Outcomes Validity and Reliability of the Modified Rankin Scale: Implications for Stroke Clinical Trials
A Literature Review and Synthesis

Jamie L. Banks, MSc, PhD; Charles A. Marotta, MD, PhD

Background and Purpose—The modified Rankin scale (mRS), a clinician-reported measure of global disability, is widely applied for evaluating stroke patient outcomes and as an end point in randomized clinical trials. Extensive evidence on the validity of the mRS exists across a large but fragmented literature. As new treatments for acute ischemic stroke are submitted for agency approval, an appreciation of the mRS’s attributes, specifically its relationship to other stroke evaluation scales, would be valuable for decision-makers to properly assess the impact of a new drug on treatment paradigms. The purpose of this report is to assemble and systematically assess the properties of the mRS to provide decision-makers with pertinent evaluative information.

Methods—A Medline search was conducted to identify reports in the peer-reviewed medical literature (1957–2006) that provide information on the structure, validation, scoring, and psychometric properties of the mRS and its use in clinical trials. The selection of articles was based on defined criteria that included relevance, study design and use of appropriate statistical methods.

Results—Of 224 articles identified by the literature search, 50 were selected for detailed assessment. Inter-rater reliability with the mRS is moderate and improves with structured interviews ($\kappa$ 0.56 versus 0.78); strong test-re-test reliability ($\kappa$ 0.81 to 0.95) has been reported. Numerous studies demonstrate the construct validity of the mRS by its relationships to physiological indicators such as stroke type, lesion size, perfusion and neurological impairment. Convergent validity between the mRS and other disability scales is well documented. Patient comorbidities and socioeconomic factors should be considered in properly applying and interpreting the mRS. Recent analyses suggest that randomized clinical trials of acute stroke treatments may require a smaller sample size if the mRS is used as a primary end point rather than the Barthel Index.

Conclusions—Multiple types of evidence attest to the validity and reliability of the mRS. The reported data support the view that the mRS is a valuable instrument for assessing the impact of new stroke treatments. (Stroke. 2007;38:1091-1096.)

Key Words: cerebrovascular accident ■ disability evaluation ■ randomized controlled trials ■ rankin scale ■ reproducibility of results
disability evaluation, statistics and randomized controlled trials. Other search terms including lesion volume, stroke type, end point and design, were used to search for additional articles. Two reviewers selected articles for assessment based on relevance to the study objectives and (1) the use of appropriate statistical methods in assessing validity, and (2) analysis of the mRS as a primary end point in RCTs of acute ischemic stroke treatments. The initial search identified 224 articles of which 50 were selected for a detailed assessment (40 relating mainly to validity/reliability, 2 to the original Rankin and related scales, 8 related to RCTs). Additional articles providing information on the limitations of the mRS, and others were identified from bibliographies of the retrieved articles.

## Results

### Historical Development

The history of the mRS dates from publication of the original Rankin Scale in 1957. Developed by Dr John Rankin in Glasgow, Scotland, the RS comprises 5 grades of stroke severity ranging from “no significant disability” to “severe disability” (Table 1), and was intended as a descriptive categorization of functional recovery of cerebrovascular disease patients (>60 years) at the time of discharge or transfer from the index hospitalization. The mRS was published in 1988 and consists of 6 categories (grades 0 to 5) rather than 5 for the RS; an additional category, grade “6” denoting death, is usually incorporated into the mRS for RCT purposes (Table 1). In the mRS: grade 1 of the original RS (“no significant disability”) is replaced by 2 grades, 0 and 1, with grade 0 describing patients without symptoms and grade 1 describing patients without significant disability “despite symptoms.” This finer discrimination of mild strokes increases the usefulness of the mRS in evaluating RCTs of acute stroke interventions. Additionally, grade 2 in the mRS (“unable to perform all previous activities”) is more definitive compared with that grade of the RS (“unable to carry out some of previous activities”).

The mRS is heavily weighted toward global disability (in particular, physical disability) and the need for assistance. As a global disability measure, the broad categories of the mRS (Table 1) subsume instrumental activities of daily living (IADL; eg, meal preparation, shopping, handling money) and basic ADLs (BADL; eg, walking, dressing, grooming) with emphasis on compromised motor function. The global nature of the mRS thereby allows the clinician to consider nonphysical attributes essential to a person’s self-maintenance and well-being, such as cognition and language, social functioning, and poststroke mood disturbances, particularly depression, that may contribute to perceived disability. This allowance for consideration of IADLs and other nonphysical characteristics distinguishes the mRS from BADL-specific measures, such as the Barthel Index (BI).

