Comparison of Psychometric Properties of Three Mobility Measures for Patients With Stroke

I-Ping Hsueh, MA; Chun-Hou Wang, BS; Ching-Fan Sheu, PhD; Ching-Lin Hsieh, PhD

Background and Purpose—This study compared the validity, responsiveness, and interrater reliability of 3 mobility measures in stroke patients from the acute stage up to 180 days after stroke onset. The 3 measures were the Rivermead Mobility Index (RMI), a modified RMI (MRMI), and the Mobility Subscale of the Stroke Rehabilitation Assessment of Movement (STREAM).

Methods—The validity and responsiveness of the 3 mobility measures were prospectively examined by monitoring 57 stroke patients with the measures and the Barthel Index at 14, 30, 90, and 180 days after stroke onset. Two individual raters used the 3 measures to evaluate a different sample of 40 patients on 2 separate occasions to determine the interrater reliability.

Results—The Spearman $\rho$ between STREAM and MRMI was $\geq 0.92$; the intraclass correlation coefficient (ICC, a measure of agreement) between them was $\geq 0.89$, indicating high concurrent validity of both measures. RMI showed a moderate to high relationship and agreement with STREAM and MRMI ($\rho \geq 0.78$, ICC $\geq 0.5$). Responsiveness of the 3 measures was high before 90 days after stroke onset (standardized response mean $\geq 0.83$) and low at 90 to 180 days after stroke onset (0.2 $\geq$ standardized response mean $\geq 0.4$). The score changes of the 3 measures at each stage were significant ($P \leq 0.05$), except for RMI and MRMI at 90 to 180 days after stroke onset. The interrater agreement of the 3 measures was high (ICC $\geq 0.92$).

Conclusions—All 3 measures examined showed acceptable levels of reliability, validity, and responsiveness in stroke patients. The psychometric characteristics of STREAM were slightly superior to those of the other 2 measures among our patients. We prefer and recommend STREAM for measuring mobility disability in stroke patients. (Stroke. 2003; 34:1741-1745.)

Key Words: cerebrovascular accident • disability evaluation • psychometrics

Mobility is essential to developing an independent lifestyle after stroke and is perhaps the particular ability that patients consider the most important.1 Therefore, improving mobility is 1 of the major goals of stroke rehabilitation.2 To identify mobility disabilities and manage their associated problems, both clinicians and researchers need a measure of mobility that is simple to administer and has sound psychometric properties (eg, reliability, validity, and responsiveness).3

The Rivermead Mobility Index (RMI), one of the few clinical mobility measures specifically designed for stroke patients, is the only mobility measure endorsed by the US Agency for Health Care Policy and Research.4 RMI has been used in several studies to examine treatment effect on mobility.5 According to Lennon and Hastings,6 the responsiveness of RMI is poor because its items are scored on a dichotomous (yes/no) basis. Therefore, the modified RMI (MRMI)2 was developed to increase the responsiveness of the measure by extending the scoring level to 6 points. However, the psychometric properties of the MRMI have yet to be systematically explored. The Mobility Subscale of the Stroke Rehabilitation Assessment of Movement Measure (STREAM)7 that assesses mobility after stroke is a more recently developed measure. STREAM is simple to administer and is reliable and valid in stroke patients.7,8 Although the 3 measures appear to be accepted by both clinicians and researchers to measure mobility disability after stroke, to the best of our knowledge, no empirical data exist to determine which measure is most psychometrically sound.

Comparison of the psychometric properties of clinical mobility measures can provide useful guidelines for both clinicians and researchers to determine an objective and scientific measure.8 The purpose of this study was to compare the reliability, validity, and responsiveness of RMI, MRMI, and STREAM in stroke patients.

Received October 29, 2002; final revision received January 17, 2003; accepted January 28, 2003.

From the School of Occupational Therapy, College of Medicine, National Taiwan University, Taipei, Taiwan (I.-P.H., C.-L.H.); School of Physical Therapy, College of Medical Technology, Chun-Shan Medical University, Taichung, Taiwan (C.-H.W.); and Department of Psychology, DePaul University, Chicago, Ill (C.-F.S.).

