly, adjusting for presenting stroke deficit severity as a baseline covariate is a crucial step in the analysis of treatment trials and quality of care registries. However, the relationship between early stroke deficits and final outcome is likely to be time-dependent. Two major potential contributing processes to time dependence can be identified: (1) In the first few hours after onset, stroke deficits are unstable, prone to sudden improvements with recanalization or sudden worsenings with clot propagation or collateral failure; whereas, beyond the first 6 to 10 hours, most of the injury has already accrued and deficits tend not to change suddenly.4,5 (2) In the subacute period from 12 hours to 7 days, most patients without complications experience steady modest deficit improvement, because of resolution of diaschisis and the beneficial effects of early neuroplasticity.6

This study sought to delineate the degree to which time of measurement alters the relationship of early stroke deficit severity to final functional outcome by analyzing the serial measures available in the public data set of the 2 National Institute of Neurological Disorders and Stroke (NINDS) tissue-type plasminogen activator (tPA) treatment trials.

Methods
Analysis was performed on placebo control patients enrolled onto the 2 NINDS-tPA trials.7 In these trials, NIHSS measures that could predict final functional outcome were obtained at 1 to 3 hours, 3 to 5 hours, 24 hours, and 7 to 10 days postonset. We used as the primary outcome measure at 90 days global disability status on the modified Rankin Scale (mRS).

The course of change in the NIHSS was assessed in 3 ways. The median NIHSS was calculated at each time point. The proportions of patients who improved or worsened on the NIHSS in the early time
window between 1 to 3 hours to 24 hours were calculated, using thresholds of changing by 2 or more and 4 or more scale points. In addition, the proportion of patients who improved or worsened on the NIHSS in the early time window between 1 to 3 hours to 24 hours by normalized changes of 25% and 50% were calculated. The normalized change is the observed change divided by total possible change.8

Following the convention in the NINDS-tPA trials, and in acute stroke trials generally, patients who had died by an evaluation time point were assigned the worst possible NIHSS, 42, at that time point. For the few patients known alive, but who were not neurologically examined at a time point, the main analysis was performed on the NIHSS scores that included the imputed values provided in the public data set. In the placebo group, there were no such patients at 1 to 3 hours, 3 to 5 hours, or 24 hours; 6 patients (1.9%) at 7 to 10 days; and 6 patients (1.9%) at 90 days. Sensitivity analysis censoring missing data showed similar findings as did the main analysis.

Associations between early course NIHSS measures and final mRS were analyzed using correlation coefficients. For comparison, the association of day-90 NIHSS with day-90 mRS was also analyzed. The association of concurrent NIHSS with mRS indicates the maximum possible explanatory value of stroke deficit severity in accounting for global disability.

To assess the impact of NIHSS score evolution on identifying patients likely to have excellent outcomes, who not be informative if enrolled in clinical trials, we identified the NIHSS thresholds that identified patients with a >75% chance of reaching a 90-day mRS score of 0 to 1, and the proportion of the population these thresholds would exclude.

**Results**

Evolution of the NIHSS score over time is shown in Table 1. Among the 312 control patients, median stroke deficit severity improved throughout the 90-day observation period. For example, the median NIHSS improved from 15 (interquartile range [IQR], 9.5–20) at 1 to 3 hours, to 7 (IQR, 2–19) at 90 days. In the early time period between 1 to 3 hours to 24 hours, the proportion of patients who spontaneously improved outnumbered the proportion who spontaneously worsened. Improvements by 4 or more points occurred in 39.1% versus worsening in 17.6% ($P<0.001$). Improvements by 2 or more points occurred in 53.8% versus worsening in 23.7% ($P<0.001$). Similarly, normalized change improvements by 50% or more occurred in 20.8% versus worsening in 2.2% ($P<0.001$), and normalized change improvements by 25% or more occurred in 39.7% versus worsening in 8.0% ($P<0.001$). The distribution of the NIHSS became progressively more bimodal over time, as more patients experienced fatal outcome, with extreme assigned NIHSS score of 42, and surviving patients experienced progressive deficit improvement (Figure 1).

