Diagnostic Accuracy of the Barthel Index for Measuring Activities of Daily Living Outcome After Ischemic Hemispheric Stroke: Does Early Poststroke Timing of Assessment Matter?
Gert Kwakkel, Janne M. Veerbeek, Barbara C. Harmeling-van der Wel, Erwin van Wegen and Boudewijn J. Kollen

Stroke 2011, 42:342-346: originally published online December 23, 2010
doi: 10.1161/STROKEAHA.110.599035
Stroke is published by the American Heart Association. 7272 Greenville Avenue, Dallas, TX 72514
Copyright © 2010 American Heart Association. 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/42/2/342

Subscriptions: Information about subscribing to Stroke is online at
http://stroke.ahajournals.org/subscriptions/

Permissions: Permissions & Rights Desk, Lippincott Williams & Wilkins, a division of Wolters Kluwer Health, 351 West Camden Street, Baltimore, MD 21202-2436. Phone: 410-528-4050. Fax: 410-528-8550. E-mail: journalpermissions@lww.com

Reprints: Information about reprints can be found online at
http://www.lww.com/reprints
Diagnostic Accuracy of the Barthel Index for Measuring Activities of Daily Living Outcome After Ischemic Hemispheric Stroke

Does Early Poststroke Timing of Assessment Matter?

Gert Kwakkel, PhD; Janne M. Veerbeek, MSc; Barbara C. Harmeling-van der Wel; Erwin van Wegen, PhD; Boudewijn J. Kollen, PhD; on behalf of the Early Prediction of functional Outcome after Stroke (EPOS) Investigators

Background and Purpose—This study investigated the diagnostic accuracy of the Barthel Index (BI) in 206 stroke patients, measured within 72 hours, for activities of daily living at 6 months and determined whether the timing of BI assessment during the first days affects the accuracy of predicting activities of daily living outcome at 6 months.

Methods—Receiver operating characteristic curves were constructed to determine the area under the curve and optimal cutoff points for BI at Days 2, 5, and 9. OR, sensitivity, specificity, positive predictive value, and negative predictive value were calculated to predict BI ≥19.

Results—The area under the curve ranged from 0.785 on Day 2 to 0.837 and 0.848 on Days 5 and 9. Comparison of the receiver operating characteristic curves showed that the area under the curve was significantly different between Days 2 and 5 (P<0.001) and between Days 2 and 9 (P<0.001). No significant difference was found between Days 5 and 9 (P=0.08). Using a BI cutoff score of 7, the positive predictive value gradually increased from 0.696 on Day 2 to 0.817 on Day 5 and 0.864 on Day 9, whereas negative predictive value declined from 0.778 on Day 2 to 0.613 on Day 9.

Conclusions—Assessment of the BI early poststroke showed good discriminative properties for final outcome of BI at 6 months. However, Day 5 proved to be the earliest time for making an optimal prediction of final outcome of activities of daily living. The BI should be measured at the end of the first week in hospital-based stroke units for early rehabilitation management. (Stroke. 2011;42:342-346.)

Key Words: ADL ■ Barthel Index ■ stroke prognosis ■ stroke units

A number of prospective epidemiological studies in the Western countries found that approximately 60% of all stroke victims will regain independency in basic activities of daily living (ADL) within 6 months poststroke.1 According to the American Heart Association, approximately 14% of these stroke survivors achieve full recovery in their basic ADLs, between 25% and 50% require at least some assistance in ADLs, and approximately half experience severe long-term dependency.2 In particular, forced by rising healthcare costs, there is a growing need for early accurate prediction of outcome after stroke to (1) set realistic and attainable treatment goals; (2) inform clients and their relatives properly; (3) facilitate discharge planning; and (4) anticipate possible consequences such as implementing home adjustments and address the need for community support. Unfortunately, there is no consensus on which measurements should be used in stroke units nor about the most appropriate poststroke timing to perform these assessments.3,4 A commonly used measurement tool to assess ADL independency in stroke units is the Barthel Index (BI).5 The BI scale measures patients’ actual performance in basic ADLs by inquiry and/or observation and contains 10 items, which are scored using arbitrary

Received August 6, 2010; final revision received September 15, 2010; accepted September 27, 2010.

