Validity and Reliability of the Barthel Index Administered by Telephone

Gian Luca Della Pietra, RN; Katia Savio, MD; Elodie Oddone, RN; Monica Reggiani, MD; Francesco Monaco, MD; Maurizio A. Leone, MD

Background and Purpose—We aimed to evaluate validity and reliability of the Barthel Index administered telephonically compared with face-to-face assessment in clinically stable patients with stroke.

Methods—One hundred thirty-one patients were interviewed twice by 2 registered nurses with identical training. Half of the patients were randomized to receive the telephone interview followed by the face-to-face interview and half the contrary. The sequence of interviewers was randomized.

Results—The median value of the Barthel Index score was 30 (first to third interquartile range, 15 to 80) by telephone and 35 (15 to 75) by face-to-face ($P=0.29$). The weighted $\kappa$ was 0.90 (95% CI, 0.85 to 0.94); $\kappa$ values ranged from 0.70 (0.58 to 0.82) for bowel control to 0.91 (0.83 to 0.99) for bathing.

Conclusions—Telephone assessment of stroke disability with the Barthel Index is reliable in comparison to direct face-to-face assessment. (Stroke. 2011;42:00-00.)

Key Words: Barthel Index ■ interview ■ reliability ■ stroke ■ telephone ■ validity

The Barthel Index (BI) is an ordinal scale for the functional assessment of disability that has been widely used in stroke outcome research.1 Telephone assessment of the BI has been evaluated in only 3 studies,2–4 showing a high agreement between telephone and face-to-face (f-to-f) assessments. We aimed to evaluate the validity and reliability of the BI administered by telephone compared with f-to-f assessment in clinically stable patients with stroke.

Methods

We assessed 157 patients with stroke consecutively admitted to the Department of Neurology, “Maggiore della Carità” Hospital of Novara—a first-referral hospital in Northern Italy—during a 9-month period. The study was approved by the Hospital’s Ethical Committee. Inclusion criteria were diagnosis of stroke, age $\geq$18 years, clinical stability, and written informed consent. Informed consent was obtained from the patient or from the proxies when the patient was not able to give consent. The National Institutes of Health Stroke Scale was scored every day by a neurologist blind to the BI score. Clinical stability was defined as nonworsening of the score for 3 consecutive days. Each patient was interviewed twice by 2 registered nurses with identical training in the use of stroke scales. Half of the patients were randomized to receive the telephone interview followed by the f-to-f interview and half the contrary. The sequence of interviewers was randomized, the randomization list was concealed, and the 2 nurses were blind to each other’s scores. The 2 interviews were administered with a 2-day interval regarded as long enough to ensure that the first responses would be forgotten and short enough to ensure that the clinical condition would not change. The caregivers of patients unable to be interviewed by telephone were interviewed as proxy respondents.

F-to-f BI was obtained by observation of patients, whereas bladder and bowel management was assessed from the subject’s history.1 Telephone interviews were unstructured; operators were free to ask what they thought important to understand the scores for the individual items. Telephone calls were made from a room outside the ward to the patient’s room telephone.

Data were analyzed with SAS. The $\kappa$ or weighted $\kappa$ ($wK$) statistics5 were used to evaluate agreement (by quintiles of the BI score). Student $t$ test, Wilcoxon signed-rank test, and Wilcoxon rank-sum test were used where appropriate.

Results

Nineteen patients died before reaching clinical stability. Seven (mean age, 76.6 years; SD, 9.9) refused to give consent. We investigated 131 patients with a mean age of 73.9 years (SD, 13.3). Nineteen patients were diagnosed with hemorrhagic stroke and 112 ischemic. The mean interval

<table>
<thead>
<tr>
<th>Barthel Index Cutoff</th>
<th>$&lt;$20</th>
<th>$\geq$60</th>
<th>$\geq$80</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sensitivity</td>
<td>88%</td>
<td>96%</td>
<td>100%</td>
</tr>
<tr>
<td>Specificity</td>
<td>94%</td>
<td>98%</td>
<td>97%</td>
</tr>
<tr>
<td>Positive predictive value</td>
<td>88%</td>
<td>96%</td>
<td>91%</td>
</tr>
<tr>
<td>Negative predictive value</td>
<td>94%</td>
<td>98%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Received January 10, 2011; accepted January 28, 2011.
From the SCDU Neurologia, AOU Maggiore della Carità and Università del Piemonte Orientale “A. Avogadro,” Novara, Italy.
Correspondence to Maurizio A. Leone, MD, SCDU Neurologia, Ospedale Maggiore della Carità, C. Mazzini 18-28100 Novara, Italy. E-mail maurizio.leone@maggioreosp.novara.it
© 2011 American Heart Association, Inc.
Stroke is available at http://stroke.ahajournals.org DOI: 10.1161/STROKEAHA.111.613521
from stroke onset to scoring was 7.5 days (SD, 5.2). Interview with caregivers was needed for 45 patients (34.4%).

