The AMC Linear Disability Score (ALDS) : measuring disability in clinical studies

Weisscher, N.

Citation for published version (APA):
could not discriminate between H&Y grading 1 and 2. In contrast to the UPDRS-ADL, the ALDS turned out to be sufficiently sensitive to discriminate between H&Y stages 2 and 3.

The ALDS item bank is constructed using the modern psychometric technique of IRT. An important advantage of this approach is that when assessing the ability to perform ADL, it is possible to present more difficult items to less disabled patients and easy items to more severely disabled patients, while the scores obtained remain comparable across the whole patient group. Combined with other attractive features, for example improving the clinical interpretation of scores and the possibility to use computer adaptive testing, the ALDS is a promising new instrument to assess the level of disability in patients with PD.

References


Abstract

Background and Purpose: The aim of this study was to examine the clinimetric properties of a new generic disability measure based on item response theory; the AMC Linear Disability Score (ALDS).

Methods: A prospective cohort study; 213 stroke patients were evaluated using the NIH Stroke Scale (NIHSS), Barthel Index (BI), modified Rankin scale (mRs) and the ALDS at admission and discharge of a stroke unit, and after 6 months follow-up.

Results: The internal consistency and test-retest reliability of the ALDS was good (range Cronbach $\alpha = 0.90 – 0.93$; ICC = 0.85). The ALDS was highly correlated with other disability measures (range $r = 0.75$ to 0.89) and less with the NIHSS ($r = 0.32$). The mean ALDS scores were significantly different between stroke types ($p = 0.008$), and between mild, moderate or severe stroke (NIHSS; $p < 0.001$) at hospital admission. Disability level based on the ALDS improved significantly over time ($p < 0.001$), whereas the responsiveness of the ALDS was moderate to large (Cohen $d$ effect size $= 0.77$ to 0.83). Distribution of the ALDS and BI scores for each mRs levels showed the increased sensitivity of the ALDS over the BI at the lower levels of disability.

Conclusion: The results of this study show that the use of different subsets of items from the ALDS item bank has good clinimetric properties across the entire continuum of stroke severity without floor or ceiling effects.

Introduction

Although most stroke intervention trials use disability outcome measures, several problems with the existing scales remain. None of the currently used instruments are sensitive across the entire continuum of stroke severity. Therefore, disability after stroke is hard to measure. The most common disability measure is the Barthel Index (BI), which has ceiling effects where patients with the best score may still have substantial disabilities. The BI is ordinal sum score-based, which means that it is only possible to ascertain whether there has been a change in functional status. Hence, a functional health change from 5 to 10 points is not the same as a change from 10 to 15. Another approach is using the modified Rankin scale (mRs), a global functional health index, in which each ranking describes a clearly defined clinical condition. In the acute setting, the mRs can have a floor effect, and scores are often dichotomized in clinical trials. A ‘good’ outcome defined as mRs grade $\leq 1$ or grade $\leq 2$ is estimated to be more powerful than dichotomization at higher grades. Concerns over dichotomized endpoints concentrates on the risk of failure to detect the impact of treatment.

To address the problems regarding floor and ceiling effects, ordinal character of sum scores and clinical sensitivity in general, item response theory (IRT) was proposed as an alternative to the use of summed scores or indexes. In conjunction with IRT it is possible to develop an item bank; a collection of items, for which the measurement properties of each item are known. Using IRT we developed a linear, generic, non-disease specific outcome scale which can be used in both mildly and in severely affected patients. The AMC Linear Disability Score (ALDS) is an item bank consisting of a large number of ADL items hierarchically ordered from simple (‘put on a T-shirt’) to complex (‘carry a shopping bag upstairs’). A major strength of an item bank is its adaptive use. By using a small number of items, tailored to the ADL level of patients, a sufficiently detailed clinical picture can be obtained. Even if different sets of items are used for different patient groups, ALDS scores still can be compared within or between medical specialties. The current version of the ALDS item bank consists of 77 items. The objective of this study was to examine the reliability and validity of the ALDS item bank in stroke, and to compare the ALDS with the BI and mRs in terms of responsiveness and floor and ceiling effects.

Materials and Methods

Subjects

A total of 213 consecutive patients were included in this study who were treated at the stroke unit of the Academic Medical Center, Amsterdam, the Netherlands between January 2004 and May 2005. All patients had a confirmed stroke, defined as a rapid onset neurological deficit reflecting a focal disturbance of cerebral function of vascular origin, persisting for > 24 hours. Stroke was diagnosed by clinical assessment and by a CT/MRI scan. Both first and recurrent strokes were included.

