Comparison of 2 Extended Activities of Daily Living Scales With the Barthel Index and Predictors of Their Outcomes
Cohort Study Within the South London Stroke Register (SLSR)

Shah-Jalal Sarker, PhD; Anthony G. Rudd, MD; Abdel Douiri, PhD; Charles D.A. Wolfe, MD

Background and Purpose — Basic activities of daily living measures are often supplemented by extended activities of daily living. We compared the Frenchay Activities Index (FAI) and Nottingham Extended Activities of Daily Living (NEADL) with the Barthel Index (BI) in terms of distribution of scores, concurrent validity, reliability, and their agreement and investigated the predictors of scales outcomes.

Methods — Two hundred thirty-eight patients from the population-based South London Stroke Register were assessed with the BI, FAI, and NEADL 3 months after a first-ever stroke. The pairwise relationship was studied using correlations, fractional polynomial regression, and Bland and Altman plot; the baseline predictors, for example, sociodemography, case severity: National Institutes of Health Stroke Scale, and 7-day Abbreviated Memory Test, comorbidities, and acute treatments by negative binomial regression.

Results — The BI was highly affected by a ceiling effect (33% had the highest score), FAI was only affected by floor effect (19%), but NEADL was symmetrical with only 4% highest and lowest score. Despite high concurrent validity of the scales (r = 0.80, P < 0.001), they agreed poorly only for the highest and the lowest level of activities. The association and agreement of NEADL with BI was higher than that of FAI with BI. Severe stroke patients (National Institutes of Health Stroke Scale > 13) had 28% lower BI (79% lower FAI and 62% lower NEADL) score than nonsevere patients (P ≤ 0.001). Cognitively intact patients (Abbreviated Memory Test: 8–10) had 2.3 times greater FAI values (65% higher NEADL) compared with impaired patients (P < 0.001).

Conclusions — The NEADL scale was symmetrical, concurrently valid with no floor and ceiling effects. It corresponded better with BI than FAI did confirming its basic activities of daily living properties, yet it is a more sensitive tool for extended activities of daily living without the floor and ceiling effects. Future functional status could be predicted by the acute stage National Institutes of Health Stroke Scale score, whereas only extended activities of daily living status could be predicted by the Abbreviated Memory Test score. Predicting future functional status at the acute stage may decrease unnecessary length of stay in acute care settings. (Stroke. 2012;43:1362-1369.)

Key Words: agreement ■ Barthel Index ■ Frenchay Activities Index ■ Nottingham Extended Activities of Daily Living ■ NIH Stroke Scale ■ stroke register

The Barthel Index (BI) is a commonly used measure of activities of daily living (ADL) for patients with stroke.1,2 It covers basic self-care activities and is known to have a marked ceiling effect (the percentage of subjects with the maximum possible score) and also a floor effect (the percentage of subjects with the minimum possible score) depending on the time of poststroke measurement.1 The use of an instrument that assesses only basic ADL might give a restricted view of a patient’s functional status.3 It is often necessary to supplement basic ADL measurements with extended activities of daily living (EADL) measures to also assess more complex activities in the home environment.3,4 There are no universally acknowledged measures regarding EADL. The Frenchay Activities Index (FAI) is a frequently used EADL scale for measuring stroke outcome.5 Another proposed EADL scale...
is the Nottingham Extended Activities of Daily Living (NEADL) scale.5,6

A useful measure should be reliable, valid, and responsive and should not contain floor or ceiling effects.1 Only few comparisons of psychometric properties (mainly responsiveness) of the FAI have been made with those of the BI in the literature.3,7,8 It has been reported that the FAI supplements the BI with minimal overlap in content.3 The use of the BI in the subacute phase of stroke was recommended, whereas the FAI was recommended in the longer term.8 However, most of the previous studies were performed on selected groups of patients such as recruited from a hospital or rehabilitation setting, possibly limiting the external validity of their findings.3,8 Unlike the FAI, the properties of the NEADL scale have not been studied and compared with those of the BI in the literature and to the best of our knowledge, no studies have been reported assessing the agreement of FAI and NEADL with BI. Additionally, little is known about baseline variables affecting poststroke BI, FAI, and NEADL scores.7

The aims of this present study are to (1) compare the FAI and NEADL scales with the BI in terms of distribution of scores, concurrent validity, reliability, and their agreement; and (2) investigate for common baseline predictors of poststroke scales outcomes. Data were used from the population-based multiethnic South London Stroke Register (SLSR).