Correspondence to Ching-Lin Hsieh, School of Occupational Therapy, College of Medicine, National Taiwan University, No 7, Chung-Shan S Rd, Taipei 100, Taiwan, ROC. E-mail mike26@ha.mc.ntu.edu.tw

© 2003 American Heart Association, Inc.

Stroke is available at http://www.strokeaha.org DOI: 10.1161/01.STR.0000075295.45185.D4

1741
Methods

Subjects
Subjects were recruited from stroke patients consecutively admitted to National Taiwan University Hospital between April 1 and December 31, 2001. Patients were included in the study if they met the following criteria: (1) diagnosis (International Classification of Diseases, ninth revision, clinical modification [ICD-9] codes) of cerebral hemorrhage (ICD-9, 431) or cerebral infarction (ICD-9, 434), (2) first onset of stroke without other major disease and the absence of a preexisting disability, (3) stroke onset within 14 days before admission, (4) ability to follow instructions, (5) willingness to participate in this study, and (6) residence in the greater Taipei area. The clinical diagnosis of stroke was confirmed by neuroimaging examination. Patients who suffered another stroke or had another major disease during the follow-up period were excluded.

Procedures
The study protocol consisted of 2 parts. The first part was a validity and responsiveness study. The 3 mobility measures and the Barthel Activities of Daily Living index (BI) were administered by occupational therapists at 14, 30, 90, and 180 days after stroke onset. The BI was used as the external criterion to examine convergent and predictive validity. Initial stroke severity was ascertained with the Canadian Neurological Scale, applied retrospectively to medical records. Degrees of responsiveness of the mobility measures were calculated from the changes occurring between 14 to 30, 30 to 90, 90 to 180, 14 to 90, and 14 to 180 days after stroke onset.

When necessary, patients were allowed to rest during the testing protocol, which lasted ~1 hour. All of the above assessments were made by occupational therapist A, according to previously published standardized methods of administration. The second part of the protocol was an interrater reliability study. Forty stroke patients at a rehabilitation unit participated in this part of the study. The 3 mobility measures were administered separately by 2 occupational therapists (A and B). To minimize the effects of possible recovery, assessments were administered within a 24-hour period according to a counterbalanced sequence. The therapists were blinded to the results of each other’s assessments during the study period.

Instruments
RMI, which covers a range of hierarchical activities from turning over in bed to running, comprises 14 questions and 1 direct observation. Each patient’s mobility performance is rated primarily by interviewing the patients and/or their primary caregiver. The highest score, 15, indicates the highest mobility status. Although previous studies have found that RMI had good psychometric properties in stroke patients, sample sizes in 2 of these studies were modest (<23). Limiting generalization of their results.

MRMI has 8 test items: turning over, changing from lying to sitting, maintaining sitting balance, going from sitting to standing, standing, transferring, walking indoors, and climbing stairs. Scores of the MRMI range from 0 to 40. One main characteristic of the MRMI is that patients are scored by direct observation of their movement. Lennard and Johnson have previously reported the psychometric characteristics of the MRMI by examined further.

STREAM measures mobility after stroke by direct observation. It contains 10 four-point items: rolling, bridging, going from supine to sitting, changing from sitting to standing, standing, placing affected foot onto first step, stepping backward, stepping to affected side, walking 10 m, and walking down stairs. Scores of STREAM range from 0 to 30. Although the reliability and validity of STREAM are high in stroke patients, its responsiveness has not been reported.

The BI is a weighted scale of 10 items of basic activities of daily living (ADL). The highest score, 100, indicates that the patient is fully independent in physical function; the lowest score, 0, represents a totally dependent, bedridden state. The reliability, validity, and responsiveness of the BI are well established in stroke patients. Its responsiveness has not been reported. The BI was used as the internal criterion to examine convergent and predictive validity. Initial stroke severity was ascertained with the Canadian Neurological Scale, applied retrospectively to medical records. Degrees of responsiveness of the mobility measures were calculated from the changes occurring between 14 to 30, 30 to 90, 90 to 180, 14 to 90, and 14 to 180 days after stroke onset.

