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

Weisscher, N.

Citation for published version (APA):
Chapter 9

The challenge of the studies included in this thesis was bringing IRT and adaptive testing to the clinical setting. No matter how superior IRT is compared to the well-known classical test theory, it is important IRT does not become a ‘black box’ for clinicians, since they have to interpret the provided outcomes in clinical research and practice. The ALDS item bank will be improved further as new items are added from studies and assessments are fully (web based) computer adaptive. Combined with other merits of IRT based item banks it is expected that this technique gradually will suppress the very large number old classical constructed paper and pencil questionnaires.

References


(8) Holman R. How does item selection procedure affect power and sample size when using an item bank to measure health status? Quality of Life Newsletter. 2004:9-11.


The efficacy of medical interventions is preferably measured using objective physiological parameters, for example blood pressure or MRI scan. These parameters can be measured accurately and reliably, and results are expressed in well-known and concrete units (mm/HG, lesion size). However, these parameters do not tell the whole story about how the disease process affects patients and their life. Since these limitations are recognized, interest has moved towards more patient-relevant outcomes to determine the efficacy of medical interventions in clinical research.

In order to structure and describe disease outcomes in a systematic and hierarchical manner the World Health Organization developed the International Classification of Impairments, Disabilities, and Handicaps (ICIDH, in 2001 replaced by the International Classification of Functioning, Disability and Health (ICF)). The domains contained in the ICIDH can be seen as health-related domains and distinguishes impairment, disability (or activity) and handicap (or participation). The last decades also a strong accent lies on the health-related quality of life (HRQL) model; this model emphasizes the subjective health perception of the patient.

This thesis focuses on the disability level defined as activities of daily life (ADL). The ability to carry out ADL is necessary for independent living in society or to live in an appropriate care setting. For example, dress and bathe, eating, mobility in and around the house, and light household tasks. Many generic and disease-specific disability measures were developed using fixed-length questionnaires. Responses to all items on a scale are required to calculate a sum score. Recently, interest moved to a modern psychometric method, the item response theory (IRT). This theory is adapted from educational measurement to determine the cognitive ability of schoolchildren (the so-called CITO-toets in the Netherlands). Typical of this method is the possibility to measure the same cognitive ability every year, even though the specific questions differ each year. This is possible because the questions are adapted from a calibrated item bank; a collection of items which measure the whole range of the underlying ability (cognitive functioning, disability status) and of each item the psychometric properties are known.

Chapter 2 of this thesis showed the results of a systematic review investigating the interchangeability of disability and HRQL measures. Thirty-one studies investigating the relation between the patient’s impairment level and disability status as well as the relation between the impairment level and HRQL were included. Meta-analytic results revealed that patient’s impairments were higher correlated with disability measures (pooled ES(z) = 0.69) than with HRQL measures (pooled ES (z) = 0.38). The physical component score (pooled ES(2) = 0.43) and disease-specific HRQL (pooled ES(2) = 0.46) were stronger associated with impairments than the mental component score (pooled ES(2) = 0.28) and generic HRQL measures (pooled ES(2) = 0.36). The results of this study showed that measures of disability and various HRQL domains were not equally related to impairment. New drugs, new surgical or radiological interventions are developed from a pathophysiologic perspective and are primarily directed at reducing impairments. Patient’s impairments are better reflected in disability measures than in HRQL instruments. Therefore we conclude that to assess the efficacy of a new treatment, disability is an important functional endpoint in clinical studies.

Chapter 3 describes the development of the AMC Linear Disability Score (ALDS) item bank, using IRT in a mixed patient population. Of 170 items, 115 were judged to be clinically relevant. Of these 115 items, 77 were retained in the item bank following the item response theory analysis. Of the 38 items that were excluded from the item bank, 24 had either been presented to fewer than 200 respondents or had fewer than 10% or more than 90% of responses in the category ‘can carry out’. A further 11 items had different measurement properties for younger and older or for male and female respondents. Finally, 3 items were excluded because the item response theory model did not fit the data. The results showed the ALDS item bank has promising measurement characteristics.

Chapter 4 provides the results of a clinimetric evaluation the ALDS item bank in a population of rheumatoid arthritis (RA) patients. Hundred twenty-nine patients completed the ALDS and Health Assessment Questionnaire Disability Index (HAQ-DI) at baseline, and after 8 and 16 weeks of anti-TNF-\(\alpha\) treatment. Disease activity assessments at these time points included serum levels of C-reactive protein (CRP), Disease Activity Score 28 (DAS28), morning stiffness and visual analogue scales for global disease activity and fatigue. Reliability of the ALDS was good (Cronbach’s \(\alpha = 0.95\); ICC = 0.93). The ALDS results at baseline were strongly correlated with the HAQ-DI (\(r = 0.73\)). Both instruments discriminated between higher and lower disease activity (\(p < 0.01\)) and between non-, moderate and good responders (\(p < 0.01\)), indicating that both instruments differentiate between groups of patients with various severity of disease. The ALDS was moderate to highly responsive to change between baseline, and after 8 weeks and 16 weeks of treatment. No substantial floor or ceiling effects were found. These results showed that the ALDS is a promising new instrument, with at least equivalent psychometric properties compared to the HAQ-DI. An advantage of the ALDS item bank over the HAQ-DI is its linear character and that the item bank can be used adaptively depending on the ability level of the patient.

