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

Weisscher, N.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 3


Chapter 4

The AMC Linear Disability Score item bank: psychometric properties of a new generic disability measure in rheumatoid arthritis patients

Nadine Weisscher, Carla Wijbrandts, Rob de Haan, Cees Glas, Marinus Vermeulen, Paul Peter Tak

J Rheumatol. 2007;34:1222-8
Abstract

Objective: To determine the psychometric properties of the Academic Medical Center (AMC) Linear Disability Scale (ALDS) item bank in a population of patients with rheumatoid arthritis (RA).

Methods: 129 patients with RA completed the ALDS and Health Assessment Questionnaire Disability Index (HAQ-DI) at baseline, and after 8 and 16 weeks of anti-tumor necrosis factor-therapy. Disease activity assessments at these time points included serum levels of C-reactive protein, Disease Activity Score 28 (DAS28), morning stiffness and visual analog scales for global disease activity and fatigue.

Results: Reliability of the ALDS was excellent (Homogeneity, Cronbach’s $\alpha = 0.95$; test-retest, ICC = 0.93). The ALDS results at baseline were strongly correlated with the HAQ-DI ($r = -0.75$). With regard to known-group validity, both instruments discriminated between higher and lower disease activity (ALDS, $p < 0.0001$; HAQ-DI, $p = 0.002$) and between non-, moderate and good responders (ALDS, $p = 0.002$; HAQ-DI, $p < 0.0001$), indicating that both instruments differentiate between groups. The ALDS was moderate to highly responsive to change between baseline, and after 8 weeks and 16 weeks of treatment (standardized response mean, range = 0.71-1.19). No substantial floor or ceiling effects were found.

Conclusions: Our results show that the ALDS is a promising new instrument, with at least equivalent psychometric properties compared to the HAQ-DI. Advantages of the ALDS item bank are its linear structure and an item bank can be used adaptively depending on the ability level of the patient.

Introduction

The influence of a disease on the patient’s level of activities of daily living (ADL) is generally considered an important outcome measure in clinical studies. At present a lot of effort is put into the development of new measures to assess patients’ daily functioning and perception of illness. Developments such as item response theory (IRT) based item banks and computer adaptive testing for the measurement of patient-reported outcomes are of increasing interest in the rapidly developing field of outcome assessment in rheumatology.

The Health Assessment Questionnaire Disability Index (HAQ-DI) has become the most frequently used and validated functional disability scale in rheumatology. Although the HAQ-DI has shown to be an effective, reliable, and valid tool that is sensitive to change, a few issues remain. The HAQ-DI is a fixed-length instrument, meaning that all the items of the scale must be administered to all patients to calculate a total score irrespective of their level of disability, which means that able patients as well as more disabled patients will be presented the same items. This impractical approach may lead to ceiling and floor effects. Another disadvantage is that the HAQ-DI uses an ordinal instead of a linear scale, thus a similar change in HAQ-DI scores represents a different amount of change in function depending on where the patient is situated on the scale (e.g., a functional health change in HAQ-DI from 0.5 to 1.0 is not the same as a change from 2.0 to 2.5). Interest is currently moving from sum score-based methods toward the more flexible framework offered by item banks in conjunction with IRT. An item bank is a collection of items, for which the measurement properties of each item are known. Using psychometric methods based on IRT we developed an outcome scale which is not ordinal but linear and that identifies the full range of disability. The Academic Medical Center Linear Disability Score (ALDS) is a generic, non-disease specific item bank consisting of a large number of ADL items hierarchically ordered from simple to complex activities. 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 groups of patients, ALDS scores can still be compared within or between medical specialties.

The methodology used to develop the ALDS item bank has been analyzed in depth. We used subsets of items from the ALDS item bank and have expanded the psychometric evaluations to patients with rheumatoid arthritis (RA).

