

# Clinical interpretation and use of stroke scales

Scott E Kasner

No single outcome measure can describe or predict all dimensions of recovery and disability after acute stroke. Several scales have proven reliability and validity in stroke trials, including the National Institutes of Health stroke scale (NIHSS), the modified Rankin scale (mRS), the Barthel index (BI), the Glasgow outcome scale (GOS), and the stroke impact scale (SIS). Several scales have been combined in stroke trials to derive a global statistic to better define the effect of acute interventions, although this composite statistic is not clinically tenable. In practice, the NIHSS is useful for early prognostication and serial assessment, whereas the BI is useful for planning rehabilitative strategies. The mRS and GOS provide summary measures of outcome and might be most relevant to clinicians and patients considering early intervention. The SIS was designed to measure the patient's perspective on the effect of stroke. Familiarity with these scales could improve clinicians' interpretation of stroke research and their clinical decision-making.

## Introduction

The application of results from stroke trials to clinical practice needs interpretation and integration of stroke-outcome measures. The major issues that relate to stroke outcome that are amenable to measurement are neurological deficit (eg, hemiparesis or aphasia), loss of ability of perform specific tasks (eg, feeding oneself or walking), loss of ability to function in normal roles and activities (eg, employment or hobbies), and quality of life. WHO developed the International Classification of Functioning, Disability, and Health to provide a standard language for the characterisation of these domains and to broadly address the concepts of medical and social disability as a composite of body structures, functions, activity, and participation.<sup>1</sup> These issues might be conceptually distinct, and all can affect the perception of one's health, but in practice they overlap substantially, especially after stroke. Among patients with stroke, motor and language dysfunction substantially affect all these domains simultaneously, and therefore the validity and clinical use of these sharp distinctions must be called into question. However, all must be encompassed in the assessment of recovery after stroke.

Several tools exist to measure stroke outcomes, but they are used inconsistently among trials and their relevance may not be clear to practising physicians. Moreover, no single measure fully describes or predicts all dimensions of stroke recovery and disability. A review of outcome measures used in 51 studies of acute stroke showed that 14 impairment measures, 11 activity or participation measures, one quality-of-life measure, and eight miscellaneous other measures were used.<sup>2</sup> Some of the most widely used scales are the National Institutes of Health stroke scale (NIHSS), the modified Rankin scale (mRS), the Barthel index (BI), and the stroke impact scale (SIS).<sup>3-5</sup> Each scale is unique, and understanding the differences is critically important for both appropriate use in clinical practice and interpretation of results reported in a clinical trial. Every scale needs to have proven reliability and defined validity, and interpretation of results requires familiarity with the characteristics of the scale.

Scales that measure neurological deficits or specific body functions can be used especially well for triage and to guide acute-treatment decisions. The NIHSS, for example, is a valuable tool for initial assessments of patients with stroke in emergency departments, hospitals, or in the prehospital setting, and is predictive of subsequent resource use and long-term outcome.<sup>6-8</sup> The mRS and BI are commonly used to assess components of disability, such as activity and participation after stroke, and can be used to guide rehabilitation plans. The SIS was designed to gain insight into the patient's perspective on the effect of stroke.<sup>5</sup> Because no individual measure fits all these roles, a composite measure, such as a global statistic derived from the scores of multiple scales, has been advocated to improve assessment of the effect of acute interventions.<sup>9</sup>

The purpose of this review is to educate clinicians about the use and misuse of these scales in the assessment of stroke research and in practice.

## Stroke scales

### National Institutes of Health stroke scale

The NIHSS is a 15-item impairment scale, which provides a quantitative measure of key components of a standard neurological examination (panel 1).<sup>8,10</sup> The scale assesses level of consciousness, extraocular movements, visual fields, facial muscle function, extremity strength, sensory function, coordination (ataxia), language (aphasia), speech (dysarthria), and hemi-inattention (neglect).<sup>11,12</sup> An additional item that measures distal motor function has been used in a few drug trials, but is not widely used in ongoing research or in clinical practice. The NIHSS was designed to assess differences in interventions in clinical trials, although its use is increasing in patient care as an initial assessment tool and in planning post-acute care disposition.<sup>6,13</sup>

### Reliability and validity

The NIHSS has established reliability and validity for use in prospective clinical research, and predictive validity for long-term stroke outcome.<sup>10,14,15</sup> The reliability of a rating scale is a quantitative measure of its reproducibility.

*Lancet Neurol* 2006; 5: 603-12

Comprehensive Stroke Center,  
Department of Neurology,  
University of Pennsylvania  
Medical Center, Philadelphia,  
PA 19104, USA (S E Kasner MD)

Correspondence to:  
Scott E Kasner  
kasner@mail.med.upenn.edu

**Panel 1: Current form of the NIHSS****1a Level of consciousness\***

- 0=Alert
- 1=Not alert, arousable
- 2=Not alert, obtunded
- 3=Unresponsive

**1b Questions**

- 0=Answers both correctly
- 1=Answers one correctly
- 2=Answers neither correctly

**1c Commands**

- 0=Performs both tasks correctly
- 1=Performs one task correctly
- 2=Performs neither task

**2 Gaze**

- 0=Normal
- 1=Partial gaze palsy
- 2=Total gaze palsy

**3 Visual fields**

- 0=No visual loss
- 1=Partial hemianopsia
- 2=Complete hemianopsia
- 3=Bilateral hemianopsia

**4 Facial palsy\***

- 0=Normal
- 1=Minor paralysis
- 2=Partial paralysis
- 3=Complete paralysis

**5a Left motor arm**

- 0=No drift
- 1=Drift before 10 s
- 2=Falls before 10 s
- 3=No effort against gravity
- 4=No movement

**5b Right motor arm**

- 0=No drift
- 1=Drift before 10 s
- 2=Falls before 10 s
- 3=No effort against gravity
- 4=No movement

**6a Left motor leg**

- 0=No drift
- 1=Drift before 5 s
- 2=Falls before 5 s
- 3=No effort against gravity
- 4=No movement

**6b Right motor leg**

- 0=No drift
- 1=Drift before 5 s
- 2=Falls before 5 s
- 3=No effort against gravity
- 4=No movement