### Test-Retest Reliability

Reliability refers to the extent to which a scale consistently and reproducibly measures the attributes it was intended to measure. Test-retest reliability evaluates the consistency of results over time in the absence of changes in the subject population and the raters. The \( \kappa \) statistic indicates the extent of agreement among different sets of results not occurring by chance; a weighted \( \kappa \) adjusts for the extent of disagreement, eg, differences of 1 grade versus 2 grades of the scale.

Strong test-re-test reliability of the mRS was reported in 2 independent studies (Table 2). In 1 investigation, 2 raters each graded 48 patients on 2 separate occasions with excellent consistency.

### Inter-Rater Reliability

Inter-rater reliability evaluates the consistency of results among raters. Because mRS categories are broad and the assessments are subjective, variability may occur across clinical raters. In 3 separate studies, inter-rater reliability of the mRS ranged from moderate to nearly perfect, indicated by the weighted \( \kappa \) (Table 2). In addition to those studies, strong inter-rater reliability has also been reported for a German version of the mRS \( \kappa = 0.76 \). Structured interviews have been shown to improve inter-rater reliability of the mRS. When used in a RCT setting,

--

\*The mRS was not developed for the acute assessment of stroke. For example, the scale would have limited value if it were used in the acute setting as illustrated by assigning a patient with a mild stroke mRS grade 4 or 5 on the basis of the need for nursing care and supervision.

\*\*The Barthel Index ADL items are: feeding, bathing, grooming, dressing, bowels, bladder, toilet use, transfers, mobility, stairs.

\*\*Strength of agreement for the \( \kappa \) statistic has been categorized as follows: <0=poor; 0–0.20=slight; 0.21–0.40=fair; 0.41–0.60=moderate; 0.61–0.80=substantial; 0.81–1.00=almost perfect.

### Table 1. Original and Modified RSs

<table>
<thead>
<tr>
<th>Grade</th>
<th>Original RS</th>
<th>mRS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>NA</td>
<td>No symptoms at all</td>
</tr>
<tr>
<td>1</td>
<td>No significant disability: able to carry out all usual duties</td>
<td>No significant disability: despite symptoms, able to carry out all usual duties and activities</td>
</tr>
<tr>
<td>2</td>
<td>Slight disability: unable to carry out some of previous activities but able to look after own affairs without assistance</td>
<td>Slight disability: unable to perform all previous activities but able to look after own affairs without assistance</td>
</tr>
<tr>
<td>3</td>
<td>Moderate disability: requiring some help but able to walk without assistance</td>
<td>Moderate disability: requiring some help but able to walk without assistance</td>
</tr>
<tr>
<td>4</td>
<td>Moderately severe disability: unable to walk without assistance, and unable to attend to own bodily needs without assistance</td>
<td>Moderately severe disability: unable to walk without assistance and unable to attend to own bodily needs without assistance</td>
</tr>
<tr>
<td>5</td>
<td>Severe disability: bedridden, incontinent and requiring constant nursing care and attention</td>
<td>Severe disability: bedridden, incontinent and requiring constant nursing care and attention</td>
</tr>
<tr>
<td>6</td>
<td>NA</td>
<td>Death</td>
</tr>
</tbody>
</table>

NA indicates not applicable.

*Used commonly in clinical trials.
training and certification of raters should be considered to reduce inter-rater variability.

Validity
Validity is the degree to which an instrument measures the concept it was intended to measure. Construct validity is applied when a gold standard does not exist. This type of validity assessment uses multiple sources of comparison to test how accurately a measure captures the outcome it claims to measure in different contexts. Convergent (criterion) validity, a fundamental aspect of construct validity, measures the degree of correlation between different measures of the same construct. Other forms of validity include predictive validity (ability to predict future events)6,14 and theoretical validity (degree to which results are consistent with a priori expectations).18 In this report we focus on construct and convergent validity, clinical sensitivity and limitations, and then consider the application of the mRS in clinical trials of acute ischemic stroke treatments.