Methods

A population-based stroke register, recording first-ever strokes in patients of all age groups for a defined area of South London, was started in January 1995 with follow-up interviews at 3 months, 1 year, and annually thereafter.9,10 To maximize case ascertainment, 16 different overlapping sources were used and standardized methods for ensuring completeness of case ascertainment were established. The detailed methods of notification of patients and data collection have been described elsewhere.11 For this study, consecutive patients who had their stroke from August 2002 to October 2004 were selected because data on NEADL were available only for this period. Follow-up information up to 3 months was used in this study.

The World Health Organization definition of stroke was used. Classification of the pathological subtype (cerebral infarction, primary intracerebral hemorrhage, and subarachnoid hemorrhage) was based on results from at least 1 of the following: brain imaging within 30 days of stroke onset (CT or MRI), cerebrospinal fluid analysis, or postmortem examination; 94% of the patients received their CT scan within 7 days of their stroke.

Information collected at initial assessment included self-definition of ethnic origin (1991 census question). Ethnic origin was stratified of ethnic origin (1991 census question). Ethnic origin was stratified into black, white, and other. Stroke severity (time of maximum impairment) was assessed at acute stage/baseline with the National Institutes of Health Stroke Scale (NIHSS; categorized as: severe, >13; moderate/mild, ≤13).12 Comorbidities included atrial fibrillation, diabetes, and hypertension (>140/90 mm Hg).13 Prestroke disability was measured by prior BI (categorized as: independent, BI=20; dependent, BI <20). Cognitive impairment was measured using 7-days Abbreviated Memory Test (AMT; categorized as: cognitively intact, 8–10 and cognitively impaired, 0–7).14 The service provision variable categorized patients into those who were treated in a stroke unit at any time point, in a general medical ward only, and those who were not admitted to a hospital.15 Three months after the stroke, survivors were interviewed by trained researchers to assess the BI, FAI, and NEADL scales.

Statistical Methods

The BI, FAI, and NEADL have different ranges: BI from 0 to 20, FAI from 0 to 45, and NEADL from 0 to 66 hampering their comparison. Therefore, the scores from each scale were linearly transformed to a 0 to 100 range using the formula: 100×(observed score−minimum possible score)/score range. Distributions of total outcome scale scores were displayed using notched box-and-whisker plot. Concurrent validity was measured using Spearman rank correlation coefficient. The intraclass correlation coefficient (ICC), the most suitable and most commonly used reliability parameter for continuous measures, was used (based on 2-way random effects model with absolute agreement) to measure reliability.16 Concordance correlation coefficient, a methodologically better measure than ICC because it takes into account of the precision (how far each observation deviates from the best-fit line) and accuracy (how far the best-fit line deviates from the 45° line through the origin), was also used for comparison purposes.17 Another useful measure coefficient of variation for duplicate measurements was used to compare the scales pairwise variation.18 The relationship between each pair of scales was studied using scatterplot and fractional polynomial with 95% CI.19 Fractional polynomials of degree 2, 3, and 1 were fitted for the pairs FAI and BI, NEADL and BI, and NEADL and FAI, respectively, based on the results of deviance tests and relative probability values were used to decide the final degree. The agreement between each pair of scales was studied using Bland and Altman plots.20 Also a folder empirical cumulative distribution plot/mountain plot (created by computing a percentile for each ranked difference between a new method and a reference method) was drawn as a useful complementary plot to the Bland and Altman plot to compare different distributions easily.21

Due to overdispersion, multivariable negative binomial regression models were used for identifying predictors of the 3 outcome scales based on age, ethnicity, comorbidities, prestroke BI, subtype, NIHSS, AMT, and stroke unit care using untransformed (original) outcome scales scores as regression is dependent on change of scale.22 Sex was left out of the model because it did not have any significant effect in this analysis and also in our previous analysis of the register data.15,23 Also, for some predictors, categories with few cases were merged together to increase the validity of the model.

Characteristics of dropouts were compared with the included patients to assess the generalizability of our results and also sensitivity analysis was done to assess selection bias in multivariable analysis results due to complete case analysis. Probability values <0.05 were defined as significant based on a 2-tailed test. All calculations were carried out using the statistical software package Intercooled STATA 10.1 (Stata Corp, College Station, TX) for Windows except mountain plot, notched box-and-whisker plot, coefficient of variation, and correlations that were done using MedCalc software for Windows Version 12 (MedCalc Software, Mariakerke, Belgium).