**7 Ataxia\***

- 0=Absent
- 1=One limb
- 2=Two limbs

**8 Sensory**

- 0=Normal
- 1=Mild loss
- 2=Severe loss

**9 Language**

- 0=Normal
- 1=Mild aphasia
- 2=Severe aphasia
- 3=Mute or global aphasia

**10 Dysarthria\***

- 0=Normal
- 1=Mild
- 2=Severe

**11 Extinction/inattention**

- 0=Normal
- 1=Mild
- 2=Severe

\* These items are dropped in the modified NIH stroke scale (mNIHSS). The sensory item is scored as 0=normal and 1=abnormal in the mNIHSS.<sup>22</sup> Reproduced from *Stroke*, by permission of Lippincott Williams and Wilkins.<sup>21</sup>

Intraobserver reliability of the scale was shown at a high level, where an initial rating and then rerating 3 months later showed an interclass correlation coefficient (ICC) of 0.93.<sup>16</sup> The ICC can be interpreted as a weighted kappa ( $\kappa$ ) statistic, where an ICC of 1 suggests perfect reliability and an ICC of 0.8 is generally deemed to represent excellent reliability.<sup>10</sup> Interobserver reliability was also high, with an overall ICC of 0.95.<sup>16</sup> This high reliability has been

published mainly in studies in which the raters underwent formal training and certification of the scale with a standard videotaped programme. Retrospective assessments of clinical trial data indicate that 11 of the 15 items of the original NIHSS could have significantly greater reliability and validity than previously reported, whereas four items were identified as poorly reproducible or redundant (level of consciousness, facial weakness, ataxia, and dysarthria). The 11-item modified NIHSS (mNIHSS) consists of ten items with excellent reliability and one item with good reliability (panel 1).<sup>3</sup> For prehospital assessment of stroke severity, an 8-item and a 5-item NIHSS have undergone preliminary evaluation, and additional studies could prove these shortened scales to be useful in paramedics' prehospital screening for suspected stroke.<sup>17</sup> The NIHSS and its derivatives are designed purely as observational scales and measurement by self-report or telephone is highly unlikely to be possible. However, measurement by video telemedicine seems reliable and could offer a method for remote assessment.<sup>18</sup>

One measure of validity of the NIHSS scale is its correlation with infarct volumes (concurrent validity). This measure has been reported in several studies, using both CT and MRI, yielding correlation coefficients of 0.4–0.8, which suggests a high degree of validity.<sup>10,19–22</sup> Not surprisingly, however, additional factors contribute to the association between deficit and infarct volume, including age and stroke location, etc.

The clinical predictive validity of the NIHSS has been shown in several investigations.<sup>7,12–14</sup> In a post-hoc analysis by stroke subtype of 1268 patients enrolled in an acute stroke trial, baseline NIHSS scores strongly predicted outcome at 7 days and at 3 months. An excellent outcome was achieved by almost two-thirds of patients with a score of three or less at day 7; however, very few patients with baseline scores of more than 15 had excellent outcomes after 3 months.<sup>14</sup> In a retrospective analysis,<sup>12</sup> the mNIHSS showed a high correlation with other scales, with an identical internal structure to the NIHSS (in terms of content validity) and validity for detecting a treatment effect. As shown by retrospective chart review studies, the NIHSS and mNIHSS can also be applied to medical records for analyses with high degrees of reliability and validity.<sup>23–25</sup>

**Role in clinical practice and research**

Neurologists spend several years in training to learn the fine points of the neurological examination. However, the first clinician to assess a patient with stroke, typically in an emergency department, is seldom a neurologist. Non-neurologist physicians, medical students, nurses, and other health practitioners often have difficulty mastering and performing the neurological examination, and as a result, their attempts to perform a neurological examination on a patient with acute stroke can be prolonged and unfocused. The NIHSS offers a more

expeditious approach since it can be effectively implemented by all types of health-care providers, with excellent reliability and validity, after only a few hours of training.<sup>16</sup> Clinicians can then use the scale for initial evaluation, providing quick and accurate assessments of stroke-related deficits, which are easy to communicate with other clinicians, ultimately saving valuable time in triage and treatment of the patient. Video training and certification of NIHSS administration with digital video are available on DVD (digital videodisc) or free online and seem highly effective.<sup>16</sup> Widespread training also offers an advantage in multicentre clinical trial logistics, where standardised use by all investigators is critical both at baseline and for measuring outcomes.<sup>15</sup>

For serial monitoring of patients after stroke who are at risk for neurological worsening, the NIHSS provides a sensitive tool, although many hospitals have traditionally used the less sensitive Glasgow coma scale (GCS) and pupillary examination for this purpose. A change of 2 points or more on the NIHSS was used in trials to suggest a potentially clinically relevant change in a patient's status.<sup>9</sup> Although this specific cutoff has not been independently validated for this purpose, a quantifiable change in neurological examination can be quickly recognised and can prompt further diagnostic studies or treatment. Consequently, many hospitals are training nurses to use the NIHSS in the bedside monitoring of patients with acute stroke.

The predictive value of the scale can also aid in planning a patient's rehabilitation or long-term care needs, even as early as the day of admission.<sup>6</sup> More than 80% of patients whose score is five or less at the time of admission will be discharged home, whereas those with scores between six and 13 usually require acute inpatient rehabilitation, and those with scores of 14 or more frequently need long-term care in nursing facilities.<sup>6,7</sup>

For research purposes, the scale seems to be more sensitive than other indices, such as the mRS or BI, in measuring a simulated treatment effect.<sup>27</sup> In clinical research, this finding provides more power for detection of a difference between interventions, potentially allowing for a smaller sample size. Whether this improved sensitivity to a treatment effect simulated by shifts in admission NIHSS status is shown in trials evaluating treatment interventions remains to be ascertained. At present, the NIHSS is a key measurement tool in studies related to thrombolytic therapy.

#### Pitfalls

Limitations of the NIHSS must be considered to ensure that it is used and interpreted appropriately. In a post-hoc analysis of data from the NINDS rt-PA Stroke Trial, when right-hemisphere and left-hemisphere strokes with equivalent NIHSS scores were compared, the median volume of right-hemisphere strokes was slightly larger than the median volume of left-hemisphere strokes.<sup>28</sup> Furthermore, the scale does not include a

detailed assessment of the cranial nerves, and relatively low scores can occur in patients with disabling infarctions of the brainstem or cerebellum. For example, a cerebellar stroke or Wallenberg syndrome (lateral medulla) can have total NIHSS scores of only 2–4 points, but these strokes can be quite disabling and are occasionally life threatening. Additionally, the NIHSS offers no discriminating value in the identification of the actual cause of the neurological deficit. For accurate diagnosis and optimum treatment of patients presenting with suspected stroke, a complete history, neurological examination, and neuroimaging are ultimately needed to exclude other disorders that can mimic stroke. The NIHSS is a strong predictor of a patient's post-acute care disposition; however, the score is not directly and specifically associated with an individual's ability to functionally compensate for a neurological deficit, and therefore it is not an ideal solitary measure of outcome after stroke.