Methods

Patients

The patients with RA in our study were participants in a clinical trial with anti-tumor necrosis factor-therapy (infliximab). As part of this study the course of functional status in relation to clinical response to anti-TNF-therapy was studied. The data presented here was obtained at baseline assessment and follow-up after 8 weeks and 16 weeks of treatment. All patients were included in the study after failing at least 2 disease modifying anti-rheumatic drugs (DMARD) including methotrexate (MTX), and having active disease defined as a Disease Activity Score 28 (DAS28) score $\geq 3.2$. Patients were taking maximal tolerable MTX treatment (5-30 mg/week), which had to be stable for at least 4 weeks prior to baseline. Oral corticosteroids ($\leq$10 mg/day) and non-steroidal anti-inflammatory drug were allowed if stable for at least 1 month prior to baseline. The study protocol was approved by the Medical Ethics Committee of the Academic Medical Center, University of Amsterdam. All patients gave written informed consent and were interviewed by trained research nurses.
Assessment

Demographic variables (age and sex) and disease status in terms of disease duration, presence of erosions, and rheumatoid factor positivity were registered at baseline. Disease activity was assessed by the DAS28 score, erythrocyte sedimentation rate (ESR), and serum levels of C-reactive protein (CRP). The DAS28 was based on the following 4 core variables: swollen joint count, tender joint, ESR and a 100 mm patient’s global disease activity visual analog scale (VAS). Other variables assessed were the HAQ-DI, the ALDS, early morning stiffness in minutes, and VAS for fatigue. All variables were obtained at baseline and after 8 weeks and 16 weeks of anti-TNF-α therapy.

The ALDS item bank

The ALDS item bank was developed to quantify functional status in terms of the ability to perform ADL using a 2-parameter logistic IRT framework (see Appendix 2 for methodological details). The current version of the item bank consists of 77 items, ranging from relatively easy to difficult (Appendix 3). Each patient was assessed with one of 4 increasingly difficult item sets, depending on the ability level of the patient. Beforehand the 4 different item sets were chosen by 2 of the authors (NW, RdH) and contained 14 items on average. Mildly disabled patients were presented with more difficult item sets because those were likely to provide more information. On the other hand, easier item sets lower on the scale were presented to more disabled patients. The items were administered by trained nurses and had 2 response options: ‘I can carry out the activity’ and ‘I cannot carry out the activity’. Participants were asked to indicate whether they could perform the activities now in the same way that they would do at home or in the location they were residing. If an activity had never before been performed by a patient (e.g., ‘travel by local bus or tram’), or the patient’s response was ‘I do not know’, ‘not applicable’ was recorded. The score ranges from 0 to 100, with lower scores representing more disability.

Psychometric evaluation

IRT was used to construct the ALDS item bank, but in our study the psychometric properties of the ALDS were studied and presented in classical terms of reliability (i.e., internal consistency and test-retest reliability), validity (i.e., construct and known-group), responsiveness, and the presence of ceiling and/or floor effects.

Internal consistency was assessed at the baseline measurement using Cronbach’s α. The Cronbach’s α coefficient is based on the (weighted) average correlation of items within a scale. The criterion used for good overall internal consistency was α ≥ 0.80. Test-retest reliability was assessed using intraclass correlation coefficients (ICC). Since test-retest reliability is preferably assessed in stable patients, the HAQ-DI was chosen to identify patients whose responses did not change across baseline and 8 weeks later. A delta HAQ-DI score of < 0.22 was considered stable. ICC [and the corresponding 95% confidence interval (CI)] was calculated for the ALDS score at the 2 time points for the stable patients only. The threshold for this type of reliability was defined as ICC ≥ 0.70.

Construct validity was assessed at baseline by measuring the extent to which the ALDS correlates with a measure addressing the same concept (HAQ-DI) or measures (CRP, VAS fatigue, morning stiffness, VAS disease activity, DAS28) that reflect different aspects of health. We labeled the strength of the association; absolute values of r = 0.00-0.19 were regarded as very weak, 0.20-0.39 as weak, 0.40-0.59 as moderate, 0.60-0.79 as strong, and 0.80-1.0 as very strong correlation. We assumed that for the ALDS to be valid, the ALDS scores had to correlate at least moderately with the HAQ-DI. In addition, we would expect the ALDS scores to show lower associations with the clinical measures of disease activity (DAS28, VAS disease activity, morning stiffness, and VAS fatigue) and even lower with laboratory measures of disease activity (CRP).

A scale demonstrates known-group validity if it discriminates between groups of patients with known differences in clinical status. Group differences were determined by comparing ALDS scores between more or less disease activity (DAS28 dichotomized on 5.1) and non-, moderate, and good responders according to the European League Against Rheumatism (EULAR) criteria after 16 weeks of anti-TNF-α therapy.