The correlation between the NIHSS and the final mRS increased over the 90-day period (Table 1). The 1-to-3-hour-postonset NIHSS accounted for 28% of the variance in final disability outcome, the 24-hour NIHSS 52%, and the day-90 NIHSS 76%.

In a subgroup analysis, the control cohort was split between patients who had their baseline NIHSS measured within the first 90 minutes ($n=156$) and patients who had their baseline NIHSS measured between 91 to 180 minutes ($n=156$). The proportion of patients with early spontaneous substantial improvement (by 4 or more points) between baseline and 24 hours was nominally higher among patients with first NIHSS measured at 0 to 90 minutes than among those with first NIHSS measured at 91 to 180 minutes; however, the difference did not reach statistical significance (42.4% versus 36.5%; $P=0.29$). Similarly, the correlation coefficients for predicting final 90-day mRS were nominally lower among those with first NIHSS measured at 0 to 90 minutes (0.47 [95% CI, 0.33–0.59]), than among those with first NIHSS measured at 91 to 180 minutes (0.54 [95% CI, 0.42–0.64]), but the difference did not reach statistical significance.

A shift to lower values over time was noted in the NIHSS scores associated with individual final mRS ranks (Table 2, Figure 2). For example, patients with a final mRS of 2 had the following median NIHSS scores: 1 to 3 hours, 12 (IQR, 8–16.5); 3 to 5 hours, 10 (IQR, 7–13); 24 hours, 9 (IQR, 5.5–12); 7 to 10 days, 6 (IQR, 4–8.5); and 90 days, 3 (IQR, 2–4). Conversely, a shift to higher values was noted in the final mRS scores associated with the same NIHSS scores obtained at different time points (Table 3).

The NIHSS values below which 75% of population were destined for an excellent final mRS outcome of 0 to 1 were 8 at 1 to 3 hours, 5 at 3 to 5 hours, 6 at 24 hours, and 5 at 7 to 10 days (versus 3 at 90 days). The proportion of the placebo population that would be excluded by these thresholds were 7.7% at 1 to 3 hours, 10.3% at 3 to 5 hours, 21.8% at 24 hours, and 28.5% at 7 to 10 days (versus 27.2% at 90 days).

**Discussion**

This analysis of the serial NIHSS assessments performed in the control arm of the 2 NINDS-tPA trials demonstrates 3 key findings critical for clinical trial and quality improvement activities.

First, stroke deficit severity is highly labile in the first few hours postonset. Both early improvements and early worsening occur, but early spontaneous improvement is substantially more common. Substantial spontaneous improvements during the first 24 hours occur in nearly 40% of patients, and are twice as common as is spontaneous worsening. This finding confirms and extends observations of frequent early improvement in small, single-center studies.4,5
Second, the prognostic value of the NIHSS score for final functional outcome improves as time from stroke onset increases. The influence of very early NIHSS measures on final outcome is attenuated when subsequent fluctuations alter patient course. As stroke deficit severity lability declines over time, the predictive power of the NIHSS grows. An NIHSS assessed in the first 3 hours after onset accounts for only one quarter of the variation in final outcome. By 24

