Assessment of Scales of Disability and Handicap for Stroke Patients

C.D.A. Wolfe, MD; N.A. Taub, MSc; E.J. Woodrow, BA; and P.G.J. Burney, FFPHM

Background and Purpose: The purpose of the study is to compare the reliability of the Barthel activities of daily living score, which assesses disability, with the Rankin scale, which assesses handicap, and to determine their mutual agreement.

Methods: Fifty patients with stroke of varying severity were identified by a community-based stroke register and interviewed by two of three research nurses on two occasions that were 2–3 weeks apart.

Results: There was no evidence of a systematic difference between the first and second measurements. Repeatability was assessed using a \( \kappa \) statistic with quadratic disagreement weights (\( \kappa_w \)) to take account of extreme differences. This measure was very good for both Barthel (\( \kappa_w = 0.98 \)) and Rankin (\( \kappa_w = 0.95 \)) scales. There was also excellent agreement between raters for the Barthel scale (\( \kappa_r \geq 0.88 \)), but some indication of disagreement (\( \kappa_r = 0.75 \)) between raters for the Rankin scale. Analysis of variance confirmed these findings. A conversion from the Barthel to the Rankin scale can be derived by assigning the most common Rankin score for the subjects with a given Barthel score, producing a \( \kappa_w \) of 0.91 for agreement.

Conclusions: The Barthel scale is a more reliable and less subjective scale for assessing disability, from which a Rankin handicap score can then be derived to enable those managing stroke patients to assess aspects of handicap as well as disability. (Stroke 1991;22:1242-1244)
Subjects and Methods

Fifty patients diagnosed clinically as having suffered a stroke at least 3 months previously were visited at home or in hospital. The World Health Organization definition of a stroke was used to identify patients for the study. Patients were registered in a community-based stroke registry and were selected to represent a range of mild to severe strokes of all ages. The three observers were registered general nurses working as research assistants for the stroke registry. Before the study they had 6 months’ experience using the Barthel and Rankin scales on first-time stroke patients. Assessments were made by questioning the patient directly and also, where necessary, by questioning the nurses caring for them. None of the patients had received any rehabilitation in the interval between assessments. The Wilcoxon test for matched pairs was used to test whether there was a tendency for subjects’ scores to rise or fall between the two assessments.

The degree of repeatability between the 50 pairs of observations was calculated using the $\kappa$ statistic. If all degrees of disagreement are of equal importance, the coefficient of agreement is $\kappa=(p_o-p_e)/(100-p_e)$, where $p_o$ is the percent agreement observed and $p_e$ is the percent agreement expected by chance. $\kappa$ is an index of the agreement over and above that expected by chance alone; for most purposes, $\kappa$ values $>0.75$ may be taken to represent excellent agreement. Fourteen patients were assessed twice by the same observer within a 2-week time interval, to obtain replicate measurements. This time interval was chosen to minimize the effect of disease progression on the one hand and observer recall on the other. A weighted $\kappa$ statistic with quadratic disagreement weights ($\kappa_w$) was used to take account of extreme differences and has been shown to be approximately equivalent to the intraclass correlation coefficient for measures on a continuous scale.

The differences between raters’ mean scores were estimated by analysis of variance using Generalized Linear Interactive Modeling (GLIM). Initially the difference between each subject’s first and second scores was plotted against their mean, to check that these were approximately normally distributed and were independent of the severity of the subject’s condition.

Results

There was no evidence of a systematic difference between the first and second measurement of either score (Wilcoxon test, $p>0.1$). The Barthel score ranged from 0–4 (16%), 5–9 (21%), 10–14 (27%), 15–19 (23%), and 20 (13%). The Rankin score took the values 0 (3%), 1 (3%), 2 (16%), 3 (21%), 4 (29%), and 5 (28%).

The $\kappa$ statistic for all subjects and for each pair of raters is shown in Table 1. The percentage exact agreement between raters is also displayed in Table 1.

The high values of $\kappa$ for the subjects measured twice by the same rater, $\kappa_w=0.98$ for Barthel and $\kappa_w=0.95$ for Rankin, indicate excellent repeatability on each scale. There is excellent agreement between raters on the Barthel scale ($\kappa_w=0.88$ for each pair of raters). There is some indication of disagreement between raters on the Rankin scale, with raters 1 and 3 agreeing closely ($\kappa_w=0.96$) but differing from rater 2.

