Prestroke Modified Rankin Stroke Scale Has Moderate Interobserver Reliability and Validity in an Acute Stroke Setting

Patricia Fearon, MRCP; Kate S. McArthur, MRCP; Kevin Garrity; Laura J. Graham; Geraldine McGroarty; Sarah Vincent; T. J. Quinn, MD

Background and Purpose—The modified Rankin Scale (mRS) is the recommended functional outcome assessment in stroke trials. Utility of mRS may be limited by interobserver variability. Prestroke function, described using mRS, is often used as trial entry criterion. We assessed the reliability and validity of prestroke mRS in acute stroke.

Methods—We present two complementary analyses of the properties of prestroke mRS: (1) Paired interviewers (trained in mRS) performed independently a blinded assessment of mRS and prestroke mRS. Interobserver variability was described using percentage agreement and weighted (kw) κ statistics with 95% confidence interval (95% CI). Validity was assessed by comparing prestroke mRS with other markers of function (comorbidity; medication count; need for carers). (2) We further assessed validity using a larger retrospective dataset. We compared prestroke mRS with Charlson comorbidity index (CCI) and the Rockwood frailty index. Rank correlation coefficient or Fisher exact test were used as appropriate.

Results—Paired interviewers assessed 74 stroke survivors. Median standard mRS was 4 (interquartile range [IQR], 2–4), median prestroke mRS was 1 (IQR, 0–3; range, 0–4). Reliability for standard mRS interview was 56% agreement, kw=0.55 (95% CI, 0.39–0.71). Reliability for prestroke mRS was 70%, kw=0.70 (95% CI, 0.53–0.87). The retrospective dataset described 231 subjects. In this data set, Spearman Rho for prestroke mRS and frailty index was 0.82 (95% CI, 0.78–0.86); CCI 0.50 (95% CI, 0.40–0.59); patient age 0.45 (95% CI, 0.34–0.54); medication count 0.28 (95% CI, 0.15–0.40). There was no association between need for carers and prestroke mRS (p=0.10).

Conclusions—Interobserver reliability of prestroke mRS is limited but comparable with standard mRS. Poor correlation of prestroke mRS with certain markers of function suggests limited validity. Our data would suggest that relying on mRS alone may be a suboptimal measure of prestroke function and could potentially bias trial samples. (Stroke. 2012;43: 3184–3188.)

Key Words: clinical trial ■ disability evaluation ■ methods ■ outcomes ■ stroke
Methods

Assessment of Interobserver Variability and Validity of Prestroke mRS

Setting
The initial study was conducted in the stroke units of 2 urban teaching hospitals (Glasgow Royal Infirmary and Western Infirmary, UK). Data collection was performed over a 4-week period in June 2011. These units admit all suspected strokes regardless of age, prestroke function, or severity of stroke.

Participants
Assessors were a group of 4 medical students undertaking a period of elective study in stroke, and all were trained and certified in mRS assessment (http://trials-rankin.trainingcampus.net). Patients were consecutive, consenting, inpatient stroke survivors. We used no specific exclusion criteria, although patients had to be medically stable to allow an attempt at mRS interview. If required, assessment could be supplemented by a proxy information source.

Procedure
Paired researchers from the pool of 4 performed same-day standard and prestroke mRS assessments. The mRS assessment was based on unstructured direct interview. Researchers were blinded to each other’s scores. Choice of assessor and sequence of assessment (mRS or prestroke mRS first) was randomized using an online resource and allocation held in sealed opaque envelopes. (Figure 1).

The assessor pair not performing mRS interviews independently collected clinical and demographic details from patient’s case records, including presence of cognitive impairment and communication problems (as assessed by the parent clinical team using ward-based observation, direct and indirect screening tools, and specialist assessment as required). In addition, they collected other metrics specific to prestroke function: number of medications, need for carers/social support (as defined by subject or proxy), and number of co-morbid medical conditions (as defined by community General Practitioner electronic patient summary records).

Assessment of Prestroke mRS Validity Using a Retrospective Dataset
A complementary analysis of validity was conducted using a database registry held in the stroke unit of Glasgow Royal Infirmary.

Participants
Patients were consecutive acute stroke admissions from a geographically defined population in East Glasgow (postcode G31, G32, G33, G34, and G40) presenting between February 2011 and January 2012. Patients with transient ischemic attack were excluded.