Procedure

Baseline assessments included stroke severity by the NIH Stroke Scale (NIHSS), and stroke type (ischemic or hemorrhagic). Ischemic stroke was categorized by the TOAST classification; large artery atherosclerosis, cardioembolism, small-artery occlusion, other or unusual causes, and undetermined etiology. Experienced neurologists were certified in the use of the NIHSS scale. One of the authors (NW), in cooperation with a physical therapist assessed patients’ functional ability using the BI, mRs and the ALDS within three days of admission and at discharge. Trained nurses evaluated the mRs and ALDS by telephone 6 months post stroke. Patients who were not communicative because of severe speech, language or other cognitive disorders were rated by a proxy respondent, in the acute phase a nurse and after 6 months primarily the partner.

ALDS item bank

The item bank was developed to quantify the ability to perform ADL using a two-parameter IRT framework. Both the item difficulty and the patient’s ability are arranged on a single hierarchical
linear scale. The items were obtained from a systematic review of generic and disease specific functional health instruments. Each item describes an ADL, for example ‘get out of bed into a chair’, and ‘walk >15 minutes’. Respondents have to rate if they could carry out the activity at present and were given two response options: ‘I can carry out the activity’ and ‘I cannot carry out the activity’. If a patient had never experienced an activity (e.g., ‘driving a car’) ‘not applicable’ was recorded. Since IRT centers on the measurement properties of individual items, rather than the instrument as a whole, it is not essential for all respondents to be examined using all 77 items. Researchers can select items from the item bank that are applicable to the population that they are investigating. This means that the item bank can be used to assess patients with a wide range of conditions and functional levels, without placing an undue strain on the patients. Regardless of the ALDS items used, it is possible to compare the functional status between patients and populations.

Each patient was assessed using 14 items of the 77 items in the current version of the item bank at admission and discharge. Six months post stroke, dependent on the patients’ ability level, four increasingly difficult item sets were used. The item sets were composite by one of the authors (NW), a physical therapist and a nurse.

The methodology, the psychometrics of the ALDS in terms of dealing with missing data, differences between item measurement characteristics in relation to age and sex and the metric properties of ALDS items in mixed types of patient groups, and the statistical power to detect given effect sizes in clinical trials using IRT outcome scales have been examined in depth. The original units of the ALDS scale are (logistic) regression coefficients, expressed in logits (see Appendix 3). To make the results easier to interpret, the logit scores are linearly transformed into values between 0 (dead) and 100 (highest level of functional status).

**Clinimetric analysis**

The clinimetric properties of the ALDS were studied in terms of reliability (internal consistency and test-retest reliability), construct and clinical validity, responsiveness and the presence of floor and/or ceiling effects. To assess internal consistency, Cronbach’s α was calculated at all time points (criterion set at α ≥ 0.80). Test-retest reliability between admission and discharge was assessed using intraclass correlation coefficient (ICC). Since test-retest reliability is preferably assessed in stable patients, only patients were studied having the same BI quartile scores at both time points (criterion set at ICC ≥ 0.70).

Construct validity was assessed by measuring the extent to which the ALDS correlates with measures (BI and mRs) addressing the same health concept (disability) or a measure (NIHSS) that reflect a conceptually different aspect of health (impairment). We assumed that the ALDS to be valid, the ALDS scores had to correlate substantially with the disability measures. In addition; we would expect the ALDS scores to show a lower association with the impairment scale.

A scale demonstrates clinical validity if it discriminates between groups of patients with known differences in clinical status. Group differences were determined by comparing the ALDS between stroke severity (NIHSS) and stroke type (ischemic or hemorrhagic) at admission. NIHSS were classified into mild (NIHSS 0 to 5), moderate (NIHSS 6 to 13), or severe (NIHSS ≥14) strokes.

Responsiveness was investigated by calculating the Cohen’s d effect size (mean change in scores divided by the standard deviation of the baseline scores) for the ALDS and BI between admission and discharge and additionally for the ALDS between discharge and 6 month follow-up. An effect size value between 0.50 and 0.80 is considered moderate, and ≥0.80 as high responsiveness. As Cohen’s d cannot be calculated for ordinal scales with few categories, this effect size is not presented for the mRs.

**Statistical analysis**

Descriptive statistics were used to analyze demographics, stroke characteristics, NIHSS score, and disability status according to the BI, mRs and ALDS. Scores on the ALDS were calculated using previously published item parameters and algorithms implemented in BILOG-MG and SPSS 12.0. ALDS responses in the ‘not applicable’ category were treated as if the item had not been represented to the patient. Values of Cronbach’s α were obtained using a specific IRT method that allow for missing item responses implemented in BILOG-MG.