Validity
The median value of the BI score was 30 (first to third interquartile range, 15 to 80) by telephone and 35 (interquartile range, 10 to 80) by f-to-f ($P=0.29$). The median score of the telephone BI was not statistically different ($P=0.88$) when this modality came first (N=68; median score, 35; interquartile range, 10 to 80) or second (63, 30, 15 to 85). The median score of the f-to-f BI was not statistically different ($P=0.76$) when this modality came first (30, 15 to 85) or second (35, 15 to 75). We first investigated the validity of BI administration by telephone using f-to-f administration as the gold standard (Table 1). The frequency distribution of the scores assigned in telephone and f-to-f interviews was similar (Figure).

Reliability
The agreement between the 2 methods was excellent; wK was 0.90 (95% CI, 0.85 to 0.94) for all patients. Agreement was lower for the age class 65 to 74 years (N=31; wK 0.78; 0.62 to 0.93) than for 75 to 84 years (43; 0.87; 0.79 to 0.96), 85+ years (29; 0.87; 0.74 to 1.0), or <65 years (28; 0.93; 0.85 to 1.0). Agreement was excellent both when the telephone interview came first (0.89; 0.83 to 0.95) or second (0.90; 0.84 to 0.96). Kappa values for each item are reported in Table 2.

The National Institutes of Health Stroke Scale score was grouped in 2 categories: agreement was excellent for National Institutes of Health Stroke Scale score $\leq 8$ (N=81; wK 0.92; CI, 0.84 to 1.0) and only moderate for score $>8$ (50; 0.56; 0.31 to 0.84). The wK for the 86 self-respondent was 0.90 (0.85 to 0.95), whereas it was 0.67 (0.48 to 0.86) in the 45 patients interviewed with the caregiver.

Discussion
This study shows that telephone assessment of stroke disability with the BI is reliable in comparison to f-to-f assessment. We found excellent agreement between the 2 methods with a wK of 0.90, like in another 3 studies. We found excellent agreement in each single item of the BI

Table 2. Agreement for Each Item of the Barthel Index Between the Telephone and Face-to-Face Interview

<table>
<thead>
<tr>
<th>Face-to-Face</th>
<th>Telephone</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
</tr>
</tbody>
</table>

Kappa* (95% CI)

Feeding 0.80 (0.72–0.89)
Bathing 0.91 (0.83–0.99)
Grooming 0.86 (0.77–0.95)
Dressing 0.78 (0.70–0.87)
Bowel 0.70 (0.58–0.82)
Bladder 0.84 (0.77–0.91)
Toilet use 0.89 (0.83–0.95)
Bed/wheelchair transfers 0.85 (0.80–0.91)
Mobility 0.92 (0.87–0.97)
Stairs 0.82 (0.73–0.91)

CI indicates confidence interval.

*Weighted $\kappa$ is shown where appropriate.
with $\kappa$ values ranging from 0.70 in the bowels to 0.91 in bathing. Bowel control had the lowest agreement also in another study.\(^3\)

We administered the BI to subjects with a wide severity range. Sensitivity ranged from 88% to 100%, indicating that telephone-administered BI is as valid as f-to-f irrespective of the cutoff. However, agreement between telephone and f-to-f assessment was lower for patients with a National Institutes of Health Stroke Scale score $\geq 8$; this may be driven by a higher number of proxy respondents who had lower agreement and raises the need for caution in interpreting telephone BI in more severe patients. Follow-up studies based on telephone BI assessment should evaluate the validity of this method in a sample of more severe patients, who may account for up to 15% of the total.\(^6\) Furthermore, we used 3 categories instead of 5 for this analysis and this could even increase the $\kappa$ value.

The major limit of our study is that results from our hospital setting might not be immediately transferred to interviews obtained at a home setting, where the BI is most often used. The identification of proxy respondents in a home environment may be more difficult than in our setting and this can lead to a poorer agreement between telephone and f-to-f assessments than in our study.

**Disclosures**

None.

**References**

Validity and Reliability of the Barthel Index Administered by Telephone
Gian Luca Della Pietra, Katia Savio, Elodie Oddone, Monica Reggiani, Francesco Monaco and
Maurizio A. Leone

Stroke. published online April 28, 2011;
Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2011 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/early/2011/04/28/STROKEAHA.111.613521