Procedure
Standard clinical and demographic data for each patient were recorded from patient’s case records by a single researcher (P.F.). Patients were assessed for prestroke functional ability by the treating clinical team. Prestroke mRS and a frailty index were recorded after multidisciplinary discussion. Prestroke comorbidity and medication count, including all long-term prescriptions, were derived from in-patient case records and electronic General Practitioner patient summaries and prescribing data.

For the database analysis we used Charlson comorbidity index (CCI) and Rockwood frailty scoring. CCI is a weighted index of comorbidity originally developed to predict mortality. It has previously been validated in a stroke population for prediction of functional outcomes. The frailty index developed by Rockwood is a multidomain evaluation based on 4 levels of physical impairment. It has been validated in a community-dwelling geriatric population as a predictor of death and institutionalisation (Table 1).

Table 1. Description of Scoring of Comorbidity and Frailty Indices

<table>
<thead>
<tr>
<th>Charlson Comorbidity Index</th>
<th>Rockwood Frailty Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weighted Score</td>
<td>Condition</td>
</tr>
<tr>
<td>1</td>
<td>Myocardial infarct</td>
</tr>
<tr>
<td></td>
<td>Congestive heart failure</td>
</tr>
<tr>
<td></td>
<td>Peripheral vascular disease</td>
</tr>
<tr>
<td></td>
<td>Cerebrovascular disease</td>
</tr>
<tr>
<td></td>
<td>Dementia</td>
</tr>
<tr>
<td></td>
<td>Chronic pulmonary disease</td>
</tr>
<tr>
<td></td>
<td>Connective tissue disease</td>
</tr>
<tr>
<td></td>
<td>Peptic ulcer disease</td>
</tr>
<tr>
<td></td>
<td>Mild liver disease</td>
</tr>
<tr>
<td></td>
<td>Diabetes</td>
</tr>
<tr>
<td>2</td>
<td>Hemiplegia</td>
</tr>
<tr>
<td></td>
<td>Moderate or severe renal disease</td>
</tr>
<tr>
<td></td>
<td>Diabetes with end-organ damage</td>
</tr>
<tr>
<td></td>
<td>Any tumor</td>
</tr>
<tr>
<td></td>
<td>Leukemia or Lymphoma</td>
</tr>
<tr>
<td>3</td>
<td>Moderate or severe liver disease</td>
</tr>
<tr>
<td>6</td>
<td>Metastatic solid tumor</td>
</tr>
<tr>
<td></td>
<td>Acquired immunodeficiency syndrome</td>
</tr>
</tbody>
</table>

*Basic activities of daily living (ADL): defined as eating, dressing, bathing, and bed transfers.
Analysis
For our first aim of describing standard and prestroke mRS reliability we used a number of complementary approaches, describing reliability with unweighted $\kappa$ statistics; $\kappa$ weighted $\kappa$ (both with corresponding 95% confidence interval [95% CI]) and percentage agreement between assessors.$^{13}$ We used quadratic weighting of $\kappa$, which approximates to the intraclass correlation coefficient.

For our second aim, various paired comparisons from both analyses were used to assess validity. To fulfill the statistical assumption of independent data, data from assessor pairs in the initial study were averaged. We performed further sensitivity analyses limited to each single assessor’s data. Proportions living alone with no external care were compared for those graded independent (prestroke mRS=1) and dependent (prestroke $>1$) using Fisher exact test. We described the relationship between prestroke mRS and number of medications, co-morbidity indices, frailty index, and patient age using rank correlation coefficient.

Statistical analyses were performed using SPSS (version 18.0) and Statsdirect software (Stats Direct Ltd, UK). Data were collected in line with Scottish Stroke Care Audit procedures,$^{14}$ and corresponding data governance measures were followed. Additional data collection for the retrospective database had research ethics committee approval.

Results
Assessment of Interobserver Variability and Validity of Prestroke mRS
Clinical and Demographic Factors
We assessed mRS in 74 stroke survivors; paired mRS data were available for 71. The cohort comprised a variety of clinical stroke presentations: Total anterior circulation strokes, 13 (18%); Partial anterior circulation strokes, 27 (36%); Lacunar strokes, 15 (20%); Posterior circulation strokes, 9 (12%); transient ischemic attack or unclassified, 10 (14%). Median age was 72 years (IQR, 62–79), median time from stroke was 5 days (IQR, 3–9). Substantial cognitive impairment was documented in 7 patients (9%), and substantial communication difficulty was documented in 15 (20%). Six (8%) subjects required additional detail from a proxy.

Median standard mRS was 4 (IQR, 2–4), and median prestroke mRS was 1 (IQR, 0–3; range, 0–4); 27 (38%) of subjects assessed had a degree of prestroke disability defined as mRS $\geq 1$.