Responsiveness was investigated by calculating the standardized response mean (SRM). An SRM value between 0.5 and 0.80 is considered moderate, and ≥ 0.80 as high responsiveness. Mean ALDS values at 8 and at 16 weeks of follow-up were compared with the mean value at baseline. Wolfe stated that the HAQ-DI cannot reliably detect differences in patients below a HAQ-DI score of 0.24. Therefore, we additionally calculated the SRM between baseline and 8 weeks of treatment for a subpopulation with a HAQ-DI score ≤ 0.24.

With regard to floor and/or ceiling effects, the percentages of patients with a maximal or minimal score at all time points were presented for the ALDS. The HAQ-DI was used as benchmark and therefore, with the exception of the reliability analysis, all the above described psychometric analysis (construct and known-group validity, responsiveness, and ceiling/floor effects) were also done for the HAQ-DI.

Statistical analysis

Patient characteristics, outcome scores, and ceiling or floor effects were analyzed using descriptive statistics. Scores for the ALDS item bank were calculated using previously published item parameters and algorithms implemented in BILOG-MG version 3.0 and SPSS 12.0 for Windows. Patients’ ability parameters were estimated with maximum likelihood methods. ALDS items to which a patient did not respond or gave a response in the ‘not applicable’ category were treated as if they had not been offered to that patient. Values of Cronbach’s α were obtained using a specific IRT method that allows for missing item responses implemented in TESTFACT version 4.0.
Associations between the ALDS scores and the other measures were expressed in Pearson’s (r) or Spearman’s correlation coefficients (r_s). Differences between mean ALDS and HAQ-DI scores were analyzed using an unpaired t-test or analysis of variance (ANOVA) and Tukey’s HSD post hoc analysis, when appropriate. The SRM was calculated by dividing the mean change scores by the standard deviation of the change scores. The original ALDS logits were used in all analysis, but for clarity only the linearly transformed ALDS scores (0-100) are presented.

Results

Patient characteristics

One hundred twenty-nine patients with RA were enrolled between April 2001 and May 2004. The study group consisted of a predominantly female (73%) population with a mean age of 55 (range 23-85) years. On average, patients had failed treatment with 2.2 ± 1.5 DMARD before inclusion in the study. Oral corticosteroids were taken by 34 (26%) patients, mean dose of 2.1 ± 3.7 mg/day. Baseline patient characteristics are summarized in Table 1.

Table 1. Baseline characteristics.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Total (n = 129)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographics</strong></td>
<td></td>
</tr>
<tr>
<td>Age (years; range)</td>
<td>55 (23 - 85)</td>
</tr>
<tr>
<td>Female (n, %)</td>
<td>94 (73)</td>
</tr>
<tr>
<td><strong>Disease status</strong></td>
<td></td>
</tr>
<tr>
<td>Disease duration (months; range)</td>
<td>123 (7 - 519)</td>
</tr>
<tr>
<td>Erosive disease (n, %)</td>
<td>100 (78)</td>
</tr>
<tr>
<td>Rheumatoid factor positive (n, %)</td>
<td>95 (74)</td>
</tr>
<tr>
<td>DAS28 (median, range)</td>
<td>5.7 (3.4 - 8.0)</td>
</tr>
<tr>
<td>ESR, mm/hr (mean, SD)</td>
<td>32 (± 23)</td>
</tr>
<tr>
<td>C-reactive protein, mg/L (mean, SD)</td>
<td>21 (± 27)</td>
</tr>
<tr>
<td><strong>Drug treatments</strong></td>
<td></td>
</tr>
<tr>
<td>Previous DMARD (mean, SD)</td>
<td>2.2 (± 1.5)</td>
</tr>
<tr>
<td>Methotrexate, mg/Wk (mean, SD)</td>
<td>18.7 (± 8.3)</td>
</tr>
<tr>
<td>Corticosteroids (n, %)</td>
<td>34 (26)</td>
</tr>
<tr>
<td>NSAID (n, %)</td>
<td>63 (49)</td>
</tr>
</tbody>
</table>