From the Department of Rehabilitation Medicine (G.K., J.M.V., E.v.W.), Research Institute MOVE, VU University Medical Center, Amsterdam, The Netherlands; the Department of Rehabilitation Medicine (G.K.), Rudolf Magnus Institute of Neuroscience, University Medical Centre Utrecht, Utrecht, The Netherlands; the Department of Rehabilitation Medicine and Physical Therapy (B.C.H.v.d.W.), Erasmus MC, University Medical Centre, Rotterdam, The Netherlands; and the Department of General Practice (B.J.K.), University Medical Centre Groningen, University of Groningen, The Netherlands. EPOS Investigators: Executive Committee: Gert Kwakkel, PhD, Principal Investigator, Department of Rehabilitation Medicine, VU University Medical Centre Amsterdam, and Co-Principal Investigator: Barbara C. Harmeling-van der Wel, Department of Rehabilitation Medicine & Physical Therapy, Erasmus MC, University Medical Center Rotterdam. Steering Committee and Data Management: Janne M. Veerbeek, MSc, and Rinske Nijland, MSc, Department of Rehabilitation Medicine, VU University Medical Center, Amsterdam, and Co-Principal Investigator: Erwin van Wegen, PhD; on behalf of the Early Prediction of functional Outcome after Stroke (EPOS) Investigators. DOI: 10.1161/STROKEAHA.110.599035

Stroke is available at http://stroke.ahajournals.org

© 2011 American Heart Association, Inc.
weights (5, 10, or 15) to arrive at a total scale range of 0 to 100 or alternatively uses 0, 1, 2, or 3 weighted item scores on a 0 to 20 scale.6–8 The instrument is easy to administer, does not need formal training or certificate programs,5 and the 0 to 20 scale version has been shown to be reliable6 and concurrently valid when compared with the motor part of the Functional Independence Measure6 and the modified Rankin Scale (mRS).4,9–11 Finally, the BI has demonstrated excellent discriminative properties in organized inpatient trials.12 Moreover, a number of prospective studies have shown that the severity of disability according to the BI recorded at 5 days poststroke, even when dichotomized,13 shows a highly prognostic accuracy for death13 or dependency as a final outcome.1,13–15 As a consequence, the BI has been recommended to be used for the development of predictive risk models to estimate final outcome for those patients who were lost in trials.10

Despite the growing consensus that the BI should be implemented as a standardized tool of measuring disability in acute (multicenter) trials11 and should be used preferably in a repetitive way to assess improvement in patients over time,3 there is little consensus about the optimal timing for assessing the BI in hospital-based stroke units as a tool for monitoring severity of disability and to predict the final outcome of ADLs after stroke. Moreover, in prospective longitudinal studies, the optimal timing of assessment early poststroke is an important factor that determines the accuracy of prediction.16

The first objective of the present study was to investigate the predictive value of BI measured within 72 hours for outcome of basic ADLs assessed at 6 months poststroke. The second aim was to determine the optimal poststroke timing of BI assessment in hospital stroke units for the most accurate prediction of final outcome of ADLs at 6 months poststroke.

Subjects and Methods

Design

The Early Prediction of functional Outcome after Stroke (EPOS) study is a prospective cohort study that applies an intensive repeated-measurements design starting within 72 hours after stroke onset. The diagnosis of stroke was based on the definition by the World Health Organization. Two hundred forty-six patients were recruited for the EPOS study in 34 months. Patients were recruited from 9 hospital-based stroke units in The Netherlands (ie, Erasmus MC Rotterdam; UMC Utrecht; VU University Medical Center Amsterdam; AMC Amsterdam; UMC St Radboud Nijmegen; LUMC Leiden; Amphia Hospital Breda; Franciscus Hospital Roosendaal; and Diaconessen Hospital Leiden). The EPOS test battery was applied within 72 hours after the start of the EPOS study. Despite their experience, all assessors in participating hospitals and nursing homes were trained to implement the EPOS test battery in a 1-day (8-hour) course.

