The AMC Linear Disability Score (ALDS) : measuring disability in clinical studies
Weisscher, N.

Citation for published version (APA):
Weisscher, N. (2008). The AMC Linear Disability Score (ALDS) : measuring disability in clinical studies
Appendix 2

Methodology of the ALDS item bank
The construction of an item bank is a complicated process, which can be split into four phases: (1) definition of content; (2) choice of calibration design; (3) data collection; and (4) fitting the IRT model.

**Definition of content**

It is important to define the concept to be measured and the patient population of interest carefully. When defining the concept, it can be useful to examine definitions given in previous studies, to study theoretical models for illness and health outcomes and consider whether the definitions given are likely to result in a unidimensional construct. Similarly, a useful starting point for identifying items is a review of existing instruments, to gain insight into how others have seen the construct.1

A large number of potential items should be identified, since it will not be possible to model the response pattern to a proportion of the items using an IRT model. The number of potential items can be increased by asking patients or healthy volunteers to keep diaries of health related activities, especially the difficult items. It is also important to consider how the data will be gathered from the patients, the number of scoring categories per item and how those categories are to be assigned to the responses made by patients. Using two, or at most three, response categories results in stability of scoring across time and researchers and increases clinical interpretability.2

**Choice of calibration design**

The calibration design to be used in the construction of an item bank describes which items are assigned to the responses made by patients. Using two, or at most three, response categories from the patients, the number of scoring categories per item and how those categories are to be assigned to the responses made by patients. Using two, or at most three, response categories results in stability of scoring across time and researchers and increases clinical interpretability.2

In the ALDS project, data was collected with an incomplete anchored calibration design using different booklets, ranging from difficult to very easy. The most difficult booklet contained activities which can only be carried out by those who are relatively healthy, and presented to less disabled patients. The easiest booklet included items which can be carried out by all except the most severely disabled patients and accordingly presented to those patients. Half of all the items in a given booklet are common with the booklet above and the other half with the booklet below, meaning that each item is in two booklets and the whole design is anchored. This design was chosen because it allowed a lot of statistical information to obtained, whilst keeping the burden on patients as low as possible.

**Data collection**

The aim of calibrating an item bank is to obtain information on the measurement properties of the items and on the fit of the IRT model chosen to analyze the responses given by patients to the items. When the item bank is implemented, perhaps in conjunction with computerized adaptive testing,4 the emphasis shifts to estimating the health status of the individual patient. When using IRT models little statistical information on item parameters is obtained from patients, whose ability level is a lot different from the overall difficulty of the items. The precise point at which most information can be obtained varies according to the IRT model used and the value of the item parameters themselves.5 However, in general, an item bank can be efficiently calibrated if the difficulty of items in a booklet is roughly matched to the ability of the patients.

The data for the ALDS project was collected from disabled patients with a broad range of conditions. The patients were interviewed during a visit to one of the neurology, rheumatology, pulmonology, internal medicine, vascular surgery, cardiology, rehabilitation medicine and gerontology outpatients’ clinics at one general and two teaching hospitals in Amsterdam, the Netherlands. Each patient was presented with one of the item sets in the calibration design. In order to increase the statistical efficiency of the design, the nurses interviewing the patients roughly matched the ‘difficulty’ of a booklet to the ability level of each patient, using their clinical experience. Hence, the easiest booklet was only presented to patients with substantial disabilities and the most difficult to those with minimal impairments. In practice, if a patient was able to carry out fewer than ten or more than twenty of the activities described in each booklet to which they were allocated, the patient was re-assessed using an easier or more difficult booklet as appropriate.

**Fitting the IRT model**

Usually, a preliminary choice of IRT model will have been made before the data are collected. But before fitting an IRT model and examining its quality, it is useful to carry out a number of
preliminary analyses. An overall impression of the data can be obtained by counting the number of patients who responded in each of the response categories to each item. It is difficult to obtain accurate estimates of parameters for items, to which the vast majority of responses say over 90%, are in a single response category. In addition, items with these characteristics do not contribute to the quality of measurements obtained using the item bank. Furthermore, items to which a substantial proportion, say over 10%, are in categories such as ‘not applicable’ or ‘don’t know’ may not be suitable for the patient population being used to calibrate the item bank.

IRT consists of a family of models, each designed to describe a relationship between the patients’ ability and the characteristics of the items. Selection of an IRT model is based in part on whether the assumptions of the model make sense for the data. The different kinds of IRT models are distinguished by the functional form specified for the relationship between the underlying ability and item response probability. There is a distinction between models for polytomous or dichotomous response options. In the thesis we focus on dichotomous models. The features of the three main types of dichotomous models are summarized in the Table. Each of these models estimates an item difficulty parameter. The two- and three-parameter models also estimate an item discrimination parameter; that is, the degree to which the item discriminates between persons in different regions on the ability continuum. Finally, the three-parameter model includes a ‘guessing’ parameter, often used in educational measurement with multiple-choice questionnaires.

The one-parameter model has been shown to be unsuitable as a final model for describing data resulting from disability items because it is too restrictive; the Rasch assumption of equal item discriminations may filter out the most highly discriminating items. Therefore, the more realistic two-parameter logistic IRT model was fitted to the ALDS data. Additionally, in the two-parameter model, the probability, \( P_{\theta_i}(\theta_k) \), that patient \( k \) responds to item \( i \) in the category ‘can carry out’ is

\[
P_{\theta_i}(\theta_k) = \frac{\exp (\alpha_i (\theta_k + \beta_i))}{1 + \exp (\alpha_i (\theta_k + \beta_i))}
\]

where \( \theta_k \) represents the functional status of patient \( k \). In addition, \( \alpha_i \) denotes the discrimination parameter and \( \beta_i \) the difficulty parameter for item \( i \).

An important assumption in item banking is that the items included in an item bank have the same measurement characteristics for all subgroups in the population. Differential item functioning of the items was examined using the one-parameter Rasch model by means of investigating if the item difficulty parameter \( \beta_i \) was similar for male and female and for younger and older patients. The cutoff point between younger and older patients was the median age. Items were excluded from further analysis if the parameter was more than half of the value of the standard deviance of the underlying distribution of ability parameters \( \theta \). Dimensionality of the item bank was examined using IRT based full information factor analysis. An exploratory factor analysis was carried out on different item sets used. To examine the population as a whole, a confirmatory factor analysis was carried out using data from all respondents. In addition, Cronbach’s alpha coefficient was calculated for each item set and for all of the data. The statistical analysis and results have been described in more detail in a previous paper.

References

Appendix 3

The ALDS item bank: 77 items and their measurement properties