Reliability
Interobserver variability for standard mRS assessment of stroke survivors was moderate, $k=0.40$ (95% CI, 0.27–0.52), $\kappa=0.55$ (95% CI, 0.39–0.71). Comparing paired raters’ scores, 56% matched.

Reliability for prestroke mRS was moderate $k=0.58$ (95% CI, 0.46–0.70), $\kappa=0.70$ (95% CI, 0.53–0.87). Comparing paired raters’ scores, 70% matched.(Table 2)

Validity
Spearman’s Rho for paired prestroke mRS and co-morbidity was 0.31 (95% CI, 0.08–0.50) for prestroke mRS and for number of medications was 0.33 (95% CI, 0.11–0.52). There was no association between need for carers and dependence on prestroke mRS ($P=0.10$). Sensitivity analyses limited to single observer data were as follows: Spearman Rho mRS and comorbidity, 0.25 (0.01–0.45) and 0.18 (0.06–0.40); medications, 0.24 (0.01–0.44) and 0.45 (0.24–0.61); Fisher’s exact test ($P=0.07$ and $P=0.25$).

Discussion
Our data suggest that there is a degree of interobserver variability in the assessment of prestroke mRS, albeit this is probably no greater than the variability seen for standard mRS assessment.$^{7}$ Formal comparison of $\kappa$ statistics between the groups was not possible, because the range of standard and prestroke mRS scores differ and this will impact on the interobserver variability. For assessment of reliability, our comparator analysis described standard mRS, using mRS assessment in the first days post-stroke. Assessment of mRS in the acute period is complicated, because the wording of certain mRS grades is not suited to a hospital environment where constant assistance is available. However, acute mRS is used in trials, and a literature describing its properties is available.$^{13}$ Ideally the mRS assessment would have been performed at the usual trial end point times of 30 or 90 days. Because our primary analysis was of the properties of prestroke mRS, which is measured in the first days post-stroke, we opted to assess both prestroke and standard mRS at the same time points.

Table 2. Reliability of Paired Assessments of mRS

<table>
<thead>
<tr>
<th>Assessor A</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessor B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>28</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>10</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

mRS indicates modified Rankin scale
Data are mRS scores (0–5) from paired assessors (A, B); n=71 paired assessments.

Assessment of Prestroke mRS Validity Using a Retrospective Dataset
Clinical and Demographic Factors
There were 231 stroke admissions from our target population. Clinical stroke presentations comprised the following: Total anterior circulation strokes, 45 (20%); Partial anterior circulation strokes, 63 (27%); Lacunar strokes, 93 (40%); Posterior circulation strokes, 23 (10%). Median age was 74 years (IQR, 63–82), and 5% were institutionalized before admission.

Median baseline mRS was 3 (IQR, 1–4), and median prestroke mRS was 1 (IQR, 0–2). Prestroke disability defined as mRS $\geq 1$ was seen in 113 patients (54%).

Median prestroke CCI score was 3 (IQR, 2–5), and only a minority of patients (n=28; 13%) had no prestroke comorbidity defined as CCI score of zero. Median prestroke frailty score was 0 (IQR, 0–2), and 74 patients (38%) had at least 1 prestroke marker of frailty.

Validity
Spearman’s Rho for prestroke mRS and frailty index was 0.82 (95% CI, 0.78–0.86) for prestroke mRS and 0.50 (95% CI, 0.40–0.59) for CCI (Figure 2). Spearman’s Rho for mRS and patient age was 0.45 (95% CI, 0.34–0.54) and for number of medications was 0.28 (95% CI, 0.15–0.40).

Table 2. Reliability of Paired Assessments of mRS

<table>
<thead>
<tr>
<th>Assessor A</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessor B</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>28</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>10</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

mRS indicates modified Rankin scale
Data are mRS scores (0–5) from paired assessors (A, B); n=71 paired assessments.
Our validity analyses suggest that prestroke mRS may be a suboptimal measure of function. Correlation with markers of comorbidity was moderate in both samples. Correlation with our frailty index was reasonable, however this is to be expected because certain measures are common to both the mRS and Rockwood score (ie, continence and mobility). Agreement was not perfect; only a third of patients had evidence of frailty, whereas more than half scored as dependent on prestroke mRS. Modest but important numbers had high levels of frailty and comorbidity but low stroke disability as measured by mRS. The discrepancy between prestroke mRS and proportions living independently at home is concerning, as dependence is used in the definition of certain mRS grades.