Ethics

Ethical approval was obtained from the ethics committee of Guy’s and St Thomas’ Hospital Trust, King’s College Hospital, and Westminster Hospital (London).

Results

Among the 395 patients with stroke with a first-ever stroke, 64 patients died before a 3-month poststroke interview. Finally, a total of 238 patients were considered for the analysis after further excluding 93 patients who had missing information on FAI, NEADL, or BI mainly due to loss to follow-up.

Characteristics of Analyzed Patients

Of the 238 patients, 205 had ischemic strokes, 20 primary intracerebral hemorrhage, 9 subarachnoid hemorrhage, and 4 unclassified. The sociodemographic and prestroke risk factors
of the sample are shown in Table 1. Overall 162 (69%) of the patients were white, 55 (23%) were black, and 21 (8%) were from other ethnic groups or missing ethnicity. The mean age was 69 years (SD, 14 years). Seventy-four percent of the patients had at least 1 present comorbid disease, whereas 26% did not have any. Eighty-one percent of the patients had no abnormalities in their cognitive tests (AMT: 8–10). Seventy-seven percent of the patients were treated on a stroke unit, whereas 16% were treated on general medical wards and 7% patients were not admitted to hospitals.

Distribution of Outcome Scale Scores
At the 3-month follow-up, a significantly large number (19%) of patients had a minimum value in FAI (floor effect) as compared with only 2% in BI and 4% in NEADL (P<0.001), whereas significantly more patients (33%) had maximum BI values (ceiling effect) compared with the proportion of patients with maximum NEADL (4%) or FAI (0%; P<0.001; Figure 1). Figure 1 also shows that median transformed total scores (interquartile range) for FAI, BI, and NEADL were, respectively 24 (4–53), 90 (65–100), and 49 (16–81) with 100 the maximum score achievable. Hence, the distribution of FAI was positively skewed (skewness= 0.54), whereas the distribution of BI was negatively skewed (skewness = −1.38) and the distribution of NEADL was symmetrical.

The Spearman rank correlation coefficient showed that the 3 scales were highly significantly concurrently valid with respect to each other (r_s=0.80, P<0.001; Table 2). However, ICC showed lower pair wise correlations (reliability) and only correlation between NEADL and FAI (r_p=0.75; CI, 0.66–0.91) was significant in this case. Concordance correlation coefficient gave almost identical values as of ICC but more precise. The correlations were also significantly different from each other like the Spearman rank correlation. All 3 measures showed that the correlation between NEADL and BI was higher than that between FAI and BI. This is consistent with the lower coefficient of variation between NEADL and BI than that of between FAI and BI (Table 2).

As expected, the correlation between the 2 EADL scales, FAI and NEADL, was the highest.

Agreement Between the Outcome Scale Scores
Results of the deviance test including model R^2 for the fitted fractional polynomial models in Figures 2 and 3 are given in online-only Data Supplement Table I (http://stroke.ahajournals.org). There was a nonlinear relationship between FAI and BI (Figure 2A) as well as between NEADL and BI (Figure 2C). Only very few points lied along the line of equality suggesting a poor agreement of FAI and NEADL with the BI. Bland-Altman plot was V-shaped in both pairs suggesting that the best agreement occurred for patients with a low or high level of ADL. However, scatterplots showed that for those measured with high BI, both FAI and NEADL had huge ranges; similarly for low FAI or NEADL, there were huge ranges of BI. BI measured higher ADL compared with the FAI and NEADL, particularly for average active patients. The mean difference between FAI and BI was −47 (Figure 2B) and between NEADL and BI was −29 (Figure 2D) suggesting that the agreement between NEADL and BI was higher than that between FAI and BI.

The relation between NEADL and FAI appeared to be quadratic (Figure 3A). The agreement between these 2 scales was fair. NEADL estimated a higher activity than FAI, particularly for average active patients (Figure 3B). The mean difference (NEADL–FAI) was 18 and the 95% limits of agreement (−11 and
48) were the smallest of all the comparisons indicating the highest agreement between these 2 scales. The inverse V-shaped Bland-Altman plot (Figure 3B) suggested the best agreement occurred for patients with a low or high level of ADL.