#### Barthel index

The BI is a scale that measures ten basic aspects of activity related to self-care and mobility (panel 2).<sup>29,30</sup> The normal score is 100, and lower scores indicate greater dependency.<sup>30</sup> Since the introduction of the BI in 1965,<sup>30</sup> it has been widely used in clinical trials to measure outcomes. The BI is determined by observation of patients in a number of tasks, although items regarding bowel and bladder continence are obtained by history.

#### Reliability and validity

There are no formal training or certification programmes to ensure proper use of the BI, so scoring must be done in accord with the scale exactly as written. The internal consistency of the BI has been reported to be extraordinarily high (as indicated by Cronbach's  $\alpha$  of 0.98),<sup>31</sup> which could suggest that some of the items might be redundant or unnecessary, especially items for bathing and personal hygiene. Intraobserver and interobserver reliabilities are also quite high, with Pearson  $r$  scores ranging from 0.89 to 0.99.<sup>31</sup> Ideally, patients should be observed in all of the items of the BI, apart from those related to bowel and bladder function, but telephone assessments of the BI using structured interviews have shown similarly high correlations with direct measurements.<sup>32</sup> When disability was dichotomised by a score of 60, the reliability of the telephone interview was very good ( $\kappa=0.68$ ), and validity relative to direct measurement yielded sensitivity and specificity of 63% and 98%, respectively. Furthermore, there were no significant differences in ratings by patient versus proxy, nor when the interview was done by trained lay persons or health-care professionals, irrespective of the mode of interview.<sup>32</sup> However, purely self-reported scores tend to be less accurate than direct measurements, especially in patients with cognitive dysfunction, serious illness, and age older than 75 years.<sup>33</sup>

For online training for the NIHSS see <http://asa.trainingcampus.net/>

**Panel 2: Barthel index****Bowels**

0=Incontinent (or needs to be given enema)  
 5=Occasional accident (once/week)  
 10=Continent

**Bladder**

0=Incontinent, or catheterised and unable to manage  
 5=Occasional accident (max once per 24 h)  
 10=Continent (for more than 7 days)

**Grooming**

0=Needs help with personal care  
 5=Independent face/hair/teeth/shaving (implements provided)

**Toilet use**

0=Dependent  
 5=Needs some help, but can do something alone  
 10=Independent (on and off, dressing, wiping)

**Feeding**

0=Unable  
 5=Needs help cutting, spreading butter, etc  
 10=Independent (food provided in reach)

**Transfer**

0=Unable, no sitting balance  
 5=Major help (one or two people, physical), can sit  
 10=Minor help (verbal or physical)  
 15=Independent

**Mobility**

0=Immobile  
 5=Wheelchair independent, including corners, etc.  
 10=Walks with help of one person (verbal or physical)  
 15=Independent (but may use any aid—eg, stick)

**Dressing**

0=Dependent  
 5=Needs help, but can do about half unaided  
 10=Independent (including buttons, zips, laces, etc)

**Stairs**

0=Unable  
 5=Needs help (verbal, physical, carrying aid)  
 10=Independent up and down

**Bathing**

0=Dependent  
 5=Independent (or in shower)

**Total (0–100)**

Reproduced from International Disability Studies, by permission of Taylor and Francis.<sup>29</sup>

The BI has moderate concurrent validity with respect to infarct volumes (correlation coefficients of 0.3–0.5).<sup>19,20,34</sup> Clinically, the BI has proven value in discriminating between groups of patients (construct

validity) and predicting outcome (predictive validity).<sup>35</sup> Of 117 patients with scores of 0–40, 70% had died or were living in long-term care facilities 6 months post-stroke. By contrast, 94% of 206 patients with scores of 81–100 were living in the community 6 months post-stroke. Additionally, patients with stroke who had scores greater than 60 after rehabilitation were more likely to be active in their homes and communities, have more social interaction, and be more satisfied with life in general than those with scores of 60 or less.<sup>35</sup>

**Role in clinical practice and research**

In patient care, the BI can be used repeatedly to assess improvement in patients over time, as was its original purpose, and can therefore be used to establish the effectiveness of rehabilitative therapies.<sup>30,36</sup> The tool is relatively easy to administer, and many investigators are familiar with it, which are advantages in conducting multicentre studies. The high reliability, even with telephone assessments, makes this tool potentially useful when study patients are unable to return for direct follow-up assessments.

**Pitfalls**

The BI measures several key activities of daily living and specific physiological deficits, but many aspects of functional independence are not included, such as cognition, language, visual function, emotional impairment, and pain. These items are explicitly excluded from the BI, although each of these components could have a significant effect on independence. For example, a patient with severe aphasia might be completely normal in all items of the BI, yet be unable to function outside of the home without assistance of another person or to even call for help if needed. Thus, the scale suffers from a “ceiling effect”, wherein the maximum score can be achieved even in many disabled patients, and therefore does not differentiate disability well among patients with higher levels of functioning.<sup>4</sup>

In a stroke rehabilitation unit, Dromerick and colleagues<sup>4</sup> studied the sensitivity to change in disability of four stroke scales: the mRS; BI; international stroke trial measure; and functional independence measure. A ceiling effect was evident for the BI, with 27% achieving a total score of 95 or 100, but none of these patients obtained the highest rating on any of the other three measures. These findings confirm that higher scores on the BI might not be relevant to overall functional ability.

In the setting of an acute stroke, the BI is not especially helpful as it is also highly susceptible to a “floor effect”. Most patients, even those with a minor stroke, are bed-bound in the first few hours after stroke, either by their deficit or by medical directive,<sup>37,38</sup> and therefore will initially receive low scores. Consequently, the BI cannot be used to measure initial stroke severity or, by extension, to stratify patients by severity in acute stroke trials.