Construct Validity: Relationship to Stroke Severity
Construct validity of the mRS has been affirmed by multiple studies in which it has been consistently observed that the location, type and extent of stroke injury are closely related to short and longer-term disability (Table 3; detail in supplemental Table I, available online at http://stroke.ahajournals.org). Numerous investigations have reported an increased risk of poor outcome (defined as mRS ≥2 or >3) from discharge to 6 months for more severe types of stroke.19–28 For example, a study of 198 younger ischemic stroke patients showed that total anterior circulation infarction was an independent predictor of mRS grade ≥2 or death at 3 months (P<0.011).22 Similar results have been reported for older stroke populations.27,28 Studies in small patient series have consistently shown significant relationships between lesion volume (measured by diffusion-weighted and other imaging methods) and mRS grades (Table 3; detail in supplemental Table I), with larger lesions predicting more severe disability.20,29–34 As would be reasonably expected, improved brain perfusion and recanalization after thrombolytic therapy are also associated with improved mRS disability outcomes.21,33,35,36 For example, in a study of 177 acute ischemic stroke patients, recanalization within 5 hours post–recombinant tissue plasminogen activator treatment (in addition to National Institutes of Health Stroke Scale [NIHSS] baseline scores and other factors)

### TABLE 2. Reliability of the mRS With and Without Structured Interviews

<table>
<thead>
<tr>
<th>Structured Interview</th>
<th># Raters</th>
<th>N</th>
<th>Interval</th>
<th>Unweighted</th>
<th>Weighted</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test-Re-Test Reliability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>2</td>
<td>48</td>
<td>1–2 weeks</td>
<td>0.81, 0.95</td>
<td>0.94, 0.99</td>
<td>15</td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td>0.84, 0.97</td>
<td>0.96, 0.99</td>
<td></td>
</tr>
<tr>
<td>Inter-Rater Reliability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>34</td>
<td>100</td>
<td>NA</td>
<td>0.56</td>
<td>0.91</td>
<td>2</td>
</tr>
<tr>
<td>No</td>
<td>2</td>
<td>63</td>
<td>NA</td>
<td>0.44, 0.68</td>
<td>0.78, 0.94</td>
<td>5</td>
</tr>
<tr>
<td>Yes</td>
<td>58</td>
<td></td>
<td></td>
<td>0.70</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>15</td>
<td>117</td>
<td>NA</td>
<td>0.25</td>
<td>0.71</td>
<td>15</td>
</tr>
<tr>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td>0.74</td>
<td>0.91</td>
<td></td>
</tr>
</tbody>
</table>

NA indicates not applicable

### TABLE 3. Summary of mRS Construct Validity

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Predictors/Correlators of mRS Grades*†</th>
<th>Less severe</th>
<th>More severe</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stroke type</td>
<td>Proximal MCA occlusion</td>
<td>TACI, POCI, cortical and cardioembolic infarction, tandem ICA, MCA, and other arterial infarction</td>
<td>20, 23, 22, 24–26, 19, 27, 28, 46</td>
<td></td>
</tr>
<tr>
<td>Lesion volume/degree of perfusion</td>
<td>Recanalization within 300 minutes post–rt-PA, infarct size, improved brain perfusion, smaller lesion volume</td>
<td>Larger lesion volume, infarct size, continued volume expansion, lack of recanalization, absence of MCA blood flow</td>
<td>21, 29, 35, 32, 33, 3420, 29, 31–34, 36</td>
<td></td>
</tr>
<tr>
<td>Impairment</td>
<td>Lower NIHSS score at baseline, lower baseline LAMS score</td>
<td>More severe NIHSS (with or without rt-PA), LAMS, CNS, Mathew Scale, Orgogozo Scale, and SSS scores; and, impairment items: impaired consciousness at admission, limb weakness, hemianopia, apraxia, perceptual deficits, aphasia, consciousness commands, leg motor impairment, left arm palsy</td>
<td>11, 20–23, 26, 28, 31, 33, 37–41, 43–45</td>
<td></td>
</tr>
</tbody>
</table>

TACI indicates total anterior circulation infarction; ICA, internal carotid artery; MCA, middle cerebral artery; POCI, posterior circulation infarction; PACI, partial anterior circulation infarction; rt-PA, recombinant tissue plasminogen activator; CNS, Canadian Neurological Scale; LAMS, Los Angeles Motor Scale; ESS, European Stroke Scale; SSS, Scandinavian Stroke Scale.