Statistical Analyses
A mobility measure should be able to reflect the whole range of mobility disability after stroke. We calculated the floor and ceiling effects, representing the percentage of subjects achieving the lowest and highest scores possible, respectively. Floor and ceiling effects exceeding 20% of sample size are considered to be significant, indicating that the measure can represent only a limited range of mobility disability.

Concurrent validity is usually established by demonstrating a high correlation or agreement between the measure and a gold standard. Because each of the 3 measures used has a different score range, the scores from each measure were converted to a range of 0 to 100. The relationship and agreement between the 3 mobility measures at 4 time points were examined by use of the Spearman correlation coefficient (ρ) and the intraclass correlation coefficient (ICC), respectively.

The convergent validity of the mobility measures was assessed by examining the relationships between the total scores of the mobility measures and those of the BI at all 4 time points after stroke using Spearman’s ρ. The predictive validity of the mobility measures was assessed by examining the associations between results of the mobility measures at 3 time points (14, 30, and 90 days after stroke onset) and those of the BI at 180 days after stroke onset using the Spearman ρ.

Responsiveness was examined with the standardized response mean (SRM). A type of effect size. SRM was calculated by dividing the mean change score by the SD of the change score in the same subjects. An effect size ≥0.8 is usually considered large; 0.5 to 0.8, moderate; and 0.2 to 0.5, small. Wilcoxon matched-pairs signed-rank tests were performed to determine the statistical significance of the change scores.

The interrater agreement on individual items of the mobility measures was analyzed with the quadratic weighted κ statistic. The interrater agreement of the total score of the mobility measures was analyzed with the ICC statistic. The fixed effect of ICC model was used to compute the ICC value for interrater reliability. Both weighted κ and ICC values ≥0.80 indicate very good agreement; 0.60 to 0.79, good agreement; 0.40 to 0.59, moderate agreement; 0.20 to 0.39, fair agreement; and 0 to 0.2, poor agreement.

Results
A total of 59 patients met the selection criteria. We excluded 157 patients because of stroke onset that was more than 14 days before admission, the patient living outside the greater Taipei area, the occurrence of recurrent stroke, and/or communication difficulties. Two patients who met the selection criteria declined to participate in the validity and responsiveness study. Of the remaining 57 patients, 43 completed the follow-up at 180 days after stroke. The Canadian Neurological Scale scores showed that patients had a broad range of severity (from mild to severe stroke) at admission. In addition, the BI scores indicated that patients were severely disabled at baseline. Nonetheless, 42% of the patients were nearly independent at 180 days after stroke onset (BI ≥95). Characteristics of the study sample are presented in Table 1.

The interquartile score range of the RMI at baseline was quite limited (Table 1). Furthermore, Table 2 shows that, except for RMI, none of the mobility measures exhibited significant floor or ceiling effects at the 4 time points after stroke. These results indicate that MRMI and STREAM demonstrated acceptable distribution from the acute stage up to 180 days after stroke onset.
Table 3 shows that the correlation ($\rho=0.92$) and agreement (ICC=0.89) between STREAM and MRMI were high, indicating that both measures had high concurrent validity. RMI showed moderate to high concurrent validity ($\rho=0.78$, ICC=0.50) when evaluated against STREAM and MRMI. Table 4 shows that the 3 mobility measures had high convergent validity ($\rho=0.72$) and acceptable predictive validity ($\rho=0.5$).

The 3 mobility measures were highly responsive in detecting changes before 90 days after stroke onset (14 to 30 days, SRM=1.14; 30 to 90 days, SRM=0.83; Table 5). At 90 to 180 days after stroke onset, the levels of responsiveness of these measures, as expected, were low ($0.2 \leq SRM \leq 0.4$). The changes shown by the 3 measures at each stage were all significant ($P<0.05$), except for those shown by RMI and MRMI at 90 to 180 days after stroke onset ($P>0.14$).