Figure 3C clearly shows that NEADL agrees more with the BI than FAI does because the fitted fractional polynomial line for the NEADL is closer to the line of equality. In the folded empirical cumulative distribution plot (Figure 3D), the mountains are not centered over 0 and hence the 2 scales, NEADL and FAI, were not unbiased with respect to BI. The differences of BI with NEADL tend to be smaller (median difference, 29) than the differences of BI with FAI (median difference, 49). Therefore, NEADL corresponds better with the BI than does FAI.

**Predictors for the Functional Outcomes**

When looking at predictors for the 3 scales outcomes, the NIHSS score was a significant predictor for all 3 scales at the 3-month follow-up (Table 3). A patient with severe stroke...
(NIHSS >13) had a 28% lower BI (79% lower FAI and 62% lower NEADL) score compared with a patient with mild to moderate stroke (P<0.001). The cognition score was a significant predictor for the EADL scales: FAI and NEADL but not for BI. Cognitively intact patients (AMT: 8–10) had FAI values 2.31 times greater (65% higher NEADL) compared with cognitively impaired patients (P<0.001). The prestroke BI score was a predictor for BI and NEADL only. These multivariable results were based on only 140 (excluding 98 from 238) patients due to missing values in baseline variables. However, sensitivity analysis assured that there was no bias in the prediction of future functional status based on the NIHSS score (see the online-only Data Supplement for details).

**Relationship of NIHSS Score With the Functional Outcomes**

Because NIHSS turned out to be a universal predictor for the future functional status, fractional polynomial models of degree 2 were fitted to the see the relationship between NIHSS and outcome scales in online-only Data Supplement Figure I. An inverse relationship between NIHSS scores and scales outcomes was found, which was expected. Although NIHSS scores increased from 0 to 24, the predicted scores for the BI, FAI, and NEADL decreased roughly from 95% to 35%, 58% to 5%, and 75% to 18%, respectively, indicating the ceiling effect of BI and floor effect of FAI influenced by the acute stage NIHSS scores. It also showed that NEADL was more sensitive (to NIHSS scores) than FAI.

**Representativeness of the Analyzed Patient Data**

We analyzed 238 patients after excluding 93 because they had missing values at least in 1 of the outcome scales due to loss to follow-up. The mean age, prior stroke disability (BI <20), severe stroke, and cognitive impairment for the excluded versus included patients were, respectively 71 years versus 69 years (P=0.15), 24% versus 20% (P=0.29), 15% versus 13% (P=0.61), and 33% versus 26% (P=0.29). It clearly shows that the loss to follow-up patients were not
significantly different from the analyzed patients, which reassures the generalizability of our results.

Discussion

This study has shown that BI had a more pronounced ceiling effect compared with EADL scales, the FAI and NEADL. However, the FAI was highly affected by a floor effect, whereas the NEADL scale was symmetrical with no floor and ceiling effects underlying its usefulness for stroke research. Although the scales were highly significantly concurrently valid with respect to one another, the agreements between them were poor because the NEADL and FAI were not unbiased with respect to BI. The FAI and NEADL agreed with BI only for the highest and lowest level of activities. The NEADL corresponded better with BI than FAI. Prediction of BI, FAI, and NEADL was possible by only 1 universal predictor, the NIHSS, suggesting its predictive power for future functional status. Cognition was a significant predictor only for the EADL scales: FAI and NEADL.

One third of the patients with first-ever stroke who survived the first 3 months after a stroke were independent in their ADL. The distributional findings regarding the ceiling effect of BI (33%) and floor effect of FAI (19%) are in line with previous research. Several authors concluded that floor and ceiling effects are considered to be present if >15% of respondents achieve the lowest or highest possible score, respectively. Therefore, although BI suffered from a ceiling effect and FAI suffered from a floor effect, NEADL was free from both of these effects. The consequence of presence of either floor or ceiling effect is that the patients with lowest or highest possible score cannot be distinguished from each other; thus, reliability is reduced and also responsiveness is limited because changes cannot be measured in these patients. Another criticism of the FAI was that it does not recognize some activities such as voluntary work or telephoning. Moreover, it does not take into account newer activities such as computer work. In contrast, the NEADL takes into account telephoning. Interestingly, the distribution of NEADL scores was symmetrical further suggesting it to be a more useful outcome measure for EADL than FAI.