*Significance of multivariate regression results expressed by P<0.05 and odds ratios with 95% CIs; results include a small number of correlation studies.
†mRS measured at 2 months to 1 year poststroke
Independently predicted mRS grade ≥2 at 3 months post-stroke (odds ratio = 4.11, 95% CI = 2.42 to 6.95; P < 0.001).21

Multivariate regression analyses have demonstrated that acute impairment score (measured with the NIHSS,20–22,37 Canadian Neurological Scale,26,38 Mathew scale,31 Los Angeles Motor Scale,39 or European Stroke Scale40) independently predict mRS grade at 2 months to 1 year poststroke (detailed in supplemental Table I). For example, Demchuk and coworkers observed that patients with the mildest strokes (NIHSS score 1 to 5) compared with those with more severe strokes (NIHSS scores 11 to 15, 16 to 20, >20) had improved chances of achieving a favorable outcome defined by mRS grade 0 to 1 (95% CI = 0.02 to 0.16, P < 0.001; 95% CI = 0.13 to 0.56, P < 0.001; 95% CI = 0.20 to 0.79, P = 0.008, respectively).41 Transitions in mRS grades have also been shown to correspond to transitions in NIHSS scores during stroke recovery.11 Consistent with these observations, moderate to strong correlations (Pearson r = 0.60 to 0.86) between acute impairment and mRS grades at discharge or follow-up have also been reported.38,40,42,43 Specific impairments that appear to contribute significantly to the more severe mRS grades include those involving alertness, orientation, leg motor function, following commands, and arm and hand function (detailed in supplemental Table I).23,44,45

Convergent Validity: Relationship to Other Disability Scales

Convergent validity of the mRS has been demonstrated by comparisons with other disability scales used to evaluate stroke patients, including the American Heart Association’s Stroke Outcomes Classification (AHA.SOC), the BI, the motor component of the Functional Independence Measure (m-FIM), the Short Form-36 (SF-36), and the Stroke Impact Scale (SIS; Table 4; detail in supplemental Table II, available online at http://stroke.ahajournals.org).7,8,11,43,46–48 For example, using stroke registry data, Kwon and coworkers quantified the frequency distribution of BI scores relative to mRS grades and showed that the highest BI scores (95 to 100, indicating excellent to complete recovery) correspond to mRS grades 0, 1 and 2.46

The particular relationship between the BI and the mRS has been explored in detailed investigations,7,8,46,48 reflecting the common use of these scales as end points in acute stroke treatment RCTs. Post hoc and quantitative translations between mRS grades and BI scores have been derived to facilitate comparisons among trial results.7,48 Whereas the correlation between trial end points of both scales is strong (r = 0.89, P < 0.001),48 their different structures, domains and scoring methods provide distinctive information. As described previously, the BI measures dependence in 10 BADLs, whereas the mRS captures higher functioning in addition to aspects of self-care. Eight of the ten BI domains reflect voluntary motor functions (see footnote†). However, poststroke disability can also affect speech, language or cognitive function.49 For example, a patient with substantial communication problems may still score ≥90 on the BI. This “ceiling effect” is a major disadvantage of the BI relative to a global disability instrument.9,45,46,49–51 Data from a large prospective cohort study of stroke patients demonstrated the limitations of the BI in mild stroke patients.55 The mRS was also reported to be more sensitive for distinguishing between mild and moderate disability, which suggests it may also be more sensitive to acute stroke treatment effects.45

Clinical Sensitivity

The clinical sensitivity or responsiveness of an instrument refers to its ability to detect a clinically important change.52 Limited information is available regarding the sensitivity of the mRS to changes in disability levels after a stroke. In a rehabilitation setting the sensitivity of 2 global disability measures, the mRS and the International Stroke Trial Measure (ISTM),53 and 2 ADL measures, the BI and FIM,44 were tested and compared in a nonrandom sample of 95 moderately disabled stroke patients.52 A change of one mRS grade was considered to be clinically significant based on the range of severity covered by the scale grades. Although the mRS was shown to be more sensitive than the ISTM (P < 0.001), it was less sensitive than either the BI (P < 0.002) or the FIM (P < 0.005), with the latter being the most sensitive of the 4 instruments under the conditions of the study. On this basis, the authors recommended the use of ADL scales for stroke intervention trials. However, other analyses (see below) based on treatment effects anticipated in the acute stroke setting suggest that those effects may be better detected using the mRS.