### Modified Rankin scale

The Rankin scale was devised in 1957 for assessment of stroke outcomes, and was modified in 1988 to improve its comprehensiveness.<sup>39–41</sup> The modified version, or mRS, has since been commonly used to assess disability after a stroke. The mRS attempts to measure functional independence, incorporating the WHO components of body function, activity, and participation. The scale is defined categorically with seven different grades:<sup>4</sup> 0 indicates no symptoms, 5 indicates severe disability, and 6 indicates death (panel 3).<sup>40</sup> A 1-point shift on this scale is often deemed clinically significant because of the large category sizes. Patients may use adaptive devices and still be considered independent, but the need for supervision or even minimum aid from another person is scored as dependent.

#### Reliability and validity

The validity in stroke outcome and inter-rater reliability have been well documented for the mRS. However, the simplicity of the mRS as a 6-point scale can affect its reliability because rating scales with more items or rankings generally offer higher reliability.<sup>42</sup> Therefore, strict adherence to a series of rules is needed to ensure accuracy. A structured interview has been proposed to improve its inter-rater reliability. Wilson and colleagues<sup>43</sup> showed that with a structured interview, the unweighted  $\kappa$  was 0.74, with agreement in 81% of cases. The unweighted  $\kappa$  for inter-rater agreement without the structured interview was only 0.25. Additional methods to improve the reliability of the mRS continue to be developed.<sup>44</sup> Although the structured interview seems highly reliable when assessing a patient in person, attempts to measure the mRS by telephone with the structured interview have provided much lower reliability ( $\kappa=0.30$ ) and cannot be recommended.<sup>45</sup> Recently, a DVD-based training system was used in the Stroke–Acute Ischemic NXY Treatment (SAINT I) clinical trial of NXY-059 to ensure correct implementation of the mRS by the nearly 200 sites in the trial.<sup>46</sup>

The mRS has moderate concurrent validity with respect to infarct volumes (correlation coefficients of 0.4–0.5),<sup>20,34,47</sup> which is similar to findings for the BI and other scales. Construct validity of the mRS has been shown to have excellent agreement with other rating scales, although it most closely accords with the GOS (described below).<sup>9</sup>

A prospective study compared five outcome measures in 1530 patients 100 days after ischaemic stroke. The mRS was more responsive to changes in functional status and was a better instrument for differentiating between changes in mild-to-moderate disability, especially after mild stroke, than was the BI, probably because there is less of a ceiling effect. The mRS was also more reflective of disability in an emotional context because it is more susceptible to change by depression, and strongly correlated with the Center for Epidemiologic Studies-

### Panel 3: Modified Rankin scale

#### 0=no symptoms

#### 1=no significant disability, despite symptoms

Able to perform all usual duties and activities

#### 2=slight disability

Unable to perform all previous activities but able to look after own affairs without assistance

#### 3=moderate disability

Requires some help, but able to walk without assistance

#### 4=moderately severe disability

Unable to walk without assistance and unable to attend to own bodily needs without assistance

#### 5=severe disability

Bedridden, incontinent, and requires constant nursing care and attention

#### 6=dead

Reproduced from *International Disability Studies*, by permission of Taylor and Francis.<sup>40</sup>

Depression (CES-D) short form.<sup>48</sup> However, overall responsiveness of the mRS is poor during short-term intervals, such as from admission to discharge,<sup>4</sup> at least in part because a substantial clinical threshold exists between each point in the scale and because patients have not resumed their usual roles and activities while in hospital.

#### Role in clinical practice and research

The mRS offers an easy and rapid assessment in clinical practice of the effect of a patient's stroke on their activities and participation in a social context. Although a typical office visit after stroke might focus on residual neurological deficits, the use of probing questions in the structured mRS interview could identify broader issues that remain problematic for patients yet may be amenable to further intervention.

In clinical trials, selection of the specific primary endpoint is critical to detecting differences between interventions. Formal statistical analyses often dichotomise the mRS outcomes at an arbitrary threshold as either good or bad.<sup>49</sup> Young and colleagues<sup>50</sup> used data from the Glycine Antagonist in Neuroprotection (GAIN) international trial to explore patterns and extent of treatment effect and to estimate statistical power for a range of endpoints with the mRS or BI. The mRS endpoints generally needed substantially smaller sample sizes to achieve adequate statistical power, and the odds of achieving a result that was statistically significant increased by 89% with an mRS endpoint compared with a BI endpoint. However, since the mRS is an ordered outcome variable and each discrete grade indicates clinically relevant levels of functional status, changes of

one or more grade in either direction can indicate meaningful improvements or declines in ability over time.<sup>48</sup> Therefore, analysis across the distribution of scores rather than dichotomisation (eg, as used in the SAINT I trial<sup>46</sup> and Field Administration of Stroke Therapy–Magnesium [FAST MAG] phase III trial<sup>51</sup>), seems reasonable and probably maximises the use of this rating scale.<sup>52,53</sup> This approach offers more statistical power for detection of clinically meaningful differences in treatment effects.<sup>52</sup> Moreover, it perhaps provides a more realistic target for stroke studies, wherein patients with initially mild strokes might be expected to recover to full independence (mRS 0 or 1), whereas those with severe strokes might be expected to be ambulatory at best (mRS 3). The SAINT I trial was the first large randomised clinical trial for stroke that used the full distribution of mRS scores as its primary endpoint, and was successful in showing a significant benefit of treatment.<sup>46</sup>

#### Pitfalls

The mRS is a broad-based summary measure of impairment, activity, and participation, but it lacks specificity. Domains such as cognition, language, visual function, emotional impairment, and pain are not directly measured, nor are other sources of disability, such as a hip fracture, but these factors are implicitly included in the mRS score. This is often regarded as a disadvantage since a relatively small stroke or mild neurological deficit can lead to severe disability in some instances (such as a visual field deficit causing a truck driver to become unemployed or post-stroke depression hampering normal activities), whereas a relatively large stroke can result in mild disability in others (such as a cerebellar stroke in a sedentary person), especially when emotional and motivational factors affect the patient's recovery and perception of disability. However, these short-term and long-term complications, plus others

#### Panel 4: Glasgow outcome scale

##### 1=good recovery

Resumption of normal activities even though there may be minor neurological or psychological deficits

##### 2=moderate disability

(Disabled but independent). Patient is independent as far as daily life is concerned. The disabilities found include varying degrees of dysphasia, hemiparesis, or ataxia, as well as intellectual and memory deficits and personality changes

##### 3=severe disability

Conscious but disabled. Patient depends on others for daily support due to mental or physical disability or both