<table>
<thead>
<tr>
<th>90-Day Rankin</th>
<th>NIHSS 1 to 3-Hour Median (IQR)</th>
<th>NIHSS 3 to 5-Hour Median (IQR)</th>
<th>NIHSS 24-Hour Median (IQR)</th>
<th>NIHSS 7 to 10-Day Median (IQR)</th>
<th>NIHSS 90-Day Median (IQR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>8 (6–11)</td>
<td>6 (3–10)</td>
<td>3 (1–5)</td>
<td>1 (0–2.5)</td>
<td>0 (0–1)</td>
</tr>
<tr>
<td>1</td>
<td>10 (7–15)</td>
<td>8 (5.75–11)</td>
<td>5 (4–8.25)</td>
<td>2 (1–5)</td>
<td>1.5 (1–3)</td>
</tr>
<tr>
<td>2</td>
<td>12 (8–16.5)</td>
<td>10 (7–13)</td>
<td>9 (5.5–12)</td>
<td>6 (4–8.5)</td>
<td>3 (2–4)</td>
</tr>
<tr>
<td>3</td>
<td>14 (10.5–19)</td>
<td>13 (10–19)</td>
<td>14 (11–17)</td>
<td>10 (7.5–14)</td>
<td>7 (5–9.5)</td>
</tr>
<tr>
<td>4</td>
<td>18 (13.75–22)</td>
<td>16 (11–20)</td>
<td>17.5 (11.75–22)</td>
<td>15 (10–18.25)</td>
<td>10 (8–15)</td>
</tr>
<tr>
<td>5</td>
<td>21 (16–24)</td>
<td>20 (14–22)</td>
<td>22 (19–25)</td>
<td>21 (19–22)</td>
<td>19 (16–21.5)</td>
</tr>
</tbody>
</table>

NIHSS indicates National Institutes of Health Stroke Scale; IQR, interquartile range.
hours after onset, the NIHSS accounts for one half of the final outcome. The final NIHSS, measured concurrently with final global disability, accounts for fully three quarters of the final outcome.

Third, a marked shift in the relationship between particular NIHSS score values and final outcome occurs over time. Given that the NIHSS typically improves in almost all patients who do not have fatal outcome, the same NIHSS score will have a worse prognosis when present later in the course. Steady reductions over time in the median NIHSS associated with final 90-day mRS ranks were seen for each of ranks 0 to 4. For example, patients destined for a final independent, but disabled, 90-day outcome (mRS 2) had an NIHSS median score of 12 at 1 to 3 hours, 10 at 3 to 5 hours, 9 at 24 hours, 6 at 1 week, and 3 at 90 days. Variations in this pattern were seen for the severe outcomes levels of mRS 5 and 6. For rank 5, associated median NIHSS scores were severe and unchanging over time. For rank 6 (mortality), the associated NIHSS scores increased over time, as an increasing proportion of patients had fatal outcome and were assigned an extreme NIHSS score of 42.

These findings have important consequences for prognostic modeling in clinical trials and quality improvement activities. Useful prognostic models almost invariably include a measure of initial stroke deficit severity as a predictor variable, as early deficit is the single most important prognostic factor in acute stroke. However, most models that have been presented in the literature have treated the relationship between initial deficit severity and final outcome as unaffected by the time of initial NIHSS assessment.1–3 This study’s findings demonstrate 2 weaknesses of such a time-invariant approach. First, because the explanatory power of the NIHSS increases over time, the prognostic weight assigned to the NIHSS should similarly increase. Second, given that the relationship between particular NIHSS score values and final outcome shifts over time, the same score can portend very different outcomes if it was assessed at 1 to 3 hours, 3 to 5 hours, or at 24 hours. Prognostic models designed for broad populations therefore should incorporate time from stroke onset until measurement of the baseline NIHSS as an additional variable or interaction term.

Clinical trials of different therapeutic strategies typically enroll patients onto different time windows. Intravenous thrombolytic agents have been successfully trialed in the 0-to-4.5-hour window, whereas mechanical endovascular therapies have been generally studied in the 3-to-8-hour window, and neuroplasticity enhancing therapies have typically been studied in the 24-to-72-hour window. The findings from this study indicate that prognostic models derived from populations with baseline measures obtained in a particular time period should generally be applied only to populations having baseline measures assessed in similar time periods.

The time-specific correlations reported in this study can also be helpful in designing clinical trial entry criteria and win criteria. NIHSS thresholds to identify likely uninformative patients for a particular win criterion shift lower and increase in their population impact over time. For instance, for the win criterion of a nondisabled outcome (mRS 0–1) at 90 days, patients very likely to be uninformative because of high rates of win outcomes in the placebo group would be those with NIHSS scores of 8 at 1 to 3 hours, versus 6 at 24 hours; the 1-to-3-hour cutoff would exclude <10% of NINDS-study-like patients, but the 24-hour cutoff would exclude 22%.