The potential suboptimal validity of prestroke mRS has implications for research. Patients with prestroke disability may be erroneously scored as independent on pre-morbid mRS, and their inclusion in trials may bias results, conversely previously independent subjects may be inappropriately denied trial entry on the basis of perceived prestroke dependency. Prestroke assessment of function also has implications for clinical practice, for example prestroke dependency (usually measured by mRS) is a contraindication to use of intravenous thrombolytic therapy.16

Our data suggest that prestroke functional impairment is prevalent in an unselected stroke cohort. The proportion with prestroke impairment in this study is similar to that seen in previous observational work.17 If trial entry criteria based on mRS are used, a substantial number of stroke patients will be ineligible for recruitment, and this will impact on generalizability of study results.

There is no gold standard assessment of function that we could use as reference for comparison with prestroke mRS. In the absence of any guidance on reference standard for pre-morbid function, we chose metrics that were objective, readily available, and reflect social support, comorbidity, and frailty. For our database analysis we were able to use more sophisticated measures of premorbid function. However, we accept that comorbidity and frailty are not synonymous with the activity limitation (disability) measured by mRS. The frailty concept has no consensus definition but is considered a multidimensional construct reflecting variable vulnerability to adverse health outcomes.18 Both frailty and comorbidity have increased prevalence in older adults.

If validity of standard mRS for prestroke function is suboptimal, there may be simple interventions to allow improvement. Collateral account from family or carers may be less subject to recall bias.19 Likewise, various options to improve the properties of mRS (consensus review, a questionnaire approach) have been described and may be equally useful in prestroke assessment.2021 Other approaches could include use of a structured informant questionnaire as is common in dementia clinical trials22 or complementing prestroke mRS with information from more specific functional scales, such as the Barthel Index6 or Nottingham activities of daily living scale.23 Keeping the structure of mRS but changing the wording of certain grades to make the scale more suited to prestroke assessment would also be possible. Equally, the combination of prestroke mRS with comorbidity or frailty indices may allow a more global assessment of health and functional ability. CCI is known to be a predictor of poor outcomes after stroke.11 Although these changes seem intuitively attractive, any change in approach would require robust clinimetric assessment before incorporation into clinical trial protocols and clinical practice.

Strengths of our approach include a representative sample of real-world stroke survivors. Our intention was to describe properties of proxy mRS across a broad cohort of stroke survivors, and we included a reasonably high proportion of disabling strokes with cognitive and communication impairment. The database analysis focused on a distinct population with a high prevalence of comorbidity and social deprivation. We recognize that properties of premorbid mRS may differ across international centers with differing patterns of community disability and referral. Our study methodology and reporting followed recommended guidance on clinimetric studies.24

We accept that the present study had several methodological limitations. For our assessment of prestroke mRS interobserver variability, assessors were not experienced stroke researchers. However, all assessors had completed a certification examination in mRS grading and so had the level of basic mRS training required for stroke trials.25 It is reassuring that this group’s variability in standard mRS was similar to that seen in previous studies of mRS reliability.7 Numbers in this analysis were modest but similar to previous studies of mRS properties.1 Numbers with higher prestroke mRS grades were modest but in keeping with expected local referral patterns and UK epidemiology of major disability in community settings. Our database analysis of prestroke mRS validity was limited to data from a single observer. Patients were not directly interviewed but had comprehensive multidisciplinary assessment by the treating clinical team with prestroke mRS determined following consensus discussion. Use of the clinical team’s unstructured assessment as a measure of cognitive and language problems may have missed subtle issues, but our intention was to detect clinically important problems.

There is no formal guidance on assessment of prestroke function for trials. Our data would suggest that relying on prestroke mRS alone may be suboptimal and may bias clinical trials and practice. We encourage greater rigor in prestroke mRS assessment. Research to describe the utility of alternative mRS
methodologies or use of additional measures of prestroke function would improve assessment.

Disclosures
Dr Quinn has received grant support for studies of stroke assessment and has received payment for developing educational resources for stroke researchers. The other authors report no conflicts.

References
Prestroke Modified Rankin Stroke Scale Has Moderate Interobserver Reliability and Validity in an Acute Stroke Setting
Patricia Fearon, Kate S. McArthur, Kevin Garrity, Laura J. Graham, Geraldine McGroarty, Sarah Vincent and T. J. Quinn

Stroke. 2012;43:3184-3188; originally published online November 13, 2012; doi: 10.1161/STROKEAHA.112.670422

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/43/12/3184

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/