##### 4=persistent vegetative state

Patient exhibits no obvious cortical function

##### 5=death

Reproduced from *The Lancet*, by permission of Elsevier.<sup>55</sup>

such as depression, dementia, or falls, could be useful as references or benchmarks for overall success of care. Therefore, this measure might be especially meaningful to clinicians and patients when considering the potential overall outcomes of an acute intervention.<sup>54</sup> In the acute setting, the mRS can have a floor effect,<sup>4</sup> with scores of four or five irrespective of stroke severity because patients are inherently bed-bound in the first few hours after stroke.<sup>37,38</sup>

Scores on the mRS are often dichotomised in clinical trials, but this approach is insensitive to partial but meaningful improvements (eg, from five to three), and decreases statistical sensitivity for detection of differences between interventions, as outlined above.<sup>49,52,53</sup> In 459 patients prospectively assessed after stroke, dichotomised mRS analysis precluded detection of recovery that was partial or did not lead to complete functional independence, although ordinal analysis incorporating all scores showed improvement in the same patient population.<sup>53</sup> Capturing data indicating all levels of functional status after stroke is important, especially in clinical trials because full recovery may be unattainable for many patients.

#### Glasgow outcome scale

The GOS is another ordered scale used to assess outcomes after acute brain injury. The scale allocates patients into broad outcome categories (panel 4),<sup>55</sup> including: (1) good recovery; (2) moderate disability; (3) severe disability; (4) persistent vegetative state; and (5) death.<sup>55</sup> A key difference of the GOS versus the mRS is the lack of distinction among patients with good outcomes, since this group encompasses full recovery and mild disability.

#### Reliability and validity

As with the mRS, the reliability of the GOS is better when administered with a structured interview.<sup>49</sup> When the GOS was administered to 50 patients with head injury through a structured interview, the scores assigned by two different raters had an overall agreement of 92%. The GOS correlates with the mRS in most patients with stroke (concurrent validity).<sup>9</sup>

#### Role in clinical practice and research

The GOS is simple to administer in the patient care setting with a structured interview. Because the scale has been widely used to monitor neurological changes in patients with head injury and stroke, many clinicians are familiar with it. The broad categories defined by the GOS provide differences in disability that are clinically meaningful. However, unlike the mRS, the effects of other factors that relate to social role and function are not well defined.

#### Pitfalls

The lack of comprehensiveness of the GOS limits its usefulness in stroke trials and practice.<sup>56</sup> Significant

variability and systematic bias among raters can be expected from administration of the GOS outside the context of a structured interview and without written protocols, which is typical in routine practice.<sup>57,58</sup> The categories of severe disability, moderate disability, and good recovery are multidimensional without specific definitions or parameters, and too ambiguous for formal data analysis.<sup>56,59</sup> Additionally, the scoring of the GOS emphasises physical disabilities without context for cognitive or emotional problems.<sup>57</sup>

### Stroke impact scale

The SIS was developed from the perspectives of patients and caregivers rather than investigators to specifically measure changes in emotion, communication, memory and thinking, and social role, primarily in mild-to-moderate strokes.<sup>5</sup> These dimensions are not directly assessed in the measures described above. The SIS is a work in progress: version 2.0 includes 64 items in eight domains, version 3.0 contains 59 items also in eight domains, and further refinements are underway.<sup>4,60,61</sup> 16 items from four of the eight domains of SIS 3.0 have been combined to produce a short composite physical domain score, known as the SIS-16 (panel 5).<sup>60</sup>

### Reliability and validity

The SIS is internally consistent, as measured by Cronbach  $\alpha$  coefficients of 0.83 to 0.9. Intrarater reliability is high, with ICCs ranging from 0.7 to 0.92 for the eight domains, although the emotional domain was less reliable (ICC 0.57, version 2.0).<sup>5</sup> A potential for a floor effect was noted for hand function with moderate stroke and for a ceiling effect in the communication domain with all strokes. Concurrent validity with respect to general measures of health, such as the 36-item short form health survey (SF-36), was good to excellent for measures of disability, memory, communication, and social function, but relatively low for emotional and physical role function. Predictive validity for global recovery was significantly related to SIS domain scores for physical function ( $p=0.0001$ ) and emotion ( $p=0.0002$ ), but did not reach statistical significance for participation ( $p=0.058$ ).<sup>5</sup>

The SIS-16 was developed to assess levels of disability 1–3 months post-stroke. Although quite similar to the BI, the SIS-16 showed increased sensitivity for detection of different levels of disability, with less ceiling effect at 1 month (5% vs 28% for BI) and 3 months (8% vs 38% for BI) post-stroke.<sup>60</sup> In telephone assessments, the SIS was effective at differentiating between groups of stroke patients with different disability levels in a community setting.<sup>62</sup>

### Role in clinical practice and research

The SIS offers the possibility of measuring a number of dimensions of health-related quality of life that are not specifically addressed in other scales. However, its role in

multicentre trials and clinical practice has yet to be established.

### Pitfalls

A major drawback of the SIS is the need for self-reporting or the use of a proxy.<sup>63</sup> This requirement substantially limits its use in aphasic patients and in those with denial of their deficit or illness. Proxy responses have been reported to differ significantly from those of patients, tending to overstate the severity of the patients' conditions.<sup>63</sup> Although the SIS has been studied intensively by its developers, it has not been applied broadly to other centres and trials, so its generalisability remains in question. However, other sites are beginning to study this tool.<sup>64</sup> Currently, formalised training and certification are not available, but this will probably become necessary if this scale is to become incorporated into clinical trials and practice.

### Other stroke scales

Several other scales have been developed to measure outcomes after stroke. In addition to the NIHSS, other neurological impairment scales in clinical trials are the Scandinavian stroke scale, the Orgogozo neurological scale, and the Canadian neurological scale.<sup>2</sup> The modified Mathew scale has also been used in stroke outcome trials, although it has poor inter-rater reliability and questionable validity in patients with stroke.<sup>65,66</sup> Other than the mRS and the BI, only a few disability scales have established validity and reliability in stroke, including the Nottingham activities of daily living scale,

#### Panel 5: Stroke impact scale-16 questionnaire

In the past 2 weeks, how difficult was it to:

- a. Dress the top part of your body?
- b. Bathe yourself?
- c. Get to the toilet on time?
- d. Control your bladder (not have an accident)?
- e. Control your bowels (not have an accident)?
- f. Stand without losing balance?
- g. Go shopping?
- h. Do heavy household chores (eg, vacuum, laundry, or garden work)?
- i. Stay sitting without losing your balance?
- j. Walk without losing your balance?
- k. Move from a bed to a chair?
- l. Walk fast?
- m. Climb one flight of stairs?
- n. Walk one block?
- o. Get in and out of a car?
- p. Carry heavy objects (eg, bag of groceries) with your affected hand?