Independent Variables

Baseline characteristics were assessed 72 hours after stroke and included: (1) age; (2) gender; (3) type of stroke (according to the Bamford Oxford Community Stroke Project classification)19; (4) number of days between the stroke and first assessment; (5) left or right hemispheric stroke; (6) urinary incontinence (bladder item of mRS); (10) the 15-item version of the National Institutes of Health Stroke Scale21,22; and (11) the BI measured within 72 hours on Days 5 and 9 poststroke. Finally, Functional Ambulation Categories23 and the Action Research Arm test24 are presented descriptively at 6 months poststroke.

Procedure

The research protocol started within 72 hours after stroke onset, and follow-up assessments took place after 5 and 9 days at each hospital stroke unit. The final measurement, scheduled at 6 months poststroke, was performed at the residence (ie, home, nursing home, or rehabilitation center) or at the outpatient clinic of the hospital. All assessments in each stroke unit had familiarized themselves with the EPOS test battery before the start of the EPOS study. Despite their experience, all assessors in participating hospitals and nursing homes were trained to implement the EPOS test battery in a 1-day (8-hour) course.

Data Analyses

On the basis of sensitivity/1-specificity and maximum area (AUC) under the receiver operating characteristic (ROC) curve, the optimal dichotomization of BI was estimated for each assessment day to predict the dichotomized (ie, ≥19 points) outcome of BI at 6 months. In case of a missing value at the second or third assessment, the last observation was carried forward. The ROC curves of 3 models were graphically displayed and tested to assess if the AUC at Day 2 was significantly different from that of Days 5 and 9 and if AUC of Day 5 differed significantly from Day 9. For each comparison between 2 ROC curves, a z-statistic was calculated by the equation: 
\[
z = \frac{AUC_1 - AUC_2}{\sqrt{SE_1^2 + SE_2^2 - 2SE_1SE_2 r}},
\]
with r representing the Pearson product moment correlation coefficient between the 2 models.25 The calculated z-statistic was evaluated to be significant if \(z \geq 1.96\). Subsequently, on the basis of the optimal cutoff score for the first BI measurement <72 hours poststroke, a bivariate logistic regression analysis was performed between initial BI and ADL independency on the BI (ie, ≥19 points) at 6 months to estimate the OR with 95% CIs. The same analysis was repeated for the data.
collected at 5 and 9 days poststroke. Finally, 2-way contingency tables were used to calculate sensitivity, specificity, and positive and negative predictive values, including their 95% CIs, for each model within 72 hours poststroke and on Days 5 and 9 poststroke. All analyses were 2-tailed using a critical probability value for significance of 0.05 and performed with SPSS Version 15.

Results
Forty of 246 patients were lost to follow-up due to death (N=23), refusal for assessment at 6 months (N=3), recurrent stroke (N=5), or other reasons (N=9). In addition, assessments of 14 patients were missing at T1 and 22 at T2. In total, 206 patients were included in analysis representing a particular segment from within the total stroke population. Table 1 presents the main characteristics of the remaining 206 patients. The candidate determinants were measured on a mean (SD) 2.18 (1.19), 5.50 (1.52), and 9.29 (4.89) days poststroke.

The average age of patients in this cohort was 66.3 (14.0) years and 95 of the patients were male. Eighty-eight subjects had a stroke in the left hemisphere. According to the Bamford classification, 96 patients were diagnosed with lacunar circulation infarcts, 42 with total anterior circulation infarcts, and 68 patients had a partial anterior circulation infarct. The median BI score on Days 2 was 7 points (interquartile range, 3 to 12), whereas the median mRS was 4 points (interquartile range, 3.75 to 5). At 6 months, BI had a median of 19 points (interquartile range, 16.75 to 20), whereas 60.7% of the 206 patients showed full independency on the BI.