The associations between the ALDS scores and the other measures were expressed in Pearson’s r or Spearman’s rho correlation coefficients (r). To examine the clinical validity, an independent t test and one-way analysis of variance (ANOVA) with Tukey-Kramer corrections for multiple comparisons was undertaken to examine the differences in mean ALDS scores. Differences in mean ALDS scores were also expressed in Cohen’s d effect size.

Change in disability scores over time were examined using a paired t test (BI), a linear mixed model (ALDS) or a Friedman test (mRs), when appropriate. In general, the original ALDS logits were used in all analysis but for clarity only the linearly transformed scores (0-100) are presented. Finally, the floor and ceiling effects of the BI and ALDS scores across mRs levels were graphically examined.

**Results**

Table 1 summarizes the characteristics of the 213 patients at admission. Before discharge, 19 patients (9%) died; hence 194 patients were assessed at discharge. By the end of the 6 months follow-up another 37 patients (19%) died, four patients refused to participate further, and in 36 patients no outcomes could be assessed (declined or unable to cooperate, language barrier and no proxy available, moved to another country); accordingly 117 patients completed the 6 months assessments.
Reliability
Cronbach α-coefficients of the ALDS scores at all time points (admission, discharge, follow-up) ranged from 0.90 to 0.93. Fifty three percent of the patients had a stable BI quartile score between admission and discharge and were included in the analysis of test-retest reliability. The ICC for the ALDS scores was 0.85 (95% CI; 0.78 to 0.89).

Validity
The ALDS showed high correlations with the other disability measures: BI (admission; \( r = 0.80 \), discharge; \( r = 0.86 \)); mRs (admission; \( r_s = 0.75 \), discharge; \( r_s = 0.87 \), follow-up; \( r_s = 0.89 \)). The association between the ALDS and the NIHSS at admission was lower (\( r = 0.32 \)). Table 2 presents the ALDS scores in relation to stroke severity and type of stroke. Disability level based on the ALDS (Tukey HSD all \( p < 0.001 \)) was significantly different between minor, moderate and severe stroke with small to moderate effect sizes. Compared to ischemic stroke, patients with hemorrhagic stroke were significantly more disabled.

Responsiveness
Table 3 shows the score distribution for the three disability instruments at the different time points. The scores on all measures significantly improved over time. The effect size between admission and discharge indicated that both the ALDS (\( d = 0.77 \)) and the BI (\( d = 0.50 \)) were moderate responsive. The effect size for the ALDS between discharge and six months post stroke was large (\( d = 0.83 \)).

Validity

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Mean (SD), Median (range) or n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, y</td>
<td>70.7 (14.8)</td>
</tr>
<tr>
<td>Male sex</td>
<td>97 (46%)</td>
</tr>
<tr>
<td>First ever stroke</td>
<td>178 (84%)</td>
</tr>
<tr>
<td>Stroke type</td>
<td></td>
</tr>
<tr>
<td>Ischemic</td>
<td>170 (80%)</td>
</tr>
<tr>
<td>large artery atherosclerosis</td>
<td>41 (24%)</td>
</tr>
<tr>
<td>cardioembolism</td>
<td>99 (65%)</td>
</tr>
<tr>
<td>small-artery occlusion</td>
<td>43 (25%)</td>
</tr>
<tr>
<td>other or unusual causes</td>
<td>13 (8%)</td>
</tr>
<tr>
<td>undetermined etiology</td>
<td>14 (8%)</td>
</tr>
<tr>
<td>Hemorrhagic</td>
<td>43 (20%)</td>
</tr>
<tr>
<td>NIHSS</td>
<td>7 (1-40)</td>
</tr>
<tr>
<td>0-5</td>
<td>80 (38%)</td>
</tr>
<tr>
<td>6-13</td>
<td>79 (37%)</td>
</tr>
<tr>
<td>&gt; 14</td>
<td>52 (24%)</td>
</tr>
<tr>
<td>not assessed</td>
<td>3 (1%)</td>
</tr>
<tr>
<td>BI (n=204)</td>
<td>6.7 (6.9)</td>
</tr>
<tr>
<td>ALDS</td>
<td>20.1 (16.3)</td>
</tr>
<tr>
<td>mRs</td>
<td>4 (2-5)</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>4 (2%)</td>
</tr>
<tr>
<td>3</td>
<td>32 (15%)</td>
</tr>
<tr>
<td>4</td>
<td>78 (37%)</td>
</tr>
<tr>
<td>5</td>
<td>99 (46%)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>N</th>
<th>ALDS (SD)</th>
<th>p-value</th>
<th>Cohen d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Admission NIHSS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-5</td>
<td>80</td>
<td>25.8 (20.5)</td>
<td>&lt;0.001*</td>
<td>0.41</td>
</tr>
<tr>
<td>6-13</td>
<td>78</td>
<td>18.7 (13.8)</td>
<td></td>
<td>0.42</td>
</tr>
<tr>
<td>&gt; 14</td>
<td>52</td>
<td>13.9 (8.7)</td>
<td></td>
<td>0.40</td>
</tr>
</tbody>
</table>