Limitations

A number of limitations apply to the mRS when used to measure disability outcome after stroke. A substantial literature documents the negative effect of patient comorbidities (including cardiovascular disease, diabetes, and arthritis),55–58 surgery,59 and socioeconomic factors60 on physical functioning, cognitive abilities,58 and overall health status, factors that may have a direct impact on the mRS.59 This is particularly important because comorbidities are common in stroke patients and the incidence of stroke in socioeconomically disadvantaged popula-

**TABLE 4. Summary of Convergent Validity**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Disability Scale*†</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ability to discriminate mRS grades</td>
<td>SIS-Participation: 3 grades</td>
<td>47</td>
</tr>
<tr>
<td>BI, mFIM: 3–4 grades</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>SIS-16, SF-36 PF/SF: 4 grades</td>
<td>11, 43, 47</td>
<td></td>
</tr>
<tr>
<td>Association/correlation/concordance with mRS grades and grade shifts</td>
<td>BI, AHA.SOC (for moderate stroke), Lawton IADL</td>
<td>11, 43</td>
</tr>
<tr>
<td>Cross-calibration of scores</td>
<td>BI</td>
<td>7</td>
</tr>
</tbody>
</table>

PF indicates Physical Functioning; SF, Social Functioning.

*Statistical significance for all results reported at P < 0.05.

†mRS measured at 1–6 months poststroke
tions is especially high. It is essential for the clinician to take these various attributes into account to avoid misapplication and misinterpretation of the mRS.

End Point in Clinical Trials
The mRS has been used often as an end point in RCTs of acute ischemic stroke treatments based on its straightforward application, acceptable inter-rater reliability, and ability to discriminate levels of stroke disability. Studies have found that the sample size requirements of trials using mRS-based end points are smaller than BI-based end points without the loss of statistical power. For example, the sample required for a neuroprotectant RCT using a mRS end point (dichotomized at grade ≤1) was estimated to be 38% of that required for a BI end point (dichotomized at ≥60).

Optimizing mRS end points for acute stroke treatment RCTs involves careful definition and appropriate statistical analyses. A “favorable” outcome defined as mRS grade ≤1 or ≤2 was estimated to be more powerful than dichotomization at higher grades. The importance of the cut-point for dichotomization was illustrated in a post hoc analysis of the ECASS II trial of alteplase. This study found that expanding the definition of “favorable” outcome from mRS grade ≤1 to grade ≤2 changed a statistically insignificant result to a significant one. Concerns over dichotomized end points center on the risk of failure to detect the impact of treatment. In a prospective study of 459 stroke patients, 116 subjects had transitioned from a baseline mRS score of 5 to a 3-month score of 4 or 3; or, from a baseline score of 4 to a 3-month score of 3. If those observations apply to RCTs, the reported shifts in mRS grades of more severe stroke patients toward reduced disability may not be captured by an end point defined as mRS ≤2, even though all mRS grade transitions are considered to be clinically meaningful.

Transition across the entire mRS grade spectrum has been proposed as a more comprehensive measure of impact for acute stroke interventions. This type of analysis uses the entire data set and has the advantage of reflecting simultaneously the risk and benefit of an intervention compared with an analysis based on a dichotomized end point. The SAINT I trial of the neuroprotectant NXY-059 for acute ischemic stroke exemplifies this approach. In this trial, the primary end point was prespecified as the mRS grade at 90 days (or last rating) with the analysis based on the overall difference in grade distribution of the 2 treatment groups.

Summary and Conclusions
A large and diverse literature supports the view that the mRS is a valid and clinically relevant instrument for assessing recovery from stroke and is a valuable end point for RCTs of acute ischemic stroke treatments. The scale consists of well-defined and easily understood grades that describe the range of global disability. When properly administered, the mRS exhibits a strong relationship with clinical measurements of stroke severity in addition to other disability and outcomes end points. However, clinicians should be aware that mRS grading may be affected by a variety of factors, including patient comorbidities and socioeconomic status. Use of the mRS in RCTs may benefit from a carefully defined end point as well as from a form of statistical analysis that is most likely to accurately reflect the effect of treatment.

Acknowledgements
The authors gratefully acknowledge the financial sponsorship of AstraZeneca LP. We greatly appreciate the contributions of Chris Pashos, PhD for his review and comments and of Yoko Komatsu-zaki, MPH, Er Chen, MPP, Yu-Chen Yeh, MS and Cassie Cai, MS, MS for their administrative and editorial assistance.

Disclosures
None.

References
21. Molina CA, Alexandrov AV, Demchuk AM, Saqur M, Uchino K, Varez-Sabun J. Improving the predictive accuracy of recanalization on
Outcomes Validity and Reliability of the Modified Rankin Scale: Implications for Stroke Clinical Trials: A Literature Review and Synthesis

Jamie L. Banks and Charles A. Marotta

Stroke. 2007;38:1091-1096; originally published online February 1, 2007;
doi: 10.1161/01.STR.0000258355.23810.c6
Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2007 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/38/3/1091

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/