The Figure shows the ROC analysis for BI scores on Days 2, 5, and 9. The AUC ranged from 0.785 for Day 2 (SE=0.035; P<0.001; 95% CI, 0.715 to 0.854), 0.837 for Day 5 (SE=0.031; P<0.001; 95% CI, 0.776 to 0.899), and 0.848 for Day 9 (SE=0.030; P<0.001; 95% CI, 0.788 to 0.908). Comparison of the 3 derived ROC curves showed that the AUC was significantly different between Day 2 and Day 5 (z=3.537, P<0.001) and between Day 2 and 9 (z=3.621, P<0.001). However, no significant difference was found between the AUC of the ROC curves of Days 5 and 9 (z=1.416, P=0.08). The optimal cutoff value, with the highest sensitivity and 1-specificity, was found when BI was dichotomized into ≤6 points (ie, severe disability) and ≥7 points (ie, moderate to mild disability).

Table 2 shows the numbers of true- and false-positives and negatives as well as the OR, sensitivity, specificity, positive predictive value, and negative predictive value calculated using a cutoff value of 7 points on BI in terms of predicting dichotomized BI outcome at 6 months poststroke. ORs based on a cutoff score of ≥19 points ranged from 8.013 (95% CI, 3.75–20.325) on Day 2 to a maximum of 10.533 (95% CI, 5.458 to 20.325) on Day 5. The positive predictive value showed a gradual increase from 0.696 (95% CI, 0.645 to 0.739) on Day 2 to 0.864 (95% CI, 0.815 to 0.905) on Day 9, whereas negative predictive value declined from 0.778 (95% CI, 0.699 to 0.844) on Day 2 to 0.613 (95% CI, 0.536 to 0.676) on Day 9. The overall accuracy for correctly predicting outcome increased from 72.8% on Day 2 to 77.2% on Day 5.

Discussion
The purpose of the present study was to determine the discriminative properties of the BI (Version 0–20) assessed at hospital-based stroke units within 72 hours poststroke for the outcome of ADL independency at 6 months. In addition, the optimal timing for early poststroke assessment of the BI to predict outcome of ADL at 6 months after stroke was explored. The present study demonstrated good discriminative properties of the BI on Days 2, 5, and 9 poststroke. However, it also suggests that the earliest, most optimal poststroke assessment is on Day 5. Assessment on Day 2 resulted in an increased number of false-negatives and consequently an underestimation of the final outcome of ADL, whereas assessment on Day 9 resulted in a relatively overes-

Table 1. Patient Characteristics Within 72 Hours After Stroke

<table>
<thead>
<tr>
<th>Patient Characteristics</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No.</td>
<td>206</td>
</tr>
<tr>
<td>Gender, female/male</td>
<td>111/95</td>
</tr>
<tr>
<td>Mean (SD) age, years</td>
<td>66.29 (14.02)</td>
</tr>
<tr>
<td>Hemisphere of stroke, left/right</td>
<td>88/118</td>
</tr>
<tr>
<td>rtPA, no. (%)</td>
<td>55 (26.7)</td>
</tr>
<tr>
<td>Mean BMI (SD)</td>
<td>26 (4.51)</td>
</tr>
<tr>
<td>Mean time interval (SD) between stroke and</td>
<td></td>
</tr>
<tr>
<td>First assessment, days</td>
<td>2.18 (1.19)</td>
</tr>
<tr>
<td>Second assessment, days</td>
<td>5.50 (1.52)</td>
</tr>
<tr>
<td>Third assessment, days</td>
<td>9.29 (4.89)</td>
</tr>
<tr>
<td>Type of stroke (Bamford)</td>
<td></td>
</tr>
<tr>
<td>LACI</td>
<td>96</td>
</tr>
<tr>
<td>PACI</td>
<td>68</td>
</tr>
<tr>
<td>TACI</td>
<td>42</td>
</tr>
<tr>
<td>NIHSS*</td>
<td>8 (4–14)</td>
</tr>
<tr>
<td>Cognitive disturbance</td>
<td></td>
</tr>
<tr>
<td>Inattention, no. (%)</td>
<td>88 (42.7)</td>
</tr>
<tr>
<td>Disorientation, no. (%)</td>
<td>50 (24.3)</td>
</tr>
<tr>
<td>Impairments of vision</td>
<td></td>
</tr>
<tr>
<td>Hemianopia, no. (%)</td>
<td>62 (30.1)</td>
</tr>
<tr>
<td>Deviation conjugue, no. (%)</td>
<td>54 (26.2)</td>
</tr>
<tr>
<td>Sensory loss, no. (%)</td>
<td>125 (60.1)</td>
</tr>
<tr>
<td>TCT (0–100)*</td>
<td>74 (36.25–100)</td>
</tr>
<tr>
<td>MI arm (0–100)*</td>
<td>39 (0–76)</td>
</tr>
<tr>
<td>MI leg (0–100)*</td>
<td>53 (22.75–76.25)</td>
</tr>
<tr>
<td>FM arm (0–66)*</td>
<td>20 (4–53)</td>
</tr>
<tr>
<td>FM leg (0–34)*</td>
<td>20.5 (9–28)</td>
</tr>
<tr>
<td>FAC (0–5)*</td>
<td>1 (0–3)</td>
</tr>
<tr>
<td>ARAT (0–57)*</td>
<td>1.5 (0–37)</td>
</tr>
<tr>
<td>BI (0–20)*</td>
<td>7 (3–12)</td>
</tr>
<tr>
<td>BI urinary incontinence, no. (%)</td>
<td>108 (52.4)</td>
</tr>
<tr>
<td>mRS*</td>
<td>4 (3.75–5.00)</td>
</tr>
</tbody>
</table>