Forty patients participated in the interrater reliability study. This sample consisted of 19 men and 21 women with a mean age of 63 years (SD, 10.2 years). The medians of the weighted $\kappa$ statistic for each item of RMI, MRMI, and STREAM were 0.71 (range, 0.37 to 0.94), 0.72 (range, 0.47 to 0.9), and 0.81 (range, 0.55 to 0.89), respectively, indicating generally acceptable interrater agreement on the item level. Four RMI items, 2 MRMI items, and 1 STREAM item had fair to moderate agreement ($0.37 \leq \kappa \leq 0.6$). The ICCs for the total scores of RMI, MRMI, and STREAM were 0.92 (95% confidence interval [CI], 0.84 to 0.96), 0.95 (95% CI, 0.90 to 0.97), and 0.97 (95% CI, 0.95 to 0.99), respectively, indicating excellent total score agreement.

### Discussion

A simple and psychometrically sound mobility measure enables both clinicians and researchers to identify, monitor, and manage mobility disability after stroke. This study is the first to compare the psychometric properties of the RMI, MRMI, and STREAM mobility measures in stroke patients concurrently and systematically. In addition, patients in the study were monitored at 4 specific time points after stroke for an extended period (up to 180 days after stroke onset) to evaluate how appropriate these measures were for use at different recovery stages after stroke. The findings of this study can provide information useful for the selection of mobility measures for both clinicians and researchers.

The score distributions of the measures of the study sample should not exhibit severe ceiling or floor effects. We found that all 3 mobility measures demonstrated acceptable distributions from the acute stage up to 180 days after stroke onset, except for RMI, which at 14 days after stroke onset showed a limited score range and a notable floor effect. These results indicate that the RMI might not adequately characterize patients’ mobility functions in the early stages of stroke, especially for patients with severe disabilities.

The validity of the 3 mobility measures has been reported in previous studies. However, given the notable heterogeneity of stroke effects, comparison with previous results

<table>
<thead>
<tr>
<th>Time After Stroke, d</th>
<th>RMI, n (%)</th>
<th>MRMI, n (%)</th>
<th>STREAM, n (%)</th>
<th>BI, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Floor</td>
<td>Ceiling</td>
<td>Floor</td>
<td>Ceiling</td>
<td>Floor</td>
</tr>
<tr>
<td>14 (n=57)</td>
<td>23 (40.4)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>30 (n=54)</td>
<td>6 (11.1)</td>
<td>2 (3.7)</td>
<td>0 (0)</td>
<td>3 (5.6)</td>
</tr>
<tr>
<td>90 (n=44)</td>
<td>1 (2.3)</td>
<td>3 (6.8)</td>
<td>0 (0)</td>
<td>6 (13.6)</td>
</tr>
<tr>
<td>180 (n=43)</td>
<td>0 (0)</td>
<td>2 (4.7)</td>
<td>0 (0)</td>
<td>7 (16.3)</td>
</tr>
</tbody>
</table>