Scoring: 1=could not do at all; 2=very difficult; 3=somewhat difficult; 4=a little difficult; 5=not difficult at all. Reproduced from *Neurology*, by permission of Lippincott Williams and Wilkins.<sup>60</sup>

the Adams disability scale, and the functional independence measure. These are seldom used in current stroke trials because of their complexity or lack of familiarity. The SF-36 is a general measure of quality of life that has been validated for use in patients with many disorders, including stroke.<sup>2</sup>

### Global statistical tests

Because the existing stroke outcome scales all measure different but related aspects of disability after stroke, a single scale does not seem sufficient to describe the spectrum of outcomes from stroke interventions. One approach to a more unified assessment is the integration of multiple scales to generate a global outcome statistic. Global tests are useful when the outcome is difficult to measure and a combination of correlated outcomes (each measuring recovery from stroke) would be informative. The global statistical test allows for categorisation of each outcome as favourable or unfavourable and essentially determines the direction and size of a treatment effect across multiple rating scales simultaneously, thereby improving statistical power and providing compelling evidence of overall treatment efficacy.<sup>67</sup>

A global statistical test was devised for use in the NINDS rt-PA Stroke Trial, which allowed for overall assessment of treatment efficacy for a combination of the NIHSS, mRS, BI, and GOS.<sup>9</sup> This study showed that intravenous thrombolysis increased the odds of a favourable global outcome by a factor of 1.9. Although this was the primary outcome of the study, this global result may be somewhat difficult for clinicians to conceptualise and apply in practice, at least in part because they tend to rely on information such as the proportion of patients who improve or return to normal after treatment. Consequently, in this particular study, the individual component scores yielded comparable results and have been more frequently communicated in subsequent reports, even though they were secondary endpoints.<sup>68,69</sup>

This approach previously required dichotomisation of each scale and therefore might have been less sensitive to incomplete but clinically relevant treatment effects. However, new methods to perform global tests using the distributions of each component score might provide even greater power for detection of clinically meaningful effects in future studies.

### Conclusions

Understanding the use of stroke scales is important for assessment of patients with stroke in both the acute and recovery phases, evaluation of published research, and selection appropriate outcome measures for intervention trials. Stroke rating scales used in clinical trials should have proven reliability and be validated for use in stroke. No single outcome measure can describe or predict all dimensions of recovery and disability after acute stroke,

### Search strategy and selection criteria

References for this review were identified by searches of MEDLINE between 1969 and April, 2006, and from references from relevant articles. Several articles were also identified through searches of the extensive files of the authors. Two groups of search terms were combined: "stroke assessment" OR "stroke outcome" OR "stroke measurement" AND the terms "National Institutes of Health Stroke Scale" OR "NIHSS" OR "Barthel Index" OR "Rankin" OR "Glasgow Outcome" OR "Stroke Impact Scale". Abstracts and reports from meetings between 2003 and 2006 were also included. Only papers published in English were reviewed. The final reference list was generated on the basis of originality and relevance to the topics covered in the review.

and each scale has a potential role in patient care and research. A composite measure, such as a global statistic derived from the scores of several scales, seems useful in measuring the multiple dimensions of outcomes after stroke. Ongoing attempts are being made to incorporate patients' perspectives, since these ultimately are critical measures of success or failure.

### Conflicts of interest

SEK has no proprietary interests in any of the stroke scales described in this paper. He has been an investigator for many stroke trials that have involved the rating scales described in this paper, including the ongoing SAINT II trial. This is the confirmatory study that follows the SAINT I trial discussed herein, and both of these trials are sponsored by AstraZeneca. SEK has also been a consultant on issues related to the design or interpretation of stroke trials for Ono Pharmaceuticals, Merck, NovoNordisk, and AstraZeneca.

### References

- 1 World Health Organization. Towards a common language for functioning, disability, and health. Geneva: WHO; 2002. <http://www.who.int/classifications/icf/en/> (accessed April 18, 2006).
- 2 Duncan PW, Jorgensen HS, Wade DT. Outcome measures in acute stroke trials: a systematic review and some recommendations to improve practice. *Stroke* 2000; **31**: 1429–38.
- 3 Meyer BC, Hemmen TM, Jackson CM, Lyden PD. Modified National Institutes of Health stroke scale for use in stroke clinical trials: prospective reliability and validity. *Stroke* 2002; **33**: 1261–66.
- 4 Dromerick AW, Edwards DF, Diringner MN. Sensitivity to changes in disability after stroke: a comparison of four scales useful in clinical trials. *J Rehabil Res Dev* 2003; **40**: 1–8.
- 5 Duncan PW, Wallace D, Lai SM, Johnson D, Embretson S, Laster LJ. The stroke impact scale version 2.0: evaluation of reliability, validity, and sensitivity to change. *Stroke* 1999; **30**: 2131–40.
- 6 Schlegel D, Kolb SJ, Luciano JM, et al. Utility of the NIH Stroke Scale as a predictor of hospital disposition. *Stroke* 2003; **34**: 134–37.
- 7 Rundek T, Mast H, Hartmann A, et al. Predictors of resource use after acute hospitalization: the Northern Manhattan Stroke Study. *Neurology* 2000; **55**: 1180–87.
- 8 Appellos P, Terént A. Characteristics of the National Institute of Health stroke scale: results from a population-based stroke cohort at baseline and after one year. *Cerebrovasc Dis* 2004; **17**: 21–27.
- 9 Tilley BC, Marler J, Geller NL, National Institute of Neurological Disorders and Stroke (NINDS) rt-PA Stroke Trial Study. Use of a global test for multiple outcomes in stroke trials with application to the National Institute of Neurological Disorders and Stroke t-PA Stroke Trial. *Stroke* 1996; **27**: 2136–42.
- 10 Brott TG, Adams HP Jr, Olinger CP, et al. Measurements of acute cerebral infarction: a clinical examination scale. *Stroke* 1989; **20**: 864–70.