This study has limitations. The 2 NINDS trial control cohorts were enrolled in the early 1990s. Supportive care

---

**Table 3.** Final 90-Day Global Disability Outcomes Associated With NIHSS Scores Between 9–12 at Different Time Points

<table>
<thead>
<tr>
<th>NIHSS 9–12 Timing</th>
<th>Excellent (0–1) 90-Day mRS</th>
<th>Good (0–2) 90-Day mRS</th>
<th>Mean 90-Day mRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 to 3 h</td>
<td>35.0%</td>
<td>50.0%</td>
<td>2.7</td>
</tr>
<tr>
<td>3 to 5 h</td>
<td>27.9%</td>
<td>47.5%</td>
<td>2.8</td>
</tr>
<tr>
<td>24 h</td>
<td>16.4%</td>
<td>42.6%</td>
<td>2.9</td>
</tr>
<tr>
<td>7 to 10 d</td>
<td>7.1%</td>
<td>16.7%</td>
<td>3.5</td>
</tr>
<tr>
<td>90 d</td>
<td>0.0%</td>
<td>5.6%</td>
<td>3.6</td>
</tr>
</tbody>
</table>

NIHSS indicates National Institutes of Health Stroke Scale; mRS, modified Rankin Scale.

---

**Figure 2.** Median NIHSS scores over time among surviving patients with different 3-month mRS global disability scores. The NIHSS scores associated with individual mRS ranks decline substantially over time. For IQRs, see Table 2.
practices have evolved in several subtle ways in the subsequent decade-and-a-half, and the serial course of current patients may be mildly different. This study examined only the NIHSS, and not other measures of stroke deficit severity. However, the NIHSS is the most widely employed formal assessment in clinical trials and in clinical practice worldwide. The data set analyzed was modest in size and did not have data on all time points of interest (eg, 6 hours, 12 hours). Confirmation and extension of our findings in larger and more varied databases is needed.

Conclusions
In conclusion, serial stroke deficit severity measurement in supportively treated patients demonstrates that spontaneous early improvement is common in acute ischemic stroke, severity of stroke deficits becomes increasingly determinative of final outcome over time, and the same deficit severity score is associated with worse outcome when present later in a patient’s course. Accordingly, time from onset to initial measurement of stroke deficit severity is an important candidate variable for inclusion in prognostic outcome models.

Sources of Funding
This study was sponsored in part by NIH-NINDS awards U01 NS 44364 and P50 NS044378 and an American Heart Association Pharmacy Roundtable Health Outcomes Research Center Award.

Disclosures
The University of California Regents receive funding for J.L.S.’s services as a scientific consultant regarding trial design and conduct to BrainsGate, CoAxia, ev3, Talecris, PhotoThera, and SYGNIS (all modest). J.L.S. is an investigator in the NIH FAST-MAG, MR RESCUE, ICES, CUFFS, CLEAR-ER, and IMS III multicenter clinical trials, for which the UC Regents receive payments based clinical trial performance; has served as an unpaid site investigator in a multicenter trials run by Lundbeck and Mitsubishi, for which the UC Regents received payments based on the clinical trial contracts for the number of subjects enrolled; is a site investigator in a multicenter registry run by Concentric, for which the UC Regents received payments based on the clinical trial contracts for the number of subjects enrolled; is an employee of the University of California, which holds a patent on retriever devices for stroke; and is funded by NIH-NINDS awards P50 NS044378 and U01 NS 44364. H.A. is an employee of BrainsGate.

References
Relationship Between Neurologic Deficit Severity and Final Functional Outcome Shifts and Strengthens During First Hours After Onset

Jeffrey L. Saver and Hernan Altman

Stroke. 2012;43:1537-1541; originally published online April 5, 2012;
doi: 10.1161/STROKEAHA.111.636928

Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2012 American Heart Association, Inc. All rights reserved.
Print ISSN: 0039-2499. Online ISSN: 1524-4628

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/43/6/1537

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Stroke 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 Stroke is online at:
http://stroke.ahajournals.org/subscriptions/