Table 2. Descriptive statistics of the instruments and construct validity at baseline assessment.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Score</th>
<th>ALDS (n = 108)</th>
<th>HAQ-DI (n = 122)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRP*</td>
<td>11 (3 – 164)</td>
<td>-0.181</td>
<td>0.151</td>
</tr>
<tr>
<td>VAS fatigue</td>
<td>56.7 (± 21.8)</td>
<td>-0.39</td>
<td>0.35</td>
</tr>
<tr>
<td>morning stiffness*</td>
<td>45 (0-1440)</td>
<td>-0.43</td>
<td>0.41</td>
</tr>
<tr>
<td>VAS disease activity</td>
<td>59.7 (± 21.8)</td>
<td>-0.45</td>
<td>0.40</td>
</tr>
<tr>
<td>DAS28</td>
<td>5.9 (± 1.1)</td>
<td>-0.47</td>
<td>0.39</td>
</tr>
<tr>
<td>HAQ-DI</td>
<td>-0.75</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Score values are mean (± SD); correlation values are Pearson’s correlation coefficient and calculated with the ALDS logits. *Due to skewed data median and range as well as Spearman’s rho correlation coefficient are presented. **All correlations are significant at the 0.01 level, except both correlations with CRP. Since higher ALDS scores indicate better functioning, signs of the coefficients are negative. ALDS: AMC Linear Disability score; HAQ-DI: Health Assessment Questionnaire Disability Index; VAS: visual analog scale.

Reliability

The internal consistency of the ALDS was good (Cronbach’s α = 0.95). Twenty-seven percent (n = 34) reported a stable disability level (delta HAQ-DI score of < 0.22) and were included in the analysis of test-retest reliability. The ICC for the baseline and 8 weeks follow-up ALDS scores was 0.93 (95% CI 0.83 - 0.97), well above the 0.70 threshold.

Validity

Table 2 summarizes the scale scores of the measures studied at baseline assessment. Convergent validity was confirmed by a strong correlation between the ALDS and the HAQ-DI. Additionally, the ALDS showed weak to moderate associations with the clinical measures of disease activity (VAS fatigue, morning stiffness, VAS disease activity, DAS28) and a very weak association with the laboratory measure of disease activity (CRP). The correlations of the HAQ-DI with the above mentioned measures were in about the same range.

Table 3 shows the results with regard to the known-group validity; both instruments discriminated between different levels of disease activity (score ≤ 5.1 and score > 5.1) according to the DAS28 (ALDS; t = 3.97, p < 0.0001; HAQ-DI; t = -3.22, p = 0.002). ANOVA showed statistically significant
Differences (ALDS; F = 6.56, p = 0.002; HAQ-DI; F = 16.96, p < 0.0001) between non-, moderate and good responders, indicating that both instruments differentiate between different groups. Post hoc analysis, however, showed no significant difference in ALDS or HAQ-DI scores between moderate and good responders (ALDS, Tukey’s HSD, p = 0.70; HAQ-DI, Tukey’s HSD, p = 0.57).

Table 3. Known group validity at baseline assessment.

<table>
<thead>
<tr>
<th></th>
<th>ALDS</th>
<th>HAQ-DI</th>
</tr>
</thead>
<tbody>
<tr>
<td>DAS28*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 5.1</td>
<td>63.3 (7.4)</td>
<td>1.1 (0.6)</td>
</tr>
<tr>
<td>&gt; 5.1</td>
<td>73.4 (15)</td>
<td>1.5 (0.7)</td>
</tr>
<tr>
<td>EULAR criteria</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-responder</td>
<td>69.6 (18.2)</td>
<td>1.7 (0.8)</td>
</tr>
<tr>
<td>Moderate responder</td>
<td>76.5 (12.2)</td>
<td>1.3 (0.6)</td>
</tr>
<tr>
<td>Good responder</td>
<td>80.4 (12.6)</td>
<td>1.3 (0.6)</td>
</tr>
</tbody>
</table>

*DAS28 was dichotomized on a cutoff score of 5.1. Score distributions are presented as mean (± SD). Differences in mean logit and HAQ-DI scores are calculated using the unpaired t-test (DAS28) and ANOVA (EULAR response criteria). ALDS scores are presented after linear transformation of the original logits.

Responsiveness
The SRM between baseline and 8 weeks of treatment indicated that the ALDS was moderate (SRM = 0.71) and the HAQ-DI (SRM = 0.88) was highly responsive. Both measures were highly responsive to change between baseline and 16 weeks of treatment (ALDS, SRM = 0.85; HAQ-DI, SRM = 1.03). Additionally, the SRM for the subpopulation with HAQ-DI score < 0.24 at baseline and 8 weeks of treatment (n = 16) was 1.19 for the ALDS and 0.98 for the HAQ-DI, respectively.