Both ICC and concordance correlation coefficient are parametric in nature and hence CIs may not be valid due to floor and ceiling effects of the 2 outcome scales. Log transformation to make the distribution of scores normal was not appropriate because the scores included 0. CI under the Spearman rank correlation coefficient seems more acceptable than others because it is based on ranks rather than actual values. All 3 scales had high concurrent validity (r ≥ 0.80). However, reliability exists between the 2 EADL scales only (r = 0.75) because ICC=0.70 is often recommended as a minimum standard for reliability. It makes sense, because reliability measures the degree to which patients can be distinguished (eg, with more or less severe disease). Irrespective of methods, the concurrent validity and reliability (if any) between NEADL and BI were higher than those of between FAI and BI.

Agreement concerns how close the scores of the 2 scales are to one another. The higher agreement of NEADL with BI than that of FAI with BI might be justified due to the lowest coefficient of variation between NEADL and BI found in this study and minimal overlap between FAI and BI reported in the literature. This suggests that NEADL may add information about ADL lacking in the BI. Some authors combined scales in stroke trials to derive a global statistics to better define the effect of acute interventions. Others showed that such combined score had a much improved distribution without strong ceiling or floor effects. In this case, FAI may be combined with the BI to cover the whole spectrum of activities due to the highest coefficient of variation and minimal overlap in terms of correlation and agreement.
Appelros found that age, NIHSS, and cognition were significantly related with FAI at 1-year poststroke. In our analysis, NIHSS was a universal predictor for all 3 scales outcomes suggesting its predictive power for future functional status. Kasner also concluded that NIHSS was useful for early prognostication. In our study, cognition was a significant predictor for FAI and NEADL but not for BI. This might be due to the requirement of preserved mental skills for many items of EADL that are obtained by FAI and NEADL such as reading newspapers or books. Unlike the findings of Appelros, age was not related to FAI. This may be due to a confounding effect in Appelros univariable analysis. Our result was based on multivariable-adjusted models and hence should be more reliable. Prestroke BI was not only a significant predictor for poststroke BI, but also for NEADL emphasizing the general correspondence between BI and NEADL.

Schlegel et al showed in a retrospective analysis that NIHSS predicted postacute care disposition among patients with stroke. Our prospective study showed that patients with severe stroke had 28% lower BI (79% lower FAI and 62% lower NEADL) score than nonsevere patients. On the other hand, cognitively intact patients had more than double FAI values (65% higher NEADL) compared with impaired patients. This means that the acute stage scores of the NIHSS can be successfully used for prediction/prognostication of future functional status of patients with stroke, whereas AMT score can only be used for EADL. Predicting future functional status at the acute stage may decrease unnecessary length of stay in the acute care setting and perhaps may also facilitate the planning of poststroke rehabilitation programs.

There are strengths and limitations of our study. We had only complete information for 72% of all stroke survivors; however, they were not significantly different from the dropouts. The fact is that more severe patients are likely to produce missing values than less severe patients, but sensitivity analysis did not warrant any significant bias to apply missing data treatment. Our data were derived from a population-based sample, which increases the external validity of the findings. All data were collected by specially trained researchers using standardized follow-up methods. However, we did not formally check interrater reliability of our data collection, although the reliability of FAI and NEADL was reported to be lower than that of BI in the literature.

Conclusions

The NEADL scale was symmetrical with no floor and ceiling effects in a general stroke population. Despite high concurrent validity of the scales, they agreed poorly only for the highest and the lowest level of activities. The NEADL corresponded better with BI than FAI did confirming its basic ADL properties, yet it is a more sensitive tool for EADL without the floor and ceiling effects. Future functional status could be predicted by the universal predictor NIHSS, whereas only EADL status could be predicted by the AMT score. Predicting future functional status at the acute stage may decrease unnecessary length of stay in acute care settings.

Acknowledgments

We thank all the patients and their families and the healthcare professionals involved. We are also very thankful for the anonymous clinical and statistical reviewers for their very helpful comments.

Sources of Funding

Funding for the Register has been provided through the Northern & Yorkshire National Health Service (NHS) R & D Programme in Cardiovascular Disease and Stroke, Guy’s and St Thomas’ Hospital Charity, Stanley Thomas Johnson Foundation, The Stroke Association, Department of Health HQIP grant, and the National Institute for Health Research Programme Grant (RP-PG-0407-10184). The authors acknowledge financial support from the Department of Health through the National Institute for Health Research (NIHR) Biomedical Research Centre award to Guy’s & St Thomas’ NHS Foundation Trust in partnership with King’s College London. Charles Wolfe is an NIHR Senior Investigator.

Disclosures

None.