Analysis of variance using GLIM demonstrated that there was no evidence of any systematic difference between the raters’ Barthel scores ($F_{2,48}=0.25$, $p>0.5$) but there was strong evidence of a systematic difference between the raters’ Rankin scores ($F_{2,48}=6.02$, $p=0.005$).

The analysis of variance indicated that raters 1 and 3 estimated the Rankin score 0.42 and 0.33 points (estimated with standard error of 0.13) higher than rater 2.

An empirical comparison of Barthel scores with Rankin scores can be derived by assigning the most common Rankin score for these subjects with a given Barthel score. There was no Barthel score that corresponded with a Rankin score of 0 or 1. Even the subjects with a Barthel score of 20 were most likely to have a Rankin score of 2 (of 13). Therefore Rankin scores 0, 1, and 2 were grouped together. The

### Table 1. Weighted $\kappa$ ($\kappa_w$) for Repeatability and Rater Agreement

<table>
<thead>
<tr>
<th></th>
<th>Barthel scale</th>
<th>Rankin scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\kappa$</td>
<td>95% CI</td>
<td>$%$ exact agreement</td>
</tr>
<tr>
<td>Raters 1 and 2</td>
<td>11</td>
<td>0.88</td>
</tr>
<tr>
<td>Raters 1 and 3</td>
<td>15</td>
<td>0.94</td>
</tr>
<tr>
<td>Raters 2 and 3</td>
<td>10</td>
<td>0.98</td>
</tr>
<tr>
<td>Replicates</td>
<td>14</td>
<td>0.98</td>
</tr>
<tr>
<td>All subjects</td>
<td>50</td>
<td>0.96</td>
</tr>
</tbody>
</table>

CI, confidence interval.
The King's Fund consensus document \textsuperscript{18} advocates a multidisciplinary approach to stroke care that would require reliable, standardized measures of disability and handicap to be used by all professionals; this study has assessed the use by research nurses of two commonly used scales. There was no evidence that the scores changed systematically between the two measurements, making a 2–3 week interval between patient assessments legitimate in this study. The results also indicated a high degree of repeatability, indicating that the measurements are not subject to random fluctuation with time.

The $\kappa$ statistics for agreement for interobserver variation were excellent and in agreement with previous studies.\textsuperscript{8,9} If such scales are to be used by members of a team, the Barthel interobserver variation is excellent, but there was some disagreement between rater 2 and the others for the Rankin score such that rater 2 scores approximately 0.4 less than raters 1 and 3 on the Rankin scale. Although this is not as good as the Barthel agreement, this magnitude of difference is unlikely to be clinically relevant. The percentage exact agreement for the Barthel scale was lower because of the subjective nature of the score. The percentage exact agreement for the Rankin scale (38 versus 80) because the former is a 20-point scale and the latter only a 6-point scale. As van Swieten\textsuperscript{9} has discussed, the Rankin score is subjectively assessed, instead of being scored on stated criteria, and misclassification is therefore possible. Rater 2 had a tendency to classify patients lower because of incorporation of asymptomatic comorbidities into the score. A suggestion of reducing the Rankin scale to 3 or 4 points to improve interobserver variation\textsuperscript{9} would be supported by our observations. The use of an ADL checklist for scoring the Rankin scale would overcome, in part, the subjective nature of the score.

Although both scales have been used in studies, the nature of the relation between them has not been addressed. Previous comparisons of ADL scores have shown them to have unweighted agreements of $\kappa=0.77$ and $\kappa=0.42$.\textsuperscript{8} To allow comparison, groups 0, 1, and 2 on the Rankin scale were combined, making it a 4-point scale; the unweighted $\kappa$ was 0.72 and $\kappa$ was 0.91, which indicate very substantial agreement. The close association between disability and handicap in this group of patients allows the conversion from one score to the other. In view of the better interobserver and intraobserver agreement for the Barthel scale, it is suggested that a Barthel score be assessed and a Rankin score derived to assess handicap. This would be in agreement with research groups in the United Kingdom who have tended to use the Barthel score as the most complete, sensitive, and widely used scale.

**Acknowledgments**

We thank Mrs. Richardson for her fieldwork, Ms. Law for statistical advice, and Ana Childs for manuscript preparation.

**References**


**KEY WORDS** • stroke assessment • stroke outcome
Assessment of scales of disability and handicap for stroke patients.
C D Wolfe, N A Taub, E J Woodrow and P G Burney

Stroke. 1991;22:1242-1244
doi: 10.1161/01.STR.22.10.1242

Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 1991 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/22/10/1242

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/