*Median values (interquartile ranges).
†Fourteen missing values.
‡Twenty-two missing values.
No. indicates no. of patients; rtPA, recombinant tissue plasminogen activator; BMI, body mass index; LACI, lacunar cerebral infarction; PACI, partial anterior cerebral infarction; TACI, total anterior cerebral infarction; NIHSS, National Institutes of Health Stroke Scale; TCT, Trunk Control Test; MI, Motricity Index; FM, Fugl-Meyer; FAC, Functional Ambulation Categories Score; ARAT, Action Research Arm Test; BI, Barthel Index; mRS, modified Rankin Scale.
timation of the final BI at 6 months. The less optimal prediction of BI at 6 months for patients assessed within 72 hours may be caused by the instability of neurological deficits as manifested by some neurological worsening during the first 24 to 48 hours after stroke observed in approximately 25% of all patients. However, in a parallel running study focused on the timing of assessment of neurological deficits by National Institutes of Health Stroke Scale, no significant differences were found among Day 2, 5, or 9. In our opinion, a more plausible explanation could be that observers find it difficult to determine the patient’s actual performance in basic ADLs when the patient is still bedridden. As a consequence, an assessment within 72 hours poststroke will underestimate the actual patients’ performance. In line with the recommendation of Kasner, our findings suggest that, even in individuals with a minor stroke who are bedridden in the first few days after stroke, the BI will underestimate outcome scores, hence making the BI not suitable for measuring disability within the first 3 days poststroke. To the best of our knowledge, the present study is the first study that underpins the limitations of BI use within the first 3 days poststroke.

It should be noted that we selected the BI tool as recommended by the Dutch stroke guidelines for rehabilitation management and because it is the most commonly used disability scale for evaluating effectiveness of stroke units. The BI use is in line with our stroke guidelines for physical therapy and stroke management as well as the recommendations of the Agency for Health Care Policy and Research Post-Stroke Rehabilitation Panel. Both of these authorities recommend to use the BI and the motor component of the Functional Independence Measure for evaluating poststroke disability. However, knowledge about the predictive value and optimal timing of assessment in hospital stroke units is lacking in the literature for other clinical useful measurement instruments such as the Functional Independence Measure and mRS.