### Table 1. Characteristics of the Stroke Patients

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>All 57 Patients</th>
<th>43 Patients Who Completed the Study</th>
<th>14 Patients Who Failed to Complete the Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex, M/F, n</td>
<td>34/23</td>
<td>26/17</td>
<td>8/6</td>
</tr>
<tr>
<td>Mean (SD) age, y</td>
<td>64.2 (11.5)</td>
<td>63.6 (11.4)</td>
<td>65.9 (12.0)</td>
</tr>
<tr>
<td>Diagnosis, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cerebral hemorrhage</td>
<td>18 (32)</td>
<td>13 (30)</td>
<td>5 (36)</td>
</tr>
<tr>
<td>Cerebral infarction</td>
<td>39 (68)</td>
<td>30 (70)</td>
<td>9 (64)</td>
</tr>
<tr>
<td>Side of lesion, right/left, n</td>
<td>29/28</td>
<td>21/22</td>
<td>8/6</td>
</tr>
<tr>
<td>Canadian Neurological Score at admission, median (interquartile range)</td>
<td>7 (4.5–6)</td>
<td>7 (4.5–8.5)</td>
<td>5.5 (4–8)</td>
</tr>
<tr>
<td>At 14 days after stroke onset, median (interquartile range)</td>
<td>RMI score</td>
<td>2 (0–5)</td>
<td>2 (0–5)</td>
</tr>
<tr>
<td>MRMI score</td>
<td>23 (13–31)</td>
<td>25 (15–32)</td>
<td>20 (9–25)</td>
</tr>
<tr>
<td>STREAM score</td>
<td>14 (7–21)</td>
<td>15 (8–21)</td>
<td>10 (5–18)</td>
</tr>
<tr>
<td>BI score</td>
<td>30 (15–50)</td>
<td>30 (15–55)</td>
<td>30 (15–45)</td>
</tr>
</tbody>
</table>
was difficult because they were not examined in the same group of patients. The validity of the 3 measures was first compared in a cohort of patients in this study. We found that STREAM and MRMI had high concurrent, convergent, and predictive validity, whereas RMI was not highly correlated with either STREAM or MRMI. It is of note that the RMI score was retrieved mainly via interview, whereas the STREAM and MRMI scores were based on direct observation of a patient’s performance. Patients may overstate their functional abilities.22 Our results suggest that among our patients, STREAM and MRMI were more valid measures of mobility after stroke than RMI. However, the differences in validity between the 3 measures may not be statistically significant.

Responsiveness is important for any measurement tool designed to measure change over time.23 Nonetheless, the responsiveness of most mobility measures has yet to be examined. The 3 mobility measures were highly responsive in detecting changes before 90 days after stroke onset. All changes in the 3 measures at each stage were significant, except for the RMI and MRMI at 90 to 180 days after stroke onset. According to the present results, STREAM was slightly more responsive than the other 2 measures. Responsiveness of the 3 measures was, as expected, low at later stages of recovery (90 to 180 days after stroke onset). One possible explanation is that patients’ improvements in mobility had reached a plateau after 90 days after stroke onset. Moreover, balance, motor, and ADL functions showed only minor improvement after 90 days after stroke onset.24,25 Another possible explanation is that these 3 mobility measures lack items sensitive enough to detect change >90 days after stroke onset.

Interestingly, MRMI was no more responsive than RMI, despite the fact that MRMI, with more scoring levels, was revised from RMI to make it more responsive.2 Some recent studies have demonstrated that an increase in the number of items or grading levels does not necessarily improve the responsiveness or difference detection between patient groups of mobility measures,26 balance measures,27 and ADL measures.16,27,28 These results support the argument that selection of a measurement tool should be based on empirical evidence, not on clinical opinion.27

Interrater agreement on individual items of mobility measures has rarely been examined. The interrater agreement of STREAM was high for individual items and the total scores. Although the total score interrater agreement of the RMI and MRMI was high, at least 2 items of both measures demonstrated only fair to moderate agreement between raters. These findings indicate that the interrater reliability of STREAM is slightly higher than that of RMI and MRMI.

Some mobility measures were not selected for comparison in this study. For example, the gait speed test (eg, 10-m walking speed test and 6-minute walking distance) is commonly used to measure mobility after stroke in both clinical and research settings. However, the gait speed test is not relevant for all patients with stroke. Mobility, by nature, is complex and multifactorial, whereas the gait speed test simply reflects 1 unique and specific dimension of mobility. Furthermore, the speed test cannot be used for the patients without the ability to walk. On the other hand, the 3 mobility measures used in this study measure patients’ performance on some tasks that reflect the multifactorial nature of mobility. Furthermore, the 3 mobility measures examined in this study are feasible for assessing most stroke patients, including those with very poor mobility.

