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 1

Introduction
An estimated 10% of the world’s population experience some form of disability. The number of people with disabilities is increasing due to population growth, ageing, emergence of chronic diseases and medical advances that preserve and prolong life. The severity of illnesses can be measured using a wide range of physiological parameters, for example blood tests and imaging techniques. These parameters can often be measured very accurately and experienced clinicians find them easy to interpret. 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.

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 and handicap. Impairments, as a direct manifestation of the disease, reflect organ dysfunctions or abnormalities of body structure, such as paresis or dysarthria after stroke. Disabilities refer to the consequences of impairments in terms of the patient’s functional performance (e.g., immobility). Handicaps are concerned with the societal disadvantages resulting from impairments and disabilities, for example the loss of economic independence.

In an effort to develop an even more complete view of the consequences of disabling conditions, some investigators have called for greater examination of how those conditions affect the patients’ quality of life. Over the years health-related quality of life (HRQL) has been defined in many ways. A broad consensus has emerged that the concept of HRQL is multidimensional and at least four dimensions should be included in HRQL assessment: physical, functional, psychological and social health. These dimensions show substantial overlap with the elements of the ICIDH / ICF model. Physical aspects refer to impairments, functional aspects are defined in terms of disabilities, while the social aspects reflect the handicap dimension. The main difference between both models is that HRQL places a stronger accent on the subjective health perception of the patient.

In this thesis the central theme is the disability concept as defined in the ICIDH / ICF model. This includes mobility, basic activities of daily living (BADL), and instrumental activities of daily living (IADL). BADLs are activities in the personal care domain, such as: eating, toileting, transfers, dress and bathe. IADLs on the other hand are necessary for independent living in society, for example: prepare own meals, do light housework, manage own money, and shopping. In short, BADLs are necessary for survival, IADLs for maintaining a participant in a given socio-cultural setting.

### Classical Test Theory

So far, numerous generic and disease-specific instruments measuring disability have been developed. Most of these instruments are multi-item questionnaires, constructed by using the classical test theory (CTT). In spite of the popularity of the CTT sum based approaches, there are several problems associated with their use. Firstly, responses to all items on a scale are required to calculate a sum score. To make these scales more practical they are often shortened which results in less detailed scales. Detailed estimates of the status of patients spread across the whole range of functional levels can only be obtained with long questionnaires (e.g., the 136-item Sickness Impact Profile (SIP)). This may cost patients, clinicians and researchers an excessive amount of time to complete such instruments. Shorter instruments either cover a wide range of possible functional status (e.g., the 20-item Health Assessment Questionnaire Disability Index (HAQ-DI)) or remain detailed, but cover a smaller range of functional levels (e.g., the 10-item Barthel Index (BI)) measuring basic ADL activities (Figure 1). Secondly, since sum scores are dependent on the items included in the instrument, it is difficult to compare scores obtained on different instruments, even if they measure the same disability concept. Thirdly, the ordinal nature of sum scores makes it difficult to interpret a given difference in scores on the scale. A difference in scores at one point on the scale does not necessarily represent the same amount of functional change as an identical difference at another point on the same scale (e.g., a change in disability status on the HAQ-DI from 1 to 1.5 is not the same change as from 2 to 2.5).

![Figure 1](image.png)

Figure 1. Measurement range of three different disability scales.
Currently, there is a lot of interest in the flexible framework offered by item response theory (IRT). IRT measures at the item level, in contrast to the sum score methods, which are based on a whole instrument. This means that disability status can be assessed in a much more flexible way and that each patient can be presented with a smaller selection of items than is possible using sum score based methods. With IRT models, which are very similar to logistic regression models, it is possible to analyze response patterns of patients to estimate both item difficulty and patient ADL ability. Using this approach it is possible to place (calibrate) items on a hierarchical difficulty with linear measurement properties. The units of the scale are the regression coefficients and are expressed in logits or logits. The ADL ability of patients is expressed on the same linear scale as the items (Figure 2). As a consequence there are a number of advantages to the use of IRT in clinical measurement. One of the most interesting, is the implementation of an adaptive testing procedure, in which more difficult ADL items (e.g., ‘walk for more than 15 minutes’) are presented to less disabled patients and easier items (e.g., ‘get out of bed into a chair’) to more severely disabled patients. The estimates of disability level will be detailed and are completely comparable, even if patients are offered different selections of items. Adaptive testing can only be applied if an item bank is available. An item bank is a collection of items, which have been calibrated by obtaining psychometric information on the measurement properties of the items from large groups of patients. (For methodological details see Appendix 2).

**Figure 2. Adaptive testing using the AMC Linear Disability Score item bank.**

More difficult items are presented to more able patients, whereas easier items are presented to more severely disabled patients.

The Academic Medical Center (AMC) Linear Disability Score (ALDS) project aimed to construct an item bank to measure the disability status of patients with a broad range of diseases. The intention of the ALDS is to be used both in daily patient care and clinical research. The methodology and IRT statistics of constructing this item bank has already been described in depth. Other papers have examined technical and practical aspects of methods to deal with missing item responses and the use of a ‘not applicable’ response category for the items. Items for inclusion in the ALDS item bank were obtained from a systematic review of generic and disease-specific ADL scales. A total of 190 items were identified and then described in detail.