Floor and ceiling effect
At all time points the ALDS showed no floor (0%) or substantial ceiling effect (baseline = 0.9%; 8 and 16 weeks = 0.8%). The same was true for the baseline scores of the HAQ-DI (floor = 0.8%, ceiling = 2.3%). In addition, the HAQ-DI showed no considerable floor effect at both follow-up measurements (8 weeks: 1.6%; 16 weeks: 0.8%), whereas the percentage patients scoring at the ceiling was substantially higher (8 weeks: 9.3%; 16 weeks: 13.2%).

Discussion
It is now widely acknowledged in the rheumatology field that assessment of patient disability should be part of the core outcome measures used in clinical trials. The HAQ was one of the first patient-reported outcome measures to be used and it has been extensively validated in a wide range of clinical trials.29,30 Since patient-reported outcomes have become increasingly important in rheumatology, the development of novel outcome measures that might supersede the old are under way.

We have shown that subsets of items from the ALDS item bank used in RA can make up a reliable instrument showing good internal consistency and test-retest stability. The strong association between the ALDS and the HAQ-DI indicates that these scales largely focus on the same disability construct. The decreasing association between the ALDS scores and the clinical and laboratory measures demonstrates that the ALDS is less focused on other aspects of health, also indicating construct validity. Although no differences in ALDS scores exist between moderate and good responders to anti-TNF-α therapy, our findings that the ALDS scores differentiate between non- and moderate responders and between higher and lower level of disease activity indicate that the ALDS is sensitive to discriminate between different patient groups. The results concerning construct and known-group validity are identical for the ALDS and the HAQ-DI. Compared to the HAQ-DI, the ALDS was shown to be sufficient, but less responsive between baseline and 8 weeks of treatment. Both instruments were highly responsive after 16 weeks of anti-TNF-α treatment. Moreover, the ALDS was shown to be less susceptible to ceiling effects.

The ALDS items can be used adaptively, meaning that more difficult items higher up the scale can be presented to relatively able patients, while the easier items can be presented to more disabled patients. Accordingly, it is possible to choose for a sample of items ranging from easy to difficult in case of unknown characteristics of the patient group. Items which have a similar difficulty grade are interchangeable and can be used to assemble scales with items that may be of interest for different patient groups. Since the metric properties of all items have been established, ALDS scores can be compared within or between medical specialties, even when different sets of items are used. Another advantage of an IRT based item bank like the ALDS over fixed-length measures is that a linear instead of an ordinal scale is used.

The use of IRT to improve patient-reported outcome received increased attention following the National Institutes of Health initiative to build a comprehensive Patient Reported Outcome Measurement System based on IRT (PROMIS, www.nihpromis.org). This initiative focuses not only on physical health but also mental health and social health. Examples of other physical functioning item banks have been developed by Haley, et al31 and Ware, et al32. However, to our knowledge item banks in use in the clinical field to date are scarce.
When validating this item bank in an RA population we were well aware that a new instrument would only be interesting if it were able to rise above the qualities of the HAQ-DI, which has almost become a ‘gold standard’. The ALDS has at least equivalent metric properties combined with attractive new and relevant features, for example, improving the clinical interpretation of scores and possibility to shorten the instrument. An advantage of the HAQ-DI is the availability of the questionnaire in several languages. The ALDS is still under development and investigation of its psychometric properties is continuing. At present versions in English and Dutch are available; French, Italian and Spanish versions are in the process of validation.

The ALDS is an interesting new instrument that is a simple to use for assessing the level of disability in patients with RA. It has been suggested that a new questionnaire should not only be shorter, and better on a theoretical basis, but it must also have psychometric properties at least as good as the original HAQ-DI.33 Our results show that use of different subsets of items from the ALDS item bank offers equivalent psychometric properties compared to the HAQ-DI. This initial study on the use of the ALDS in RA provides the basis for further testing. However, two advantages of the ALDS item bank compared to the original HAQ-DI are its linearity and that an IRT item bank can be used adaptively and will form a good foundation for computer adaptive testing, where the difficulty level is automatically adapted per question depending on the individual patient’s ability to perform the requested activity.

References

Chapter 5

The AMC Linear Disability Score in patients with newly diagnosed Parkinson disease

Nadine Weisscher, Bart Post, Rob J de Haan, Cees AW Glas, Johannes D Speelman, Marinus Vermeulen

Neurology. 2007;69:2155-61