A limitation of the present study is that the intrarater reliability of the measures was not examined. We found a high intrarater reliability of the measures. Therefore, the intrarater reliability of the 3 measures might not be an issue of great concern. Another limitation is that the sample size in this study was not large enough to further analyze the data according to type or severity of stroke. Because the type of stroke or level of severity could affect the results of these measures, further studies with larger sample sizes are necessary to analyze these effects on the psychometric characteristics of the measures. Furthermore, the psychometric properties of STREAM appeared to be slightly better than those of the other 2 measures (eg, RMI showed a notable floor effect in the early stages of stroke; the score changes of RMI and MRMI at 90 to 180 days after stroke onset were not

### Table 3. Concurrent Validity of the 3 Mobility Measures at 4 Time Points After Stroke

<table>
<thead>
<tr>
<th>Time After Stroke, d</th>
<th>RMI vs MRMI</th>
<th>RMI vs STREAM</th>
<th>MRMI vs STREAM</th>
</tr>
</thead>
<tbody>
<tr>
<td>14 (n=57)</td>
<td>0.78</td>
<td>0.50</td>
<td>0.69</td>
</tr>
<tr>
<td>30 (n=54)</td>
<td>0.90</td>
<td>0.59</td>
<td>0.87</td>
</tr>
<tr>
<td>90 (n=44)</td>
<td>0.90</td>
<td>0.53</td>
<td>0.82</td>
</tr>
<tr>
<td>180 (n=43)</td>
<td>0.93</td>
<td>0.55</td>
<td>0.85</td>
</tr>
</tbody>
</table>

### Table 4. Convergent Validity and Predictive Validity of the 3 Mobility Measures at 4 Time Points After Stroke

<table>
<thead>
<tr>
<th>Time After Stroke, d</th>
<th>Convergent Validity* (p)</th>
<th>Predictive Validity† (p)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMI</td>
<td>MRMI</td>
</tr>
<tr>
<td>14 (n=57)</td>
<td>0.72</td>
<td>0.86</td>
</tr>
<tr>
<td>30 (n=54)</td>
<td>0.88</td>
<td>0.89</td>
</tr>
<tr>
<td>90 (n=44)</td>
<td>0.86</td>
<td>0.88</td>
</tr>
<tr>
<td>180 (n=43)</td>
<td>0.88</td>
<td>0.82</td>
</tr>
</tbody>
</table>

*Relationships between the total score of the mobility measures with that of the BI at 4 time points after stroke.
†Relationships between the results of the mobility measures at 3 time points (14, 30, and 90 d after stroke) and that of the BI at 180 d after stroke.

### Table 5. Responsiveness of the 3 Mobility Measures and the BI at Different Stages of Stroke Recovery

<table>
<thead>
<tr>
<th>Time After Stroke, d</th>
<th>SRM</th>
<th>Wilcoxon z</th>
</tr>
</thead>
<tbody>
<tr>
<td>14–30 (n=51)</td>
<td>1.14</td>
<td>1.31</td>
</tr>
<tr>
<td>30–90 (n=43)</td>
<td>0.86</td>
<td>0.83</td>
</tr>
<tr>
<td>90–180 (n=43)</td>
<td>0.24</td>
<td>0.20</td>
</tr>
<tr>
<td>14–90 (n=43)</td>
<td>1.67</td>
<td>1.56</td>
</tr>
<tr>
<td>14–180 (n=43)</td>
<td>1.94</td>
<td>1.53</td>
</tr>
</tbody>
</table>

*P<0.001; †P<0.05.
significant. However, the psychometric differences among the 3 measures may not be statistically significant.

In summary, the 3 mobility measures examined demonstrated acceptable levels of reliability, validity, and responsiveness among our stroke patients. The psychometric characteristics of STREAM were slightly superior to those of RMI and MRMI. We prefer and recommend STREAM for assessing mobility disability after stroke in both clinical and research settings.

Acknowledgment

This study was supported by research grants from the National Science Council (NSC 91-2314-B-002-353 and NSC 91-2314-B-002-354).

References
Comparison of Psychometric Properties of Three Mobility Measures for Patients With Stroke

I-Ping Hsueh, Chun-Hou Wang, Ching-Fan Sheu and Ching-Lin Hsieh

*Stroke*. 2003;34:1741-1745; originally published online May 29, 2003;
doi: 10.1161/01.STR.0000075295.45185.D4

*Stroke* is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2003 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/34/7/1741