Patients are asked whether they can, rather than do, carry out the activities. Phrasing questions in terms of capacity may overestimate the disability level. On the other hand, phrasing items in terms of actual performance, may underestimate the disability level, since actual performance also depends on personal characteristics and interests. Moreover, measuring actual performance places patients in an unnatural ‘laboratory’ situation.

The ALDS uses two response options: ‘I can carry out the activity’ and ‘I cannot carry out the activity’. Patients are asked to rate if they are able to carry out the activities at this moment when they are at home or in their own area. When the patient is able to perform the activity independent, without any help from anybody else, the response ‘I can’ is recorded. Aids or devices are allowed. If a person is physically not able to perform an activity, needs help from somebody else or if the symptoms would increase, the response should be ‘I cannot’. If patients had never had the opportunity to experience an activity a ‘not applicable’ response is recorded. For example, responses from patients who have never held a driver’s license to the item ‘driving a car’ are recorded in this category. Responses in the category ‘not applicable’ are statistical treated as if the individual items had not been presented to the individual respondent.

In the development phase of the ALDS data was collected from over 1000 disabled patients with a broad range of conditions including stroke, Parkinson’s disease and chronic pain. The respondents were interviewed during a visit to inpatients’ or outpatients’ clinics, at the AMC. Also residents of supported housing, residential care or nursing homes were interviewed. Those geriatric patients had (co)morbid conditions such as movement and skeletal disorders, urologic, endocrine, gastrointestinal, pulmonary, cardiovascular and neurological diseases. The respondents attending an outpatient clinic for chronic pain filled in the ALDS questionnaire by themselves. All other respondents were interviewed by specially trained nurses or doctors.

In the calibration phase of the ALDS project the collected data was fitted using the two-parameter
logistic IRT model. This model was chosen because it allows a more realistic model for the data to be built than the more restrictive one-parameter logistic model. Additionally, the one-parameter model has been shown to be unsuitable as a final model for describing data resulting from disability items because it is to restrictive. (for more methodological details see Appendix 2)

An important IRT assumption is that the items included in an item bank have the same measurement characteristics for subgroups in the population. Differential item functioning (DIF) occurs when patients in two groups have different probabilities to respond to an item in a given way, even though they have the same level of disability status. If DIF is ignored, inaccurate measurements are obtained and true differences between patient groups may be obscured and non-existent differences ‘created’. For example, ‘preparing a meal’ may be easier for women than for men, the opposite may be true for ‘lifting heavy objects’. If these items are used to measure the disability status of a group of patients, the results for men and women are not completely comparable. Therefore, items with different measurement characteristics for younger and older, and for male and female patients were removed from the item bank.

Outline of this thesis

The main subject of this clinimetric thesis is the clinical application of the ALDS item bank in a variety of diseases. But first, in Chapter 2, we investigate if the disability continuum is the most appropriate patient-reported outcome in clinical efficacy studies. Disability and HRQL both reflect the consequences of diseases on personal level. However, they are conceptually different and not synonymous as is often thought by clinical researchers. The decision to use one of these measures can have important implications for the interpretation of the study results. Using a meta-analytic approach we examined the interchangeability of these two measures by comparing their association with other measures with known measurement properties.

In Chapter 3 we present the content and the psychometric properties of the ALDS item bank in its present state. Chapters 4 to 7 examine the use of the ALDS item bank in patients with rheumatoid arthritis (RA) and neurological diseases. Chapter 4 concerns the clinimetric properties of the ALDS item bank in a population of RA patients in comparison with the HAQ-DI. The HAQ-DI has become the most frequently used and validated functional disability scale in rheumatology. A new instrument, such as the ALDS, should not only be shorter, and better on theoretical basis, but it must also have clinimetric properties at least as good as the original HAQ-DI.

The first part of Chapter 5 focuses on the metric behavior of the ALDS in patients with newly diagnosed Parkinson’s disease compared to the often used Schwab & England scale. Whereas the second part of Chapter 5 includes a short communication concerning the concurrent validity of the ALDS with the ‘gold standard’: the ADL section of the Unified Parkinson’s Disease Rating Scale.

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, in general, in acute stroke another disability scale is preferred than six months post stroke. In Chapter 6 the use of different subsets of items from the ALDS item bank in patients in the acute phase and six months post stroke are examined.

Treatment outcomes in stroke trials are generally dichotomized reflecting ‘poor’ versus ‘good’ outcome. Concerns over dichotomized endpoints concentrates on the insensitivity for changes which are clinical relevant and using different cutoff points on a scale may lead to different interpretations of trial results. Using a sensitive range of ALDS items the clinical meaning of dichotomizing a well-known stroke outcome scale, the modified Rankin scale, is investigated in Chapter 7.

When comparing outcomes of the ALDS between patients groups, item characteristics of the ALDS should be consistent across groups. To investigate the generic character of the ALDS item bank, Chapter 8 provides a differential item functioning analysis were the item difficulties in a group of patients with various disorders has been compared.

Finally, Chapter 9 presents a general discussion on the strengths and weaknesses of the ALDS item bank and suggestions are given for future research using IRT based outcome measurements in general and the ALDS in particular. A summary in English and Dutch concludes this thesis.

References