References


Comparison of 2 Extended Activities of Daily Living Scales With the Barthel Index and Predictors of Their Outcomes: Cohort Study Within the South London Stroke Register (SLSR)

Shah-Jalal Sarker, Anthony G. Rudd, Abdel Douiri and Charles D.A. Wolfe

*Stroke*. 2012;43:1362-1369; originally published online March 29, 2012; doi: 10.1161/STROKEAHA.111.645234

*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/43/5/1362

Data Supplement (unedited) at:

http://stroke.ahajournals.org/content/suppl/2012/03/29/STROKEAHA.111.645234.DC1

**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/
ONLINE SUPPLEMENT

Comparison of two Extended Activities of Daily Living scales with the Barthel Index and predictors of their outcomes– Cohort study within the South London Stroke Register (SLSR)

Shah-Jalal Sarker, PhD¹,³; Anthony G Rudd, MD¹,²; Abdel Douiri, PhD¹,² and Charles DA Wolfe, MD¹,²

Address for correspondence
Shah-Jalal Sarker
Lecturer in Biostatistics
Old Anatomy Building, Charterhouse Square
London EC1M 6BQ
Tel: + 44 (0)20 7882 8495
Fax: + 44 (0)20 7882 5958
Email: s.sarker@qmul.ac.uk

Division of Health and Social Care Research, King’s College London, London, UK.¹
NIHR Comprehensive Biomedical Research Centre, Guy’s & St. Thomas’ NHS Foundation Trust and King’s College London, London, UK.²
Centre for Experimental Cancer Medicine, Barts Cancer Institute - a CR-UK Centre of Excellence, Queen Mary, University of London, London, UK.³
Sensitivity of the multivariable analysis results
The multivariable negative binomial regression analysis in Table 3 was based on 140 patients excluding 98 patients who had at least one baseline variable missing. The median scores for the excluded versus included patients were (85 vs 92.5, \( p=0.21 \)), (15.6 vs 26.7, \( p=0.02 \)) and (38.7 vs 52.7, \( p=0.11 \)) for the BI, FAI and NEADL respectively. The median NIHSS score for the excluded patients (\( n=43 \), median=9) was significantly higher than that of included patients (\( n=140 \), median=5), \( p<0.01 \). This shows that analysed patients had slightly less severe stroke and hence higher functional status. To check the bias in NIHSS score based prediction, univariable negative binomial regression models were fitted. The relative change in the BI, FAI and NEADL scores for the excluded (\( n=43 \)) versus included (\( n=140 \)) patients were respectively (0.69, \( p=0.03 \) vs 0.69, \( p<0.001 \)), (0.56, \( p=0.19 \) vs 0.14, \( p<0.001 \)) and (0.61, \( p=0.13 \) vs 0.29, \( p<0.001 \)). The prediction for ADL was clearly unbiased. However, it was not possible to assess the bias for EADL prediction as the model fitted for the excluded patients were not significant.

### Table S1. Results of the deviance tests for fitted fractional polynomial models: BI, FAI and NEADL

<table>
<thead>
<tr>
<th>Fractional Polynomial Model †</th>
<th>Degree of FP</th>
<th>Deviance</th>
<th>Residual Standard Deviation</th>
<th>Deviance difference</th>
<th><strong>P-value</strong></th>
<th>Model ( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FP</strong>&lt;sub&gt;FAI&lt;/sub&gt; versus BI</td>
<td>2</td>
<td>2051</td>
<td>18</td>
<td>5.39</td>
<td>0.07</td>
<td>0.59</td>
</tr>
<tr>
<td><strong>FP</strong>&lt;sub&gt;NEADL&lt;/sub&gt; versus BI</td>
<td>3</td>
<td>1988</td>
<td>16</td>
<td>0.07</td>
<td>0.96</td>
<td>0.77</td>
</tr>
<tr>
<td><strong>FP</strong>&lt;sub&gt;NEADL&lt;/sub&gt; versus FAI</td>
<td>1</td>
<td>1925</td>
<td>14</td>
<td>4.09</td>
<td>0.13</td>
<td>0.83</td>
</tr>
</tbody>
</table>

*P*-value from deviance difference comparing reported model with the respective next higher degree model
† BI=Barthel Index, FAI=Frenchay Activities Index and NEADL=Nottingham Extended Activities of Daily Living.

**Figure S1.** The relationship between NIHSS and outcome scales: Barthel Index (BI), Frenchay Activities Index (FAI) and Nottingham Extended Activities of Daily Living (NEADL) using fractional polynomial regression.