Chapter 5 describes the development of 132 patients with newly diagnosed Parkinson’s disease (PD). Neurological impairments were evaluated using the Hoehn & Yahr (H&Y) disease staging scale and the Unified Parkinson’s Disease Rating Scale (UPDRS) motor examination. The Schwab & England scale (S&E) and ALDS were used to assess disability status. HRQL was measured with the Short Form-36 and the Parkinson’s Disease Quality of Life Questionnaire. The internal consistency reliability of the ALDS was good (Cronbach’s \(\alpha = 0.95\)), with 55 items extending the sufficient item-total correlation criterion (\(r > 0.20\)). The ALDS was correlated with other disability measures (\(r = 0.50 – 0.63\)) and decreasingly associated with measures reflecting impairments (\(r = 0.36 – 0.37\)) and mental health (\(r = 0.23 – 0.01\)). The ALDS indicated that patients with more severe PD (H&Y stage...
3) were more disabled than patients with mild (H&Y stage 1) or moderate PD (H&Y stage 2) \( p < 0.01 \). Additionally, the ALDS discriminated between patients with more or less severe extrapyramidal symptoms \( p < 0.01 \) and patients with or without postural instability \( p < 0.01 \). Compared to the S&E (19%), the ALDS showed less of a ceiling effect (5%). We concluded the ALDS is a flexible, feasible and clinimetrically promising instrument to assess the level of disability in newly diagnosed PD patients.

Part 2 of Chapter 5 extends the results described above by comparing the ALDS item bank with the most often used disability scale in PD: the ALD part of the UPDRS. Both scales turned out to be able to discriminate between patients with more or less severe extrapyramidal symptoms \( p < 0.01 \) and patients with or without postural instability \( p < 0.01 \). However, in contrast to the ALDS \( p < 0.01 \), the UPDRS-ADL could not distinguish patients with moderate (H&Y stage 2) or severe (H&Y stage) PD \( p = 0.59 \).

Chapter 6 showed the results of a clinimetric evaluation of the ALDS in the acute phase of strokes as well as six months post stroke. At admission and discharge 213 stroke patients were evaluated using the NIH Stroke Scale, Barthel Index (BI), modified Rankin scale (mRs) and the ALDS. After six months the mRs and different subsets of ALDS items were assessed by telephone. The internal consistency (range Cronbach’s \( \alpha = 0.90 – 0.93 \)) and test-retest reliability (ICC = 0.85) of the ALDS was good. The ALDS was highly correlated with other disability measures \( r = 0.75 - 0.89 \) and less with the stroke scale \( r = 0.32 \). The mean ALDS scores were significantly different between ischemic or hemorrhagic strokes \( p = 0.008 \), and between mild, moderate or severe stroke at hospital admission \( p < 0.01 \). Disability level based on the ALDS improved significantly over time \( p < 0.01 \), whereas the responsiveness of the ALDS was moderate to large (Cohen’s \( d \) effect size = 0.77 – 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. The results of this study showed that the use of different subsets of items from the ALDS item bank had good clinimetric properties in the acute phase of strokes as well as later after stroke without floor or ceiling effects.

Chapter 7 presents the results of a study investigating the clinical meaning of the modified Rankin scale (mRs) score. Good outcome in stroke trials is generally defined as a mRs score of 0-1 or 0-2. A sample of 152 patients was assessed six months post stroke with the mRs and the ALDS. The mean probability to perform a selection of ALDS items per mRs grade and per type of dichotomization (0-1 or 0-2) was calculated. The ability to perform difficult ALDS items declined gradual with increasing mRs grade. When favorable outcome is defined as mRs 0-1, 15% of the cohort has a good outcome; of these patients 84% were likely to perform outdoor activities. If good outcome is defined as mRs 0-2, the percentage of patients with good outcome increased to 37%, whereas 66% of these patients were likely to perform outdoor activities. If good outcome is defined as the ability to perform outdoor activities mRs 0-1 should be chosen. If complex ADL is also considered as good outcome mRs 0-2 should be the dichotomization of choice. However, independent of which cutoff point will be chosen, the treatment effect in clinical trials must be large before good outcome is achieved. Therefore, it is likely that clinical important treatment effects can be missed in clinical trials with both these mRs endpoints.

When comparing outcomes of the ALDS between patients groups, the psychometric properties of the ALDS items should be consistent across groups of patients. IRT methods provide an ideal basis for assessing differential item functioning (DIF), defined as different probabilities of endorsing an item by respondents from two groups who are equal on the disability level. To examine the generic purpose of the ALDS item bank Chapter 8 provides such a DIF analysis. A cross-sectional, multi-centre study include 1283 in- and outpatients with a variety of diseases. The sample was divided in: 1) mainly neurological patients (vascular medicine, Parkinson’s disease, and neuromuscular diseases) and 2) patients from internal medicine (pulmonary diseases, chronic pain, rheumatoid arthritis, and geriatric patients). DIF was present in 18 items from the ALDS item bank when comparing the item difficulties between the two groups. We showed that the DIF was small and could effectively be modeled such that the ensemble of the items comprised a scale applicable in both groups. Therefore, no adjustment in terms of removing the concerning items is required.

In Chapter 9 the main findings are discussed briefly. The strengths and limitations of using IRT based methods in general, and the ALDS item bank specifically, in patient-oriented research is addressed. Finally, suggestions regarding the desired direction of the ALDS item bank in future clinical research are given.