**Table 2. ALDS admission scores in relation to stroke severity and stroke type (n=213).**

**Table 3. Disability level by Barthel Index, ALDS and modified Rankin scale in relation to time: admission, discharge and 6 months post stroke.**
Floor and ceiling effect
Distribution of the ALDS and BI scores for each mRs level shows the increased sensitivity of the ALDS over the BI in less disabled patients (Figure 1 and 2). The BI shows already a ceiling effect in the acute phase in slightly disabled patients (mRs 2). In comparison, the ALDS showed no floor or ceiling effects across the whole range of mRs levels.

Figure 1. Distribution of the Barthel Index according to the modified Rankin scale.

Figure 2. Distribution of the ALDS according to the modified Rankin scale.

Discussion
Disability scales are frequently used primary endpoints for stroke treatment trials.\textsuperscript{8, 31} As a result of the limitations of existing scales, new reliable, valid and sensitive outcome measures are welcome. This study showed the ALDS item bank has promising clinimetric properties in the acute phase and 6 months post stroke.

In clinical trials, the most widely used disability scales are the BI and mRs.\textsuperscript{9, 32} Where the BI is more appropriate for assessment of severe disability in the acute phase, the mRs is a better instrument for differentiating between changes in mild to moderate disability, especially after minor stroke.\textsuperscript{33} The ALDS expands the range of physical functioning assessment beyond the measurement range of both the BI and mRs by explicitly including both basic ADL and instrumental ADL. The ALDS items can be used adaptively, meaning more difficult items can be presented to more disabled patients, while the easier items can be presented to more disabled patient. By using the ALDS in this flexible way, a detailed clinical picture can be obtained along the whole stroke continuum, thereby reducing time, effort and patient burden. The ALDS will also form a good foundation for a computerized adaptive testing procedure, allowing for shorter questionnaires where the difficulty level is automatically adapted per question depending on the individual patient’s ability to perform the questioned activity.\textsuperscript{34}

Studies have found that sample size requirements of trials using mRs-based endpoints are smaller than BI-based endpoints without loss of statistical power.\textsuperscript{33, 35} At present, there is no consensus which mRs defined endpoint is the most appropriate ‘good’ outcome (mRs grade ≤ 1 or ≤ 2). However, in general, clinical information and statistical power are reduced when data are categorized.\textsuperscript{12} Therefore, analyzing data across all levels of functional status is important. Besides the linear nature of the ALDS, an IRT calibrated item bank offers an additional option to increase the statistical power by simply increasing the number of items to be used.\textsuperscript{24}

The use of IRT to improve outcome assessment recently got increased attention. Especially the Rasch model is often used to minimize the number of questions asked while maintaining or improving measurement characteristics of an existing instrument. For example, the Stroke Impact Scale -16 is a shorter instrument to assess physical function based on a Rasch analytical technique.\textsuperscript{1} Hattingsveld\textsuperscript{4} used Rasch analysis to improve the measurement properties and clinical interpretability of BI scores. Hsueh\textsuperscript{36} performed Rasch modeling to combine the BI and Frenchay Activity Index items. However, to our knowledge in the stroke field no physical functioning IRT calibrated item bank exists.

Measuring disability in the acute phase of stroke could be a limitation of the recent study, because
most patients are bedridden and impaired disability is not the primary medical concern. In the recent study one of the objectives was to investigate the sensitivity of the ALDS item bank across the entire continuum of stroke severity. Although the dispersal of disability status is rather small in acute stroke, the ALDS was still able to detect differences in disability level between different NIHSS categories.

In summary, the ALDS is a new instrument that is a simple and flexible tool to assess ADL disability in stroke patients. The results of this study show that the use of different subsets of items from the ALDS item bank has good clinimetric properties across the entire continuum of stroke severity without floor or ceiling effects. Obviously, this is an initial study on the use of the ALDS in stroke. More research is needed with regard to the implication for sample size calculations in stroke trials, the interobserver reliability, and the development of adaptive testing software.

References

Chapter 7

What should be defined as good outcome in stroke trials: a modified Rankin score of 0-1 or 0-2?

Nadine Weisscher, Marinus Vermeulen, Yvo B Roos, Rob J de Haan

J Neurol (in press)