The present study has some limitations. First, the day of the first assessment on Day 2 was selected to conform with the Dutch stroke guidelines that recommend to mobilize patients within 72 hours poststroke onset, whereas the other 2 days of assessment (ie, Days 5 and 9) were pragmatically selected based on clinical experience. Second, our model may not be applicable to patients with brain stem strokes, hemorrhagic strokes, or recurrent strokes, which have been shown to present with different recovery profiles. This finding suggests that the model should be re-investigated for case mix and preferably crossvalidated in a holdout group. Third, it should be emphasized that the BI is an ordinal scale. In ordinal scales, the overall score is obtained by simply adding up arbitrary numeric values assigned to a subject’s ratings on a series of items. To overcome the discrete nonlinear 10 steps, we dichotomized the BI into those that achieved independency at 6 months (ie, 19 or 20 points) and those who remained dependent in basic ADLs according to the Stroke Unit Trialist Collaboration. However, there is little consensus about the use of these cutoffs in the literature. Recently, it has been shown that this cutoff corresponds with 1 point on the mRS score. Although in general, reliability of BI is considered excellent, in older individuals with cognitive impairments and when scores obtained by patient interview are compared with patient testing, reliability may not be optimal. Furthermore, currently multiple versions of BI are in circulation, which could hamper general use and valid comparison of this measure. Moreover, although the BI is not used in all countries in the world, in most European countries, the BI is considered the gold standard for measuring ADL. In addition, missing values were imputed based on the last observation carried forward. Although relatively few (approximately 1%) missing values needed to be replaced, this may have resulted in the underestimation of the discriminative properties of the BI. Finally, the BI has been extensively tested showing good psychometric properties in terms of reliability and validity. However, this instrument is known to

Table 2. Predictive Value of Dichotomized BI Assessed on Days 2, 5, and 9 Poststroke for BI Independency After 6 Months (N=206)*

<table>
<thead>
<tr>
<th>Assessment</th>
<th>True-Positives, No.</th>
<th>True-Negatives, No.</th>
<th>False-Positives, No.</th>
<th>False-Negatives, No.</th>
<th>OR† (95% CI)</th>
<th>Accuracy</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>PPV (95% CI)</th>
<th>NPV (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Day 2</td>
<td>63</td>
<td>38</td>
<td>18</td>
<td>87</td>
<td>8.013 (4.192–15.316)</td>
<td>0.728</td>
<td>0.829 (0.768–0.879)</td>
<td>0.624 (0.560–0.677)</td>
<td>0.696 (0.645–0.738)</td>
<td>0.778 (0.699–0.844)</td>
</tr>
<tr>
<td>Day 5</td>
<td>57</td>
<td>23</td>
<td>24</td>
<td>102</td>
<td>10.533 (5.458–20.325)</td>
<td>0.772</td>
<td>0.810 (0.760–0.852)</td>
<td>0.713 (0.634–0.779)</td>
<td>0.816 (0.766–0.859)</td>
<td>0.704 (0.626–0.769)</td>
</tr>
<tr>
<td>Day 9</td>
<td>49</td>
<td>17</td>
<td>31</td>
<td>108</td>
<td>10.042 (5.082–29.842)</td>
<td>0.766</td>
<td>0.777 (0.733–0.813)</td>
<td>0.742 (0.650–0.819)</td>
<td>0.864 (0.815–0.905)</td>
<td>0.613 (0.536–0.676)</td>
</tr>
</tbody>
</table>

*BI is 0 to 20 points.
†Using a cutoff of 7 points on BI.
PPV indicates positive predictive value; NPV, negative predictive value.
be insensitive to small changes in functional status, suffers from ceiling effects,13-35 and allows for the use of compensation strategies when the nonparetic arm is used for grooming and eating.35-37

Sources of Funding
This study was part of the Early Prediction of Functional Outcome after stroke (EPOS) research project funded by the “Wetsenschappelijk College Fysiotherapie” (WCF: number 33368) of the Royal Dutch Society for Physical Therapy (KNGF), the Netherlands and by ZON-MW (Zorg Onderzoek Nederland grant 8900001) as a part of the EXplaining PLasticity after stroke (EXPLICIT)-stroke program (www.explicit-stroke.nl), EXPLICIT is registered at The Netherlands Trial Register (NTR, www.trialregister.nl., TC 1424).

Disclosures
None.

References