- 11 Lyden P, Lu M, Jackson C, NINDS tPA Stroke Trial Investigators. Underlying structure of the National Institutes of Health Stroke Scale: results of a factor analysis. *Stroke* 1999; **30**: 2347–54.
- 12 Lyden PD, Lu M, Levine SR, Brott TG, Broderick J, NINDS rtPA Stroke Study Group. A modified National Institutes of Health Stroke Scale for use in stroke clinical trials: preliminary reliability and validity. *Stroke* 2001; **32**: 1310–17.
- 13 Schlegel DJ, Tanne D, Demchuk AM, Levine SR, Kasner SE, Multicenter rt-PA Stroke Survey Group. Prediction of hospital disposition after thrombolysis for acute ischemic stroke using the National Institutes of Health Stroke Scale. *Arch Neurol* 2004; **61**: 1061–64.
- 14 Adams HP Jr, Davis PH, Leira EC, et al. 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**: 126–31.
- 15 Lyden P, Brott T, Tilley B, NINDS TPA Stroke Study Group. Improved reliability of the NIH Stroke Scale using video training. *Stroke* 1994; **25**: 2220–26.
- 16 Goldstein LB, Samsa GP. Reliability of the National Institutes of Health stroke scale: extension to non-neurologists in the context of a clinical trial. *Stroke* 1997; **28**: 307–10.
- 17 Tirschwell DL, Longstreth WT Jr, Becker KJ, et al. Shortening the NIH Stroke Scale for use in the prehospital setting. *Stroke* 2002; **33**: 2801–06.
- 18 Meyer BC, Lyden PD, Al-Khoury L, et al. Prospective reliability of the STRoKE DOC wireless/site independent telemedicine system. *Neurology* 2005; **64**: 1058–60.
- 19 Saver JL, Johnston KC, Homer D, et al. Infarct volume as a surrogate or auxiliary outcome measure in ischemic stroke clinical trials. *Stroke* 1999; **30**: 293–98.
- 20 Schiemanck SK, Post MWM, Witkamp TD, Kappelle LJ, Prevo AJH. Relationship between ischemic lesion volume and functional status in the 2nd week after middle cerebral artery stroke. *Neurorehabil Neural Repair* 2005; **19**: 133–38.
- 21 Lyden P, Claesson L, Havstad S, Ashwood T, Lu M. Factor analysis of the National Institutes of Health Stroke Scale in patients with large strokes. *Arch Neurol* 2004; **61**: 1677–80.
- 22 Derex L, Nighoghossian N, Hermier M, et al. Influence of pretreatment MRI parameters on clinical outcome, recanalization and infarct size in 49 stroke patients treated by intravenous tissue plasminogen activator. *J Neurol Sci* 2004; **225**: 3–9.
- 23 Kasner SE, Chalela JA, Luciano JM, et al. Reliability and validity of estimating the NIH Stroke Scale score from medical records. *Stroke* 1999; **30**: 1534–37.
- 24 Kasner SE, Cucchiara BL, McGarvey ML, Luciano JM, Liebeskind DS, Chalela JA. Modified National Institutes of Health Stroke Scale can be estimated from medical records. *Stroke* 2003; **34**: 568–70.
- 25 Bushnell CD, Johnston DCC, Goldstein LB. Retrospective assessment of initial stroke severity: comparison of the NIH Stroke Scale and the Canadian Neurological Scale. *Stroke* 2001; **32**: 656–60.
- 26 Lyden P, Raman R, Liu L, et al. NIHSS training and certification using a new digital video disk is reliable. *Stroke* 2005; **36**: 2446–49.
- 27 Young FB, Weir CJ, Lees KR, GAIN International Trial Steering Committee and Investigators. Comparison of the National Institutes of Health Stroke Scale with disability outcome measures in acute stroke trials. *Stroke* 2005; **36**: 2187–92.
- 28 Woo D, Broderick JP, Kothari RU, NINDS t-PA Stroke Study Group. Does the National Institutes of Health Stroke Scale favor left hemisphere strokes? *Stroke* 1999; **30**: 2355–59.
- 29 Collin C, Wade DT, Davies S, Horne V. The Barthel ADL Index: a reliability study. *Int Disabil Stud* 1988; **10**: 61–63.
- 30 Mahoney FI, Barthel DW. Functional evaluation: the Barthel Index. *Md State Med J* 1965; **14**: 61–65.
- 31 Shinar D, Gross CR, Bronstein KS, et al. Reliability of the Activities of Daily Living scale and its use in telephone interview. *Arch Phys Med Rehabil* 1987; **68**: 723–28.
- 32 Korner-Bitensky N, Wood-Dauphinee S. Barthel Index information elicited over the telephone: is it reliable? *Am J Phys Med Rehabil* 1995; **74**: 9–18.
- 33 Sinoff G, Ore L. The Barthel activities of daily living index: self-reporting versus actual performance in the old-old ( $\geq 75$  years). *J Am Geriatr Soc* 1997; **45**: 832–36.
- 34 Schiemanck SK, Post MWM, Kwakkel G, Witkamp TD, Kappelle LJ, Prevo AJH. Ischemic lesion volume correlates with long-term functional outcome and quality of life of middle cerebral artery stroke survivors. *Restor Neurol Neurosci* 2005; **23**: 257–63.
- 35 Granger CV, Hamilton BB, Gresham GE. The stroke rehabilitation outcome study—Part I: general description. *Arch Phys Med Rehabil* 1988; **69**: 506–09.
- 36 Granger CV, Dewis LS, Peters NC, Sherwood CC, Barrett JE. Stroke rehabilitation: analysis of repeated Barthel index measures. *Arch Phys Med Rehabil* 1979; **60**: 14–17.
- 37 Adams HP, Jr, Adams RJ, Brott T, et al. Guidelines for the early management of patients with ischemic stroke: a scientific statement from the stroke council of the American Stroke Association. *Stroke* 2003; **34**: 1056–83.
- 38 Wojner-Alexander AW, Garami Z, Chernyshev OY, Alexandrov AV. Heads down: flat positioning improves blood flow velocity in acute ischemic stroke. *Neurology* 2005; **64**: 1354–57.
- 39 New PW, Buchbinder R. Critical appraisal and review of the Rankin Scale and its derivatives. *Neuroepidemiology* 2006; **26**: 4–15.
- 40 van Swieten JC, Koudstaal PJ, Visser MC, Schouten HJA, van Gijn J. Interobserver agreement for the assessment of handicap in stroke patients. *Stroke* 1988; **19**: 604–07.
- 41 Sulter G, Steen C, De Keyser J. Use of the Barthel index and modified Rankin scale in acute stroke trials. *Stroke* 1999; **30**: 1538–41.
- 42 Streiner DL, Norman GR. Health measurement scales. In: A practical guide to their development and use, 2nd edn. Oxford: Oxford University Press; 1995: 111–12.
- 43 Wilson JTL, Hareendran A, Hendry A, Potter J, Bone I, Muir KW. Reliability of the modified Rankin scale across multiple raters: benefits of a structured interview. *Stroke* 2005; **36**: 777–81.
- 44 Shinohara Y, Minematsu K, Amano T, Ohashi Y. Modified Rankin scale with expanded guidance scheme and interview questionnaire: interrater agreement and reproducibility of assessment. *Cerebrovasc Dis* 2006; **21**: 271–78.
- 45 Newcommon NJ, Green TL, Haley E, et al. Improving the assessment of outcomes in stroke: use of a structured interview to assign grades on the modified Rankin scale. *Stroke* 2003; **34**: 377–78.
- 46 Lees KR, Zivin JA, Ashwood T, et al. NXY-059 for acute ischemic stroke. *N Engl J Med* 2006; **354**: 588–600.
- 47 Lev MH, Segal AZ, Farkas J, et al. Utility of perfusion-weighted CT imaging in acute middle cerebral artery stroke treated with intra-arterial thrombolysis: prediction of final infarct volume and clinical outcome. *Stroke* 2001; **32**: 2021–28.
- 48 Weimar C, Kurth T, Kraywinkel K, German Stroke Data Bank Collaborators. Assessment of functioning and disability after ischemic stroke. *Stroke* 2002; **33**: 2053–59.
- 49 Murray GD, Barer D, Choi S, et al. Design and analysis of phase III trials with ordered outcome scales: the concept of the sliding dichotomy. *J Neurotrauma* 2005; **22**: 511–17.
- 50 Young FB, Lees KR, Weir CJ, Glycine Antagonist in Neuroprotection (GAIN) International Trial Steering Committee and Investigators. Strengthening acute stroke trials through optimal use of disability end points. *Stroke* 2003; **34**: 2676–80.
- 51 Saver JL Jr, Kidwell CS, Hamilton S, et al. The Field Administration of Stroke Therapy—Magnesium (FAST-MAG) Phase 3 clinical trial. Presented at the 29th International Stroke Conference; Feb 5–7, 2004; San Diego, Calif.
- 52 Saver JL. Number needed to treat estimates incorporating effects over the entire range of clinical outcomes: novel derivation method and application to thrombolytic therapy for acute stroke. *Arch Neurol* 2004; **61**: 1066–70.
- 53 Lai S-M, Duncan PW. Stroke recovery profile and the modified Rankin assessment. *Neuroepidemiology* 2001; **20**: 26–30.
- 54 Johnson KC, Li JY, Lyden PD, et al. Medical and neurological complications of ischemic stroke: experience from the RANTTAS Trial. *Stroke* 1998; **29**: 447–53.
- 55 Jennett B, Bond M. Assessment of outcome after severe brain damage: a practical scale. *Lancet* 1975; **1**: 480–84.
- 56 Wilson JTL, Pettigrew LEL, Teasdale GM. Structured interviews for the Glasgow outcome scale and the extended Glasgow outcome scale: guidelines for their use. *J Neurotrauma* 1998; **15**: 573–85.
- 57 Maas AIR, Braakman R, Schouten HJA, Minderhoud JM, van Zomeren AH. Agreement between physicians on assessment of outcome following severe head injury. *J Neurosurg* 1983; **58**: 321–25.

- 58 Anderson SI, Housley AM, Jones PA, Slattery J, Miller JD. Glasgow outcome scale: an inter-rater reliability study. *Brain Inj* 1993; **7**: 309–17.
- 59 Grant I, Alves W. Psychiatric and psychosocial disturbances in head injury. In: Levin HS, Grafman J, Eisenberg HM, eds. *Neurobehavioral recovery from head injury*. New York: Oxford University Press; 2002: 232–61.
- 60 Duncan PW, Lai SM, Bode RK, Perera S, DeRosa J, GAIN Americas Investigators. Stroke impact scale-16: a brief assessment of physical function. *Neurology* 2003; **60**: 291–96.
- 61 Duncan PW, Bode RK, Lai SM, Perera S. Glycine Antagonist in Neuroprotection Americas Investigators. Rasch analysis of a new stroke-specific outcome scale: the stroke impact scale. *Arch Phys Med Rehabil* 2003; **84**: 950–63.
- 62 Kwon S, Duncan P, Studenski S, Perera S, Lai SM, Reker D. Measuring stroke impact with SIS: construct validity of SIS telephone administration. *Qual Life Res* 2006; **15**: 367–76.
- 63 Duncan PW, Lai SM, Tyler D, Perera S, Reker DM, Studenski S. Evaluation of proxy responses to the stroke impact scale. *Stroke* 2002; **33**: 2593–99.
- 64 Edwards B, O'Connell B. Internal consistency and validity of the stroke impact scale 2.0 (SIS 2.0) and SIS-16 in an Australian sample. *Qual Life Res* 2003; **12**: 1127–35.
- 65 Weir NU, Counsell CE, McDowall M, Gunkel A, Dennis MS. Reliability of the variables in a new set of models that predict outcome after stroke. *J Neurol Neurosurg Psychiatry* 2003; **74**: 447–51.
- 66 Gelmers HJ, Gorter K, de Weerd CJ, Wiezer HJA. Assessment of interobserver variability in a Dutch multicenter study on acute ischemic stroke. *Stroke* 1988; **19**: 709–11.
- 67 Pocock SJ, Geller NL, Tsiatis AA. The analysis of multiple endpoints in clinical trials. *Biometrics* 1987; **43**: 487–98.
- 68 Broderick JP, Lu M, Kothari R, et al. Finding the most powerful measures of the effectiveness of tissue plasminogen activator in the NINDS tPA stroke trial. *Stroke* 2000; **31**: 2335–41.
- 69 Frankel MR, Morgenstern LB, Kwiatkowski T, et al. Predicting prognosis after stroke: a placebo group analysis from the National Institute of Neurological Disorders and Stroke rt-PA Stroke Trial. *Neurology* 